A 12 años del CIBPsi

Roberto Aguirre y Maria Noel Macedo
(editores)

Montevideo, 2023

CIBPsi
CENTRO DE INVESTIGACIÓN BÁSICA EN PSICOLOGÍA

Facultad de Psicología
UNIVERSIDAD DE LA REPÚBLICA

****

Líneas de investigación vigentes a 2023 / Research lines at 2023

Atención (Attention)
Cerebro y Lenguaje (Brain and Language)
Cognición Numérica (Numerical Cognition)
Comportamiento alimentario (Food Behavior)
Evaluación del Desarrollo y Temporalidad
Mente, Acción y Lenguaje (Mind, Action, and Language)
Neurociencia Cognitiva y Salud Mental
(Cognitive Neurosciences and Mental Health)
Percepción e Interacción (Perception and Interaction)

****

Maquetación / Page layout: Facundo Antunez y Roberto Aguirre

****

Esta publicación se realizó sin fines de lucro. Su distribución es gratuita. Esta publicación está formada por versiones libres de los pre-prints u originales de acceso abierto de artículos publicados en revistas científicas arbitradas. Las adaptaciones de los artículos se hicieron con el consentimiento de los autores. El estilo editorial se ajustó al formato de los artículos de seis páginas de las reuniones anuales de la *Cognitive Science Society*. Sin embargo, algunos aspectos de la citación y las referencias se mantuvieron iguales al formato de las revistas de publicación de los artículos finales. Las imágenes fueron aportadas por los miembros del centro. La versión en pdf de este documento está accesible en la página web del centro https://cibpsi.psico.edu.uy/ con el nombre MEMORIA12AÑOS.pdf

This publication was published on a non-profit basis. Its distribution is free of charge. This memorial includes free versions of the pre-prints or open-access articles published in peer-reviewed scientific journals. Adaptations of the articles were made with the consent of the authors. The editorial style conformed to the format of the six-page papers of the annual meetings of the Cognitive Science Society. However, some aspects of citation and referencing were kept the same as the format of the final article publication journals. The members of the center contributed images. The pdf version of the document is available at the web of the center https://cibpsi.psico.edu.uy/ with the name MEMORIA12AÑOS.pdf

Índice
Index

Prólogo / Prologue

Es una gran satisfacción observar, a través de los artículos de este libro, el camino recorrido por el Centro de Investigaciones Básicas en Psicología. Recuerdo las conversaciones con Alejandro Maiche, en la Comisión Sectorial de Investigación Científica (CSIC), mientras nacía este proyecto. Eran años intensos. El rectorado de Rodrigo Arocena avanzaba y la universidad se transformaba. Nacían programas para fortalecer la investigación más exigente, como el programa de Grupos, y para apoyar a aquellos sectores más débiles como el Programa de Fortalecimiento de la Investigación de Calidad. Los programas de equipamiento y de apoyo a publicaciones atendían aspectos que hasta entonces se sostenían en esfuerzos esporádicos. Las Jornadas Ancap-Udelar y el Programa de Investigación Orientado a la Inclusión Social buscaban articular la demanda de conocimiento con las capacidades de nuestra universidad. Nacía el Programa de Apoyo a la Investigación Estudiantil (PAIE). El nuevo impulso al desarrollo en el interior imaginaba ya la instalación de numerosos grupos de investigación en diversos puntos del país. Nacía el Espacio Interdisciplinario y se fortalecía la extensión como apuestas fuertes a romper fronteras. Se discutía la Ordenanza de Grado que cambiaría casi totalmente los planes de estudio. El proyecto Flor de Ceibo juntaba estudiantes de todas las carreras en grupos que acompañaban a las familias de los niños que recibían las ceibalitas. La lista es enorme y muestra una ebullición y un entusiasmo que sigue emocionando.

Fue en ese contexto que la Facultad de Psicología transformó a fondo su estructura y su manera de ser. El nacimiento del CIBPsi está íntimamente ligado a todo ello. Cuando uno hojea los artículos seleccionados para este número casi puede sentir latir con vida propia toda esa energía. Allí están la interdisciplina, los PAIE, las ceibalitas, el equipamiento, la articulación del conocimiento con las necesidades sociales, la gente que volvió al país en ese tiempo y se sumó con entusiasmo a la patriada.

Las conversaciones con Alejandro a veces buscaban facilitar aspectos concretos del proyecto, lograr que alguien viniera al país, por ejemplo. Pero no pocas veces se convertían en discusiones sobre cómo encarar las transformaciones, qué hacer ante las dudas o las resistencias, cómo construir algo que fuera transformador y a la vez estable. Han pasado doce años, no tantos en la vida de una institución académica, y sin embargo suficientes para calibrar el camino recorrido y darse cuenta de que fue una muy buena apuesta. El CIBPsi es hoy una realidad reconocida, con fuertes vínculos adentro y afuera de la universidad, respetado y apreciado. Da gusto ver que aquellos balbuceos se han convertido en la fuerza adolescente de un proyecto que tiene mucho por delante.

It is a great pleasure that we observe, through the articles in this book, the path taken by the Centro de Investigación Básica en Psicología. I remember meetings with Alejandro Maiche, at the Comisión Sectorial de Investigación Científica (CSIC), while this project was being born. Those were intense years. Rodrigo Arocena's rectorship was advancing and the university was being transformed. New programs were born to strengthen the most demanding research, such as the Groups program, and to support the weakest sectors, such as the Programa de Fortalecimiento de la Investigación de Calidad. The equipments and publications support programs addressed aspects that until then had been sustained by occasional efforts. The Ancap-Udelar Meetings and the Programa de Investigación Orientado a la Inclusión Social sought to articulate the demand for knowledge with the capabilities of our university. The Programa de Apoyo a la Investigación Estudiantil (PAIE) was initiated. The new impulse for the university development in the countryside already envisioned the establishment of numerous research groups in different parts of the country. The Espacio Interdisciplinario was began and the outreach was strengthened as strong bets to breaking frontiers. The new Undergraduate Studies Regulation that would change the study plans throughout the University was being discussed. The Flor de Ceibo project gathered students from all majors in groups that accompanied the families of the children who received the ceibalitas*. The list is enormous and shows an ebullience and enthusiasm that continues to thrill.

It was in this context that the Facultad de Psicología deeply transformed its structure and way of being. The birth of CIBPsi is intimately linked to all this. When you leaf through the articles selected for this memorial book, you can almost feel all that energy pulsating with a life of its own. Interdisciplinary, the PAIEs, the ceibalitas, the new equipment, the articulation of knowledge in line with social needs, and those who returned to the country then and enthusiastically took part in what was happening.

The dialogues with Alejandro were sometimes aimed at helping specific aspects of the project, such as getting someone to come to the country. But they often turned into discussions on how to approach the transformations, what to do in the face of doubts or resistance, and how to build something that would be both transformative and stable at the same time. Twelve years have gone by, not so many in the life of an academic institution, but enough to gauge the road traveled and to recognize that it was a very good bet. CIBPsi is today a well-known reality, with strong links inside and outside the university, respected and appreciated. It is good to see that those stammerings have become the adolescent strength of a project with a long way to go.

Gregory Randall

* ceibalitas. El término se refiere a las portátiles entregadas a niños y docentes de la educación primaria pública como componente del proyecto mencionado. / The term refers to the laptops delivered to children and teachers in Uruguayan public primary education as components of the aforementioned project .

Introducción

El 22 de septiembre de 2010 el Consejo de la Facultad de Psicología de la Universidad de la República creó el Centro de Investigación Básica en Psicología (CIBPsi). A través del centro, un conjunto interdisciplinario de investigadores uruguayos —formados dentro y fuera del país— y algunos extranjeros empujaron y profundizaron en una rama de la Psicología —la Psicología cognitiva— con escaso desarrollo hasta el momento en nuestro país. Desde los inicios, el CIBPsi fue concebido como un espacio superador de los encapsulamientos disciplinares: un espacio promovido por la Facultad de Psicología pero que albergó desde su fundación a profesores de las facultades de ciencias, de Química, de Veterinaria, entre otras procedencias. A la interna de la Universidad, el CIBPsi significó un espacio de legitimación de la vocación de transformación académica de la Facultad de Psicología. A la interna de la Facultad de Psicología, significó el comienzo de un proceso donde el estudio de las funciones cognitivas es parte de la formación de los psicólogos y permitió estructurar la enseñanza de las técnicas y métodos de investigación propios de las ciencias empíricas.

A la interna de la Universidad, el CIBPsi significó un espacio de legitimación de la vocación de transformación académica de la Facultad de Psicología. A la interna de la Facultad de Psicología, significó el comienzo de un proceso donde el estudio de las funciones cognitivas es parte de la formación de los psicólogos y permitió estructurar la enseñanza de las técnicas y métodos de investigación propios de las ciencias empíricas. De manera más extendida, el CIBPsi ha colaborado en la emergencia de las Ciencias Cognitivas uruguayas y su impacto en ámbitos de relevancia social tales como: la educación, la salud, la inclusión o la alimentación. El entorno del estudio de la cognición en 2010 no es el mismo de hoy. Hay más actores. Hay esfuerzos nacionales que, por fortuna, se acompañaron en el tiempo con la creación del CIBPsi. Iniciativas, organizaciones y plataformas como el Núcleo Interdisciplinario de Ciencias Cognitivas, la Latin American School for Education, Cognitive and Neural Sciences, el Centro de Investigación en Cognición para la Enseñanza y el Aprendizaje (CICEA), la Maestría en Ciencias Cognitivas, la Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento (SUCCC), el Laboratorio de Psicología Experimental del CENUR sede Salto (Labpex) o el Laboratorio de registro de la conducta mediante señales de video en el CICEA son algunos de los ejemplos que dan cuenta de esfuerzos en diferentes planos que involucran la participación de personas que originalmente estuvieron vinculadas al CIBPsi.

Con alegría, podemos decir que el CIBPsi fue una pieza fundamental en algunas innovaciones científicas para el país como pueden ser las primeras intervenciones cognitivas en las escuelas (con el apoyo del Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea), los primeros estudios cognitivos con resonancia magnética funcional (con el apoyo del Centro Uruguayo de Imagenología Molecular), los primeros estudios de electroencefalografía con un equipo de 64 canales o los estudios empíricos del procesamiento mental de la lengua de señas uruguaya con medios experimentales. Hay temas y abordajes científicos —novedosos y otras veces no tanto— que llegaron a Uruguay de la mano de nuestros colaboradores y han permitido formar, poco a poco, una nueva generación de científicos cognitivos provenientes de distintas disciplinas.

Entendemos que, en la actividad científica, la comunicación a la sociedad de los resultados de investigación -en sus distintos formatos- es un fin por sí mismo y una finalidad de la ciencia. En este horizonte, la publicación de artículos científicos es un hito que requiere mucho esfuerzo, energía y destreza. En esta publicación conmemorativa queremos compartir con ustedes catorce publicaciones científicas que los miembros de nuestras actuales líneas de investigación consideraron emblemáticas, significativas y objeto de nuestro afecto compartido porque reflejan las preguntas de investigación que nos hemos hecho en estos 12 años.

Además, esperamos que esta publicación también ayude a compartir y comunicar nuestro entusiasmo y afecto por los esfuerzos realizados, la dimensión cotidiana, lúdica de dicho esfuerzo y los atisbos de un nuevo ciclo. Los artículos aquí compilados son resultado de la colaboración del CIBPsi con colegas y equipos nacionales e internacionales. Por esta razón, anexamos imágenes representativas de algunos obsequios, instrumentos, eventos, situaciones y testimonios voluntarios de algunos investigadores que hemos desarrollado una parte significativa de nuestra actividad académica en nuestro centro.

Profr. Tit. Dr. Alejandro Maiche
Primer director

Profr. Adj. Dr. Roberto Aguirre
Director 2021-2022

Introduction

On September 22, 2010, the Council of the Facultad de Psicología of the Universidad de la República created the Centro de Investigación Básica en Psicología (CIBPsi). Through the center, an interdisciplinary group of Uruguayan researchers —trained inside and outside the country— and some foreigners pushed and deepened in a branch of Psychology —cognitive Psychology— with little development so far in our country. From the beginning, the CIBPsi was conceived as a space that overcame disciplinary encapsulation: a space promoted by the Facultad de Psicología but which, since its foundation, hosted professors from the faculties of Ciencias, Química, Veterinaria, and Medicina, among other backgrounds. Within the University, the CIBPsi meant a space of legitimization of the academic transformation vocation of the Facultad de Psicología. Within these faculty, it meant the beginning of a process where the study of cognitive functions is part of the training of psychologists and allowed structuring the teaching of techniques and research methods of the empirical sciences.

Within the university, the CIBPsi meant a space of legitimization of the vocation of academic transformation of the Faculty of Psychology. Within the Facultad de Psicología, it meant the beginning of a process where the study of cognitive functions is part of the training of psychologists and allowed structuring the teaching of techniques and research methods of the empirical sciences. In a more extended way, CIBPsi has collaborated in the emergence of the Uruguayan Cognitive Sciences and its impact in areas of social relevance such as: education, health, inclusion or food. The environment for the study of cognition in 2010 is not the same as today. There are more actors. There are national efforts that, fortunately, were accompanied in time with the creation of the CIBPsi. Initiatives, organizations and platforms such as the Núcleo Interdisciplinario de Ciencias Cognitivas, the Latin American School for Education, Cognitive and Neural Sciences, the Centro de Investigación en Cognición para la Enseñanza y el Aprendizaje (CICEA), the Maestría en Ciencias Cognitivas, the Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento (SUCCC), the Laboratorio de Psicología Experimental del CENUR Salto (Labpex) or the Laboratorio de registro de la conducta mediante señales de video at CICEA are some of the examples that show efforts at different levels involving the participation of people originally linked to CIBPsi.

We can happily say that CIBPsi was a fundamental piece in some scientific innovations for the country, such as the first cognitive interventions in schools (with the support of the Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea), the first cognitive studies with functional magnetic resonance imaging (with the support of the Centro Uruguayo de Imagenología Molecular), the first electroencephalography studies with 64-channel equipment or the empirical studies of mental processing of Uruguayan sign language with experimental means. There are topics and scientific approaches -novel and sometimes not so novel- that arrived in Uruguay thanks to our collaborators and have allowed the formation, little by little, of a new generation of cognitive scientists from different disciplines.

We understand that, in scientific activity, the communication to society of research results —in its different formats— is an end in itself and a purpose of science. In this context, the publication of scientific articles is a milestone that requires a great deal of effort, energy and skill. In this commemorative publication we want to share with you fourteen scientific publications that the members of our current research lines considered emblematic, significant and the object of our shared affection because they reflect the research questions we have asked ourselves in these 12 years.

In addition, we hope that this publication will also help to share and communicate our enthusiasm and affection for the efforts made, the daily, playful dimension of this effort and the glimpses of a new cycle. The articles compiled here are the result of CIBPsi's collaboration with colleagues and national and international teams. For this reason, we attach representative images of some gifts, instruments, events, situations, and voluntary testimonies of some researchers who have developed a significant part of our academic activity in our center.

Prof. Tit. Dr. Alejandro Maiche
First Director

Profr. Adj. Dr. Roberto Aguirre
Director 2021-2022

Taza decorativa de ilusiones visuales
Visual illusions decorative mug

Ésta fue la **primera taza** que se obsequió a los voluntarios por su participación en los experimentos e investigaciones del **CIBPsi**. Entre 2014 y 2016, los investigadores del centro obsequiaban esta taza a sus voluntarios.

This was the **first decorative mug** given to volunteers for their participation in **CIBPsi** experiments and research. Between 2014 and 2016, researchers from the center gave this mug as a gift to their volunteers.

Do potential past and future events activate the Left-Right Mental Timeline?

Roberto Aguirre[1] & Julio Santiago[2]

1. **Centro de Investigación Básica en Psicología (CIBPsi)**. Universidad de la República. Montevideo, Uruguay
2. Mind, Brain, and Behavior Research Center. University of Granada. Granada, España

Roberto Aguirre https://orcid.org/0000-0001-8386-1380, Julio Santiago https://orcid.org/0000-0002-0346-2740

Correspondence concerning this manuscript should be addressed to Roberto Aguirre, Facultad de Psicología, Tristán Narvaja 1674, Montevideo 11200, Uruguay. Email: **zimmer20uy@gmail.com**:

This is a free version -with the approval of the authors- of the pre-print of the manuscript Aguirre, R., & Santiago, J. (2017). Do potential past and future events activate the Left-Right Mental Timeline? *Psicológica*, *38*(2), 231-255. Available at **https://www.redalyc.org/pdf/169/16951418004.pdf**

Abstract

Current evidence provides support for the idea that time is mentally represented by spatial means, i.e., a left-right mental timeline. However, available studies have tested only factual events, i.e., those which have occurred in the past or can be predicted to occur in the future. In the present study we tested whether past and future potential events are also represented along the left-right mental timeline. In Experiment 1, participants categorized the temporal reference (past or future) of either real or potential events and responded by means of a lateralized (left or right) keypress. Factual events showed a space-time congruency effect that replicated prior findings: Participants were faster to categorize past events with the left hand and future events with the right hand than when using the opposite mapping. Crucially, this also ocurred for potential events. Experiment 2 replicated this finding using blocks of trials comprising only potential events. In order to assess the degree of automaticity of the activation of the mental timeline in these two kinds of events, Experiment 3 asked participants to judge whether the expressions referred to factual or potential events. In this case, there was no space-time congruency effect, showing that the lateralized timeline is active only when relevant to the task. Moreover, participants were faster to categorize potential events with the left hand and real events with the right hand than when using the opposite mapping, suggesting for the first time a link between the mental representations of lateral space and potentiality.

Keywords: Conceptual Metaphor; Mental Timeline; Time; Space; Potentiality

Introduction

A large number of studies support the suggestion by Lakoff and Johnson (1980) that space is used to conceptualize time. Among other possibilities, time can be represented as flowing from left to right in space, at least in languages with a left-to-right orthography (see Santiago, Lupiáñez, Pérez, & Funes, 2007, for Spanish; Tversky, Kugelmass, & Winter, 1991, for English; Ulrich & Maienborn, 2010, for German). Santiago et al. (2007) presented verbs referring either to the future or to the past, and participants categorized their temporal reference by pressing either a left or right response key. Responses were faster when past words were responded to with the left hand and future words with the right hand in comparison to a reversed mapping condition. Space-time congruency effects such as this one have been interpreted as evidence of the use of an underlying left-to-right mental timeline.

Yet, all available studies of the lateralized mental timeline have used past and future factual events. Some studies have used single words (temporal adverbials and tensed verbs: Flumini & Santiago, 2013; Ouellet, Santiago, Israeli, & Gabay, 2010; Santiago et al., 2007; Torralbo, Santiago, & Lupiáñez, 2006; Weger & Pratt, 2008). Others have used short adverbial phrases (Casasanto & Bottini, 2014) or whole sentences (Ulrich & Maienborn, 2010). Still others have used sequences of events which can be objectively placed in temporal succession (Fuhrman & Boroditsky, 2010; Santiago, Román, Ouellet, Rodríguez, & Pérez-Azor, 2010). The aim of the present research is to test whether potential events are also able to activate the left-right mental timeline. To our knowledge, this is the first study to tap directly onto this question.

The ability to represent potential events, both past and future, is central to human cognition. Developmental studies suggest that this ability is reached only after the child enters the formal operational stage (Piaget & Inhelder, 1985). Moreover, children with learning problems often have difficulties in imagining potential events (Schlemenson, 2004). Representing potential events supports the manipulation of alternative scenarios and the evaluation of their consequences in order to make decisions about courses of action (Baumeister & Masicampo, 2010; Hegarty, 2004; Johnson-Laird, 1980). Past potential events are a necessary component of counterfactuals (e.g., “If I had been your father, I hadn't allowed you to do it”; see Gilead, Liberman, & Maril, 2012). They are also related to studies of the processing of negation (as any potential past event is something that did not happen). The mental representation of uncertain and negated events has recently arised strong interest from embodied approaches to language comprehension (Ferguson, Tresh, & Leblond, 2013; Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007; Orenes, Beltrán, & Santamaría, 2014; De Vega et al., 2014). If comprehension is mediated by detailed, modal mental simulations of linguistic content, as these approaches propose, uncertain and negated events pose an important theoretical challenge. Kaup et al. (2007) showed that, when the context implies a choice between two alternative events, speakers create a simulation of the negated actions separated from the simulation of the real events. In the same line, De Vega and Urrutia (2012) suggested that negations could momentarily activate a counterfactual representation of the negated events as if they had actually happened, followed by the representation of the real events, and De Vega et al (2014) observed commonalities in the brain activations induced by negations and counterfactuals. These findings suggest that, at least in some contexts, counterfactual and negated events may be internally simulated in the same way as real events are.

Spanish verb forms provide a great opportunity to assess the strength of space as a ground domain of temporal metaphorical mappings when the events are potential. Potentiality is encoded in syntax and morphology in many ways, but a central strategy is by means of verb aspect, in particular, mood. Spanish has a set of verb forms that stereotypically encode potentiality (e.g., for present future: “Él dormiría” - “He would sleep”, and “(Si) él durmiera...” - “(If) he would sleep...”; for past: “Él habría dormido (si)...” - “He would have slept (if)...” y “(Si) él hubiera dormido...” - “(If) he had slept”) (Ruiz Campillo, 2014). Note that English translations are only approximate, as Spanish verb forms provide clearer information regarding mood than many other languages. For example, the closest English verb form to the Spanish Pretérito Pluscuamperfecto de Subjuntivo (Subjunctive Pluperfect Past, a stereotypical verb form for events that might have happened, but did not; “si yo hubiera sonreído, ella me habría mirado”) is the Third Conditional (“If I had smiled, she would have looked at me”). The English Third Conditional is a complex construction that requires a Past Perfect verb in the conditional clause and a conditional construction in the main clause. Thus, the Past Perfect is used in English for past events occurring before a past reference point both when the event was real as well as when it was potential, and both uses are common. In contrast, the Spanish Pretérito Pluscuamperfecto de Subjuntivo is used preferentially for potential events, thereby providing a clearer potentiality marker. Additionally, Spanish verb endings and auxiliaries provide simultaneously information about potentiality (mood) and time (tense). For example, in “comió” (“he ate”), “-ió” indicates a past factual event. Although verb use in context may sometimes differ from preferred interpretations, Spanish speakers can recognize the stereotypical mood and tense of verb forms in simple sentences without much context. The present studies used this kind of sentences.

Prior research has shown that the mental simulations of concrete factual events activate a left-right mental timeline. The present study aims to shed light on whether potential events are also mentally arranged along a left-right axis. In order to answer this question, the present study used a standard space-time conceptual congruency task following Santiago et al. (2007). In Experiment 1, factual past and future events were mixed with potential past and future events. Events were presented by means of short Spanish sentences containing a pronoun and a conjugated verb. The conjugation of the verb indicated whether the event was factual or potential as well as whether it was past or future. Participants were asked to categorize all sentences as referring to past or future by means of lateralized left and right keypresses. In one block they used a congruent mapping (left-past right-future) and in another block the mapping was reversed.

Experiment 1

We expected that potential events would activate the lateralized mental timeline as well as factual events do. Therefore, we predicted an interaction between temporal reference and response side both for factual and potential events. It is important to point out that only the interaction with response side is informative for this prediction. Because the conditions defined by the factors potentiality and time were not matched in stimulus length in characters, word frequency, verb form complexity, verb form frequency, and so on, we cannot make predictions regarding main effects nor interactions between factors other than response side. Time and potentiality are between-item factors, and therefore, their main effects or two-way interaction might arise because of uncontrolled item variables. In contrast, response side is a within-item factor, and therefore, its interaction with either time and/or potentiality cannot be accounted for by differences among items.

Methods

Ethics Statement In all the experiments reported in this paper, written informed consent was obtained from all participants. The studies were approved by the Committee for Ethics in Human Research of the University of Granada and the University of La República (Montevideo).

Participants Twenty-eight Psychology undergraduate students (32.5 mean age, one left-handed, 13 women) of the Autonomous University of Barcelona volunteered to participate without any compensation. All of them were native Spanish speakers. We conducted an a priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) with power (1 - β) set at 0.90 and α=.10, two-tailed, for an effect size of .30. This showed N=23 is a sample size big enough to detect a small-sized effect with a 90% probability.

Materials Verb forms were generated from 20 intransitive regular Spanish verbs conjugated in four paradigms (80 Spanish expressions in total, see Table 1 in the Appendix). No constraints were set on the kind of meanings expressed by these verbs, as they were not relevant to current concerns. We used intransitive verbs to avoid strong feelings of semantic anomaly when those verbs appear in sentences without an object. Sentence length was between 13 and 20 characters. The factual past condition used verbs in Indicative Perfect Past form (“ella despertó” - “she woke up”) and the factual future condition used verbs in Indicative Simple Future (“nosotros dormiremos” - “we will sleep”). The potential past condition used verbs in Subjunctive Pluperfect Past (“él hubiera trabajado”- roughly corresponding to “[if] he had worked”) and the potential future condition used Indicative Conditional (“ella se dormiría” - “she would fall asleep”). Note that the selected verb forms stereotypically mark factuality and potentiality in Spanish. Indicative Perfect Past and Indicative Simple Future stereotypically refer to events which have occurred in the past or will occur in the future, respectively. Spanish Subjunctive Pluperfect Past stereotypically refers to past events that could have happened but have not (thereby past potential events). The Spanish Conditional stereotypically refers to events that will happen in the future if some condition is met, with the condition remaining uncertain (thereby future potential events). Verb forms do not exhaust all the resources available in Spanish to mark the potentiality of an event. For example, potentiality can also be marked by the use of conditional conjunctions (e.g., “if”) or modal phrases (e.g., “it is possible that”, “it’s a fact that”). However, we decided to rely only on verb forms to mark potentiality in order to introduce as few differences as possible with already published studies about the timeline in Spanish using single verbs as stimuli (see Introduction) as well as to expose participants to time and potentiality information simultaneously. Lexical frequency, imaginability (concrete vs abstract words) and phrase length were not strictly controlled because the key contrast of interest in the present design compares the experimental sentences with themselves in two response conditions (left vs. right hand). In other words, we were not interested in main effects of temporal reference nor potentiality, but in their interaction with responding hand.

Procedure The experiment was programmed in E-Prime (Schneider, Eschman, & Zuccolotto, 2002) and run in a sound attenuated room. Stimuli were presented at the centre of a computer screen (spanning 6.23º of visual angle, in white letters over a black background). The distance between screen and participant was 0.60 m. One session lasted approximately 20 minutes. Participants pressed a left (“a”) or right (“6”) response keys on a keyboard. The keys were covered by stickers of the same colour. At the beginning of each trial a fixation cross was presented for 500 ms before a randomly chosen sentence appeared on the centre of the screen. It remained on screen until the participant’s response or a maximum time of 4,000 ms. Then there was an interval of 3,000 ms. Wrong responses were followed by a 440 Hz beep that lasted 500 ms. The next trial started 3,000 ms after a correct response or the offset of the auditory feedback. There were two experimental blocks, one for the congruent time-response mapping and the other for the incongruent mapping. In the congruent condition, participants pressed the left key in response to past verb forms (both real and potential), and the right key in response to future verb forms (both real and potential). In the incongruent condition, this mapping was reversed. The order of blocks was counterbalanced over participants. The whole set of 80 verbal stimuli was used in each block. Before each block there was a practice block of eight trials per condition. Written instructions were presented on screen at the beginning of each block.

Design Latency and accuracy were analyzed by means of repeated measures ANOVAs including the factors Potentiality (real vs. potential) X Time (past vs. future) X Response side (left vs. right) X Order of conditions (congruent-incongruent vs. incongruent-congruent). The Order of conditions factor was introduced to decrease error variance. However, because of its irrelevance to present hypotheses, its effects and interactions will not be reported further. The design was a factorial design with all factors manipulated within participants.

Open Materials and Data E-prime programs and full data sets for all the experiments in this study can be downloaded from https://osf.io/cxjhr/.

Results

Due to experimenter error, three verbal stimuli in the factual condition (“Nosotros silbamos” “Nosotros dormimos” and “Nosotros soñamos”) were ambiguous as to their conjugation (they take identical forms in Indicative Past and Present), and were removed. These represented 3.0% (168) of total trials. Response errors occurred on 6.0% (258) of the remaining trials and were excluded from the latency analysis. After discarding wrong item and error response trials, in order to avoid the influence of outliers we excluded latencies below 400 ms and above 3,500 ms, which amounted to discarding an additional 1.5% (62) of correct trials. The cut-offs were set by visual inspection of the reaction time (RT) distribution, at points where it was leveling off, with the preestablished requirement of not leaving out more than 2% of correct trials. The rejection rate was kept constant across experiments. Fixed cut-offs are a standard way to deal with outliers and they have both advantages and disadvantages when compared with other methods (Ratcliff, 1993). They are the method used in many of the prior studies on the timeline for real events (e.g., Santiago et al., 2007; Torralbo et al., 2006). By establishing cut-offs that leave out the same percentage of data points in all experiments, we made sure that the trimming of latencies was consistent across experiments that may have different grand means.

Reaction Time Analysis. Table 2 in the Appendix shows cell mean latencies and number of errors. Centrally for our hypotheses, Time interacted with Response side (F(1,27)=8.71, p=.006, η^2=.24). Post-hoc comparisons showed that both the contrast between Past-Left vs. Past-Right (p<.031) and between Future-Left vs. Future-Right (p<.001) reached significance. Moreover, there was no three-way interaction between Potentiality, Time, and Response side (F<1), indicating that the size of the interaction between Time and Response side was the same for both real and potential events. This was supported by independent repeated measures ANOVA analyses of the interaction between Time and Response side for real events (F(1,27)=8.33, p=.008, η^2=.24) and potential events (F(1,27)=7.56, p=.01, η^2=.22). Figure 1 illustrates these results. Additionally, there was an interaction between Potentiality and Time (F(1,27)=7.86, p=.009, η^2=.23), and no interaction between Potentiality and Response side (F(1,27)=1.34, p=.26, η^2=.05). There were significant main effects of both Potentiality (F(1,27)=6.94, p=.01, η^2=.21) and Response side (F(1,27)=5.05, p=.03, η^2=.16). There was no main effect of Time (F<1).

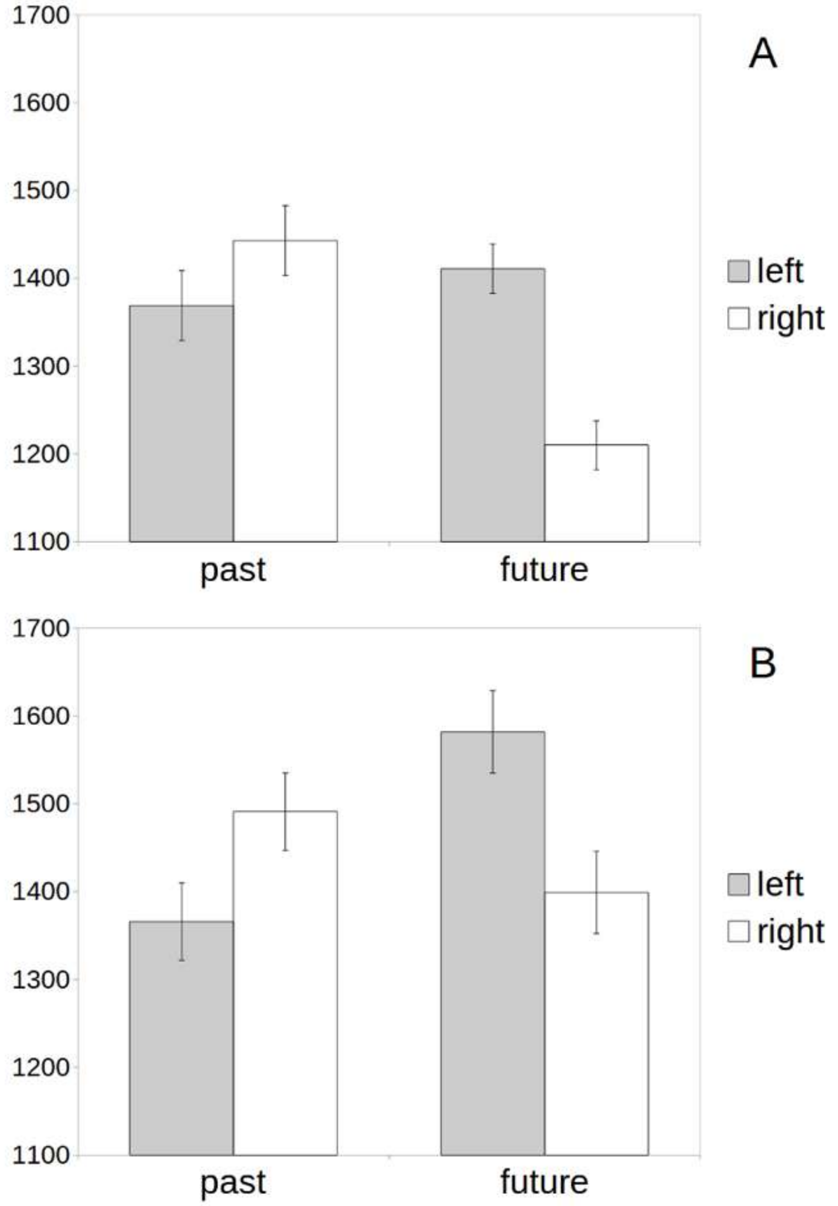


Figure 1. Mean latencies (ms) for factual and potential events in Experiment 1 (error bars show Standard Error of the Mean corrected for within-subject designs following Cousineau, 2005). Participants' task was to judge past versus future reference.

Accuracy Analysis The interaction between Time and Response side and the three-way interaction between Potentiality, Time, and Response side were non-significant (all Fs<1). There was also an unexpected interaction between Potentiality and Response side (F(1,27)=8.18, p=.008, $\eta2$=.23): accuracy was greater when potential events were responded to with the left hand and factual events with the right hand than when using the opposite mappings. This interaction was not predicted by present hypotheses and constitutes a novel finding. The interaction between Potentiality and Time (F(1,27)=13.21, p=.001, η^2=.33) was significant. There were no main effects (Potentiality: F(1,27)=3.69, p=.07, η^2=.12; Time: F<1; Response side: F(1,27)=2.20, p=.15, η^2=.08).

Discussion

Experiment 1 revealed a space-time congruency effect both for factual and potential events. Participants responded faster to both kinds of events when past was mapped to the left hand and future to the right hand than with the opposite mapping. The size of the effect was the same for both event kinds. This pattern of results suggests that past and future potential events are represented along the left-right mental timeline as well as factual events are.

However, there is an alternative explanation of the interaction between Time and Response side in the processing of potential events. On this account, by intermixing factual and potential trials and assigning response keys to past and future reference throughout the block we may have induced a carry-over of the space-time congruency effect from factual to potential trials. In other words, it is possible that potential trials only showed the left-right past-future congruency effect because they were intermixed with factual trials, which do show the effect. One way to sort out the carry-over account is to remove the factual trials altogether, keeping only the potential trials. The carry-over account is based on the possibility that factuality would play a role on activating the left-right past-future mental timeline. Additionally, accuracy data suggested the interesting possibility that potentiality by itself may be able to interact with left-right space, a previously unreported finding. We will delay a detailed discussion of this effect until we explore it further in Experiment 3, where participants were asked to discriminate factual from potential events.

Experiment 2

The aim of this experiment was to examine whether potential past and future verb forms are able to activate left and right space when presented in a context that does not include factual events. As in Experiment 1, the interaction between Time and Response side was the crucial prediction: we expected that performance would be better in the congruent conditions.

Methods

Participants Thirty-four Psychology undergraduate students of the Universidad de la República at Montevideo (26.8 mean age, 3 left-handed, 23 women) volunteered without compensation. They were all native Spanish speakers. Because Experiments 2 and 3 worked with the same parameters of Experiment 1, the previously conducted power analysis can also be used to estimate the minimum sample size in them. In both experiments, sample size was greater than this minimum.

Materials and Procedure Verbal stimuli were the 40 potential expressions of Experiment 1. Conditions regarding sound attenuation, screen size and resolution, and visual angle, were similar to Experiment 1. The procedure was identical to Experiment 1 in all other details.

Design Latency and accuracy were analyzed by means of a repeated measures ANOVA including the factors Time (past vs. future) X Response side (left vs. right), both manipulated within participants.

Results

Errors occurred on 5.23% (142) of the trials, and were excluded from the latency analysis. After inspection of the RT distribution we excluded correct trials with latencies below 335 ms and above 4,000 ms, what amounted to discarding an additional 1.7% (43 trials).

Reaction Time Analysis Centrally for the hypothesis, a significant interaction between Time and Response side emerged ($F(1,33)=6.53$, $p=.02$, $\eta^2=.17$; see Table 2 in the Appendix). Figure 2 illustrates these results. Post-hoc comparisons showed that both the Past-Left vs. Past-Right ($p=.045$) and Future-Left vs. Future-Right ($p<.003$) contrasts reached significance. We also analyzed together the data from potential trials in the two studies including Experiment as a factor. In the overall analysis, the interaction between Time and Response side was also significant ($F(1,60)=13.45$, $p=.001$, $\eta^2=.18$). Moreover, the three-way interaction between Time, Response side and Experiment was not significant ($F<1$). Thus, the space-time congruency effect had the same size in Experiments 1 and 2. There was a main effect of Response side ($F(1,33)=5.06$, $p=.03$, $\eta^2=.13$), but not of Time ($F<1$).

Accuracy Analysis The interaction between Time and Response side was not significant ($F(1,33)=3.15$, $p=.09$, $\eta^2=.09$). Neither Time ($F<1$) nor Response side ($F(1,33)=1.31$, $p=.26$, $\eta^2=.04$) produced significant main effects.

Discussion

A clear space-time congruency effect was observed when potential past and future events were presented without factual events in the experimental context: participants responded faster when past was mapped to the left hand and future to the right hand, than with the opposite mapping. The size of the effect was not different from that observed in Experiment 1. Therefore, present data rule out the possibility that the congruency effect observed for potential events in Experiment 1 was induced by the presence of factual events in the experimental materials. The results of Experiments 1 and 2 provide evidence of a genuine space-time congruency effect for potential events. Our final experiment explored the boundary conditions of this effect. Available studies suggest that the activation of the space-time association is not automatic (Ulrich & Maienborn, 2010; Flumini & Santiago, 2013): it is activated only when temporal discrimination is required by the task. In order to assess whether the activation of the left-right timeline is also automatic for potential events, in Experiment 3 we asked participants to judge the potentiality of the events, instead of their reference to past or future. An additional advantage of this manipulation is that it turns potentiality into a task-relevant dimension, thereby increasing its saliency. Under these conditions we will test whether potentiality is able to generate congruency effects with the dimension of space. If the activation of the link between potentiality and space is also non-automatic, it should now generate a significant congruency effect.

Experiment 3

This experiment had two goals: firstly, to test whether the space-time congruency effect in both factual and potential events can be found when time is a task-irrelevant dimension; and secondly, to test whether there is a space-potentiality congruency effect when potentiality is a task-relevant dimension. We presented the same expressions as in Experiment 1 and asked participants to judge whether each expression referred to a real or potential event.

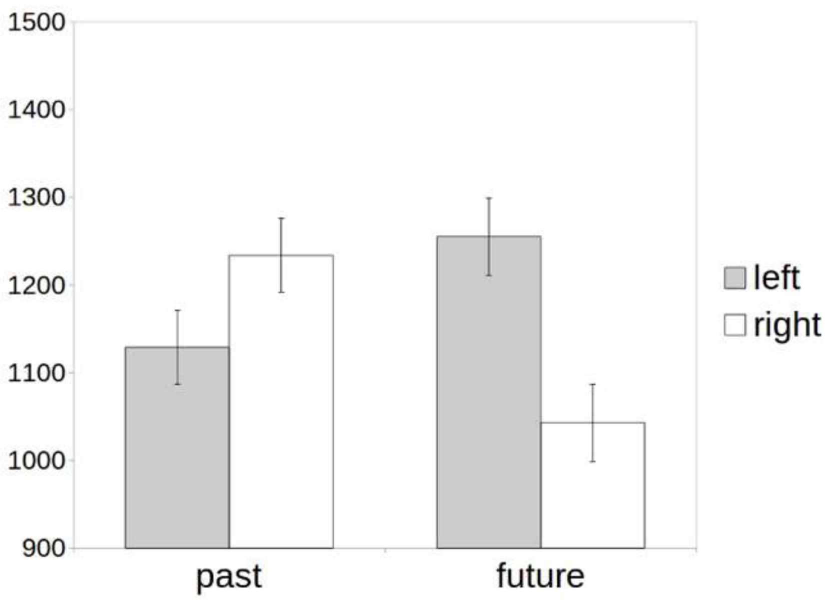


Figure 2. Mean latencies (ms) for potential events in Experiment 2 (error bars show Standard Error of the Mean corrected for within-subject designs, Cousineau, 2005). Participants' task was to judge past versus future reference.

Methods

Participants Thirty new Psychology undergraduate students of the Universidad de la República participated as volunteers (26 mean age, no left-handers, 25 women). They were all native Spanish speakers and volunteered without compensation.

Materials Verbal stimuli were the same as in Experiment 1, with four exceptions: Firstly, the ambiguous items in Experiment 1 ("Nosotros silbamos", "Nosotros dormimos" and "Nosotros soñamos") were fixed by changing their conjugation from first person plural to third person singular, which is not ambiguous. Additionally, the verb "permanecer" ("to remain") was replaced by the verb "sonreír" ("to smile") because by itself "permanecer" does not express a specific event.

Procedure The procedure followed closely Experiment 1, with the following exceptions. At the beginning of the session, we ensured that participants clearly understood what we meant by factual and potential events by means of an example. Additionally, the practice block was extended to 16 trials per condition. This was because, on pilot testing, the potentiality task was shown to be more difficult than the temporality task. In one mapping condition, participants pressed the left key in response to a factual event and the right key in response to a potential event. In the other mapping condition the assignment was reversed.

Design Latency and accuracy were analyzed by means of a repeated measures ANOVA including the same factors as in Experiment 1: Potentiality (factual vs. potential) X Time (past vs. future) X Response side (left vs. right).

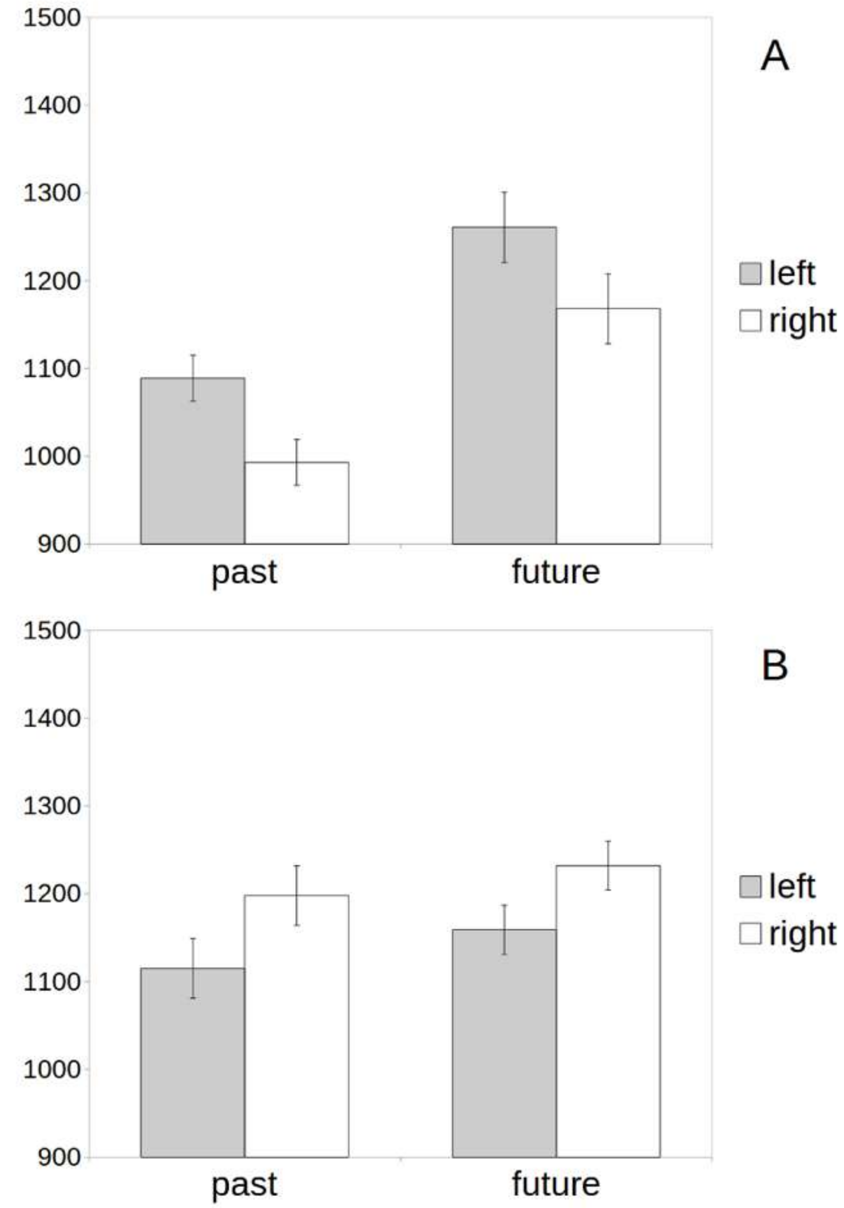


Figure 3. Mean latencies (ms) for factual and potential events in Experiment 3 (error bars show Standard Error of the Mean corrected for within-subject designs, Cousineau, 2005). Participants' task was to judge factual versus potential reference.

Results

Errors occurred on 5.4% (257) of the trials, and were excluded from the latency analysis. After inspection of the RT distribution we excluded correct trials with latencies below 450 ms and above 3,200 ms, what amounted to discarding an additional 1.6% (74 trials).

Reaction Time Analysis Centrally for our concerns, the interaction between Time and Response side disappeared (F<1) and we found an interaction between Potentiality and Response side (F(1,29)=6.99, p=.01, η^2=.19): responses were faster when potential events were mapped onto the left hand and factual events onto the right hand than when using the opposite mapping. Post-hoc comparisons showed that both contrasts, Factual-Left vs. Past-Right (p=.01) and Potential-Left vs. Potential-Right (p<.01), reached significance. Moreover, the three-way interaction between Potentiality, Time, and Response side was also non-significant (F<1; see Table 2 in the Appendix). Figure 3 illustrates these results.

Additionally, the interaction between Potentiality and Time was replicated (F(1,29)=11.98, p=.002, η^2=.30). Potentiality (F(1,29)=4.51, p=.04, η^2=.14) produced a main effect, as in Experiment 1. In contrast to that experiment, the main effect of Time was significant (F(1,29)=18.87, p<.001, η^2=.39) whereas Response side was not (F<1). We analyzed together the data from Experiments 1 and 3 with the aim of comparing the effects of the type of task (time vs. potentiality judgment) on the interactions between Time and Response side, as well as on the newly found interaction between Potentiality and Response side. The overall two-way interaction between Time and Response side was significant (F(1,56)=8.55, p=.005, η^2=.13), and it was modulated by Experiment (F(1,56)=8.21, p=.006, η^2=.13), supporting a change in the space- time congruency effect, from being present in Experiment 1 to being absent in Experiment 3. Additionally, the overall two-way interaction between Potentiality and Response side reached significance (F(1,56)=7.94, p=.007, η^2=.12), and its interaction with Experiment fell short of it (F(1,56)=3.52, p=.07, η^2=.06). The independent analyses of Experiment 1 and 3 reported above support as the most likely conclusion that it was absent in Experiment 1 and present in Experiment 3. Therefore, the task-relevant conceptual dimension in each experiment (time vs. potentiality) interacted with side of response. This analysis also uncovered some additional effects, which are not central to the concern of this research, as they do not involve interactions with the Response side factor. The overall main effect of Time (F(1,56)=7.86, p=.007, η^2=.12) was also affected by the change in task (F(1,56)=14.75, p<.001, η^2=.21), in a way that does not lend itself to a straightforward interpretation: it changed from a 25 ms advantage of future over past events in Experiment 1 to a 107 ms advantage of past over future events in Experiment 3. Furthermore, the overall interaction between Potentiality and Time was not significant (F<1), because it took opposite shapes in the two experiments (F(1,56)=18.28, p<.001, η^2=.25): when participants judged time, future factual events were faster than potential events, but past factual and potential events did not differ; when participants judged potentiality, future factual and potential events did not differ, and past real events were faster than past potential events. Finally, some effects were not affected by Experiment. The overall main effects of Potentiality ($F(1,56)=11.59$, $p=.001$, $\eta^2=.17$) and Response side ($F(1,56)=4.19$, $p=.05$, $\eta^2=.07$) were significant and remained constant across experiments (Potentiality X Experiment: $F(1,56)=1.44$, $p=.24$, $\eta^2=.03$; Response side X Experiment: $F(1,56)=1.98$, $p=.17$, $\eta^2=.03$). The overall three-way interaction between Time, Potentiality, and Response direction was not significant, nor did it interact with Experiment (both Fs<1).

Accuracy Analysis. The interactions between Time and Response ($F<1$), Time, Potentiality, and Response ($F(1,29)=2.52$, $p=.12$, $\eta^2=.08$), and Potentiality and Response side were not significant ($F(1,29)=2.65$, $p=.11$, $\eta^2=.08$). There was an interaction between Time and Potentiality ($F(1,29)=4.91$, $p=.04$, $\eta^2=.15$) and a main effect of Time ($F(1,29)=40.09$, $p<.001$, $\eta^2=.58$). The main effect of Response side also failed to be significant ($F(1,29)=3.06$, $p=.09$, $\eta^2=.10$).

Discussion

In a task using a potentiality judgment, the latency measure did not detect any space-timecongruency effect, neither for factual nor potential events. This result supports the non-automaticity of the activation of the lateralized mental timeline, as suggested by Ulrich and Maienborn (2010). Instead, there was a space-potentiality interaction: participants responded faster when potential events were mapped to the left hand, and factual events to the right hand, than when using the opposite mapping. This interaction was suggested by the accuracy analysis of Experiment 1 and it was confirmed by the overall ANOVA of latency data from Experiments 1 and 3. We discuss this finding in the following section. Experiment 3 also revealed some intriguing additional findings of the task demands: the change from a time judgment task to a potentiality judgment task reversed the main effect of temporal reference, and affected the shape of the interaction between temporal reference and potentiality. As these effects do not involve interaction with response side they are irrelevant to test the spatial representation of abstract concepts. They could be due to conceptual connections between time and potentiality which are independent from their spatial aspects and somehow dependent on the task relevance of the conceptual dimensions. Unfortunately, what could be these connections and how they give rise to the observed effects is at present unclear and will need to be explored in more detail in future research.

General Discussion

Do potential events activate the mental timeline? The present study provided an initial answer to this question: Yes, speakers can map time onto space when processing potential events. Experiment 1 showed that the space-time congruency effect for potential events was indistinguishable from the effect observed for real events. Experiment 2 showed that the effect is genuine and arises even when the experimental materials include only potential events. Finally, Experiment 3 showed that when time becomes task-irrelevant and potentiality becomes task-relevant, the space-time congruency effect vanishes away and is replaced by a space-potentiality effect, such that processing is facilitated when potential events are mapped onto the left hand and factual events onto the right hand (as compared to the opposite mapping). Therefore, both space-time and space-potentiality mappings seem non-automatic and can be modulated by task demands. This finding is relevant to the debate about the automaticity of congruency effects. The best studied congruency effect related to an abstract concept is the Space Number Association of Response Codes (SNARC), first reported by Dehaene, Bossini, and Giraux (1993). In the SNARC effect small numbers are responded to faster with the left hand and larger numbers with the right hand. It arises in tasks where accessing magnitude information is not relevant (such as deciding whether a number is odd or even). Moreover, the mere perception of a number induces shifts of visual attention toward the left or right depending on its magnitude (Fischer, Castel, Dodd & Pratt, 2003). However, this finding has been difficult to replicate (Fattorini, Pinto, Rotondaro & Doricchi, 2015) and the implicitness of the odd-even task is questionable. In this context, present data suggest that other abstract dimensions such as time and potentiality are activated in ways that are clearly non-automatic and depend on their relevance to the task. Present data thus suggest that a mental timeline is activated both when people represent sequences of factual and potential events, whenever the task does require a consideration of the order of those events. If this interpretation is correct, current findings add to the literature that shows that language comprehension implies the creation of mental simulations of linguistic content (Barsalou, 2003) even when the described events are negated or counterfactual (de Vega et al, 2014). Present findings suggest that these mental simulations not only contain analogical dimensions linked to the internal characteristics of the simulated events, but also represent in analogical fashion aspects which are external to the events, such as their order or potentiality. However, while the space-time congruency effect can be interpreted straightforwardly in the current research context, the newly attested space-potentiality effect poses interesting challenges.

What could be the causes of the space-potentiality effect? One possibility relies on the inherently predictive character of future events. For instance, speakers of Aymara refer to the future using the word for "back", and to the past using the word for "front" (Núñez & Sweetser, 2006). These authors suggested that the motivation for this conceptual mapping is the fact that the past can be "seen" clearly, as it has already happened, but the future cannot. Under this account, the potentiality of the future would support mapping both future and potential onto right space in Spanish speakers. However, present data actually show the opposite mapping (potential-left, factual-right), and therefore rule out this account. An alternative account is based on the polarity correspondence hypothesis proposed by Proctor and Cho (2006). If both potentiality and lateral space can be processed as polar dimensions, with a marked and an unmarked (default) pole, the polarity correspondence hypothesis would predict that processing should be facilitated when the poles of the same sign are mapped onto each other. It seems intuitively correct to assume that the unmarked pole of the dimension of potentiality is the factual pole, and that the unmarked pole of the dimension of lateral space is the right side (at least for right-handers). Therefore, mapping factual on the right and potential on the left would facilitate processing as compared to the reversed mapping.

This view can account for the observed space-potentiality congruency effect and at present we believe it is the best available explanation of it. However, it opens other challenging questions. Recently, Santiago and Lakens (2015) have shown that polarity correspondence cannot explain the mapping of time (nor numbers) onto lateral space. What are, then, the factors that make some conceptual dimensions, such as the potentiality dimension, able to generate a polarity correspondence effect and that distinguish it from the dimension of time, which is not? At present we believe that the answer may be related to processing characteristics linked to the horizontal left-right axis versus the vertical axis. Lakens (2012) showed that the congruency effect between social power and vertical space is affected by factors predicted by polarity correspondence. Social power is conceptualized as a vertical dimension ("power is up", see Lakoff & Johnson, 1980), as shown by expressions such as "the commands came from above". In a recent study using the same rationale as Santiago and Lakens (2015), Chang and Cho (2015) showed that, in contrast to time and number, loudness does behave as predicted by the polarity correspondence hypothesis. Loudness is also conceptualized as a vertical dimension through the "more is up" metaphor (Lakoff & Johnson, 1980), linguistically manifested in expressions such as "turn the volume up" or "turn it down". Linguistic expressions suggest that potentiality is also a vertical dimension, as something factual and certain is "tied down" whereas something potential is "still in the air". Time and numbers, instead, map onto the left-right axis (Bonato, Zorzi, & Umiltà, 2012). This pattern suggests that the relevant spatial axis is the main factor that distinguishes conceptual dimensions that generate polarity correspondence effects from those that do not. Some studies aimed to test this possibility are currently under course in our lab. If space-time and space-potentiality congruency effects finally deserve a single theoretical interpretation or they require different mechanisms is something that remains to be seen.

To conclude, the present study has shown that potential past and future events can activate the lateralized mental timeline to the same extent as factual events do. In doing so, it has also revealed an interesting new phenomenon: the mental representation of potentiality can also establish links to the lateral spatial dimension, at least under conditions in which potentiality is task relevant. Future research will address the exact nature of this relation.

Acknowledgements

This research was supported by grant PD_MAC_2012_1_8004 from the Uruguayan National Research and Innovation Agency to RA, and by grant PSI2012-32464 from the Spanish Ministry of Economy and Competitivity to JS. Portions of this research were presented at the Annual Conference of the Cognitive Science Society 2015. We thank José Placido Ruiz Campillo for his suggestions on the expression of tense and mood in Spanish verbs.

References

Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513-562.

Baumeister, R.F., & Masicampo, E.J. (2010). Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal-culture interface. *Psychological Review*, *117*, 945-71. https://doi:10.1037/a0019393

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews*, *36*, 2257-2273. https://doi:10.1016/j.neubiorev.2012.08.007

Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, *143*, 473-479. https://doi: 10.1037/a0033297

Chang, S., & Cho, Y.S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology*, *6*. https://doi:10.3389/fpsyg.2015.00683

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*, 42-45.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371. https://doi.org/10.1037/00963445.122.3.371

De Vega, M., & Urrutia, M. (2012). Discourse updating after reading a counterfactual event. *Psicologica: International Journal of Methodology and Experimental Psychology*, *33*(2), 157-173

De Vega, M., De León, I., Hernández, J. A., Valdés, M., Padrón, I., & Ferstl, E. C. (2014). Action sentences activate sensory motor regions in the brain independently of their status of reality. *Journal of Cognitive Neuroscience*, *26*, 1363-1376. https://doi:10.1162/jocn

Fattorini, E., Pinto, M., Rotondaro, F., & Doricchi, F. (2015). Perceiving numbers does not cause automatic shifts of spatial attention. *Cortex*, *73*, 298-316. https://doi.org.proxy.timbo.org.uy:443/10.1016/j.cortex.2015.09.007

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. https://doi:10.3758/BF03193146

Ferguson, H. J., Fresh, M., & Leblond, J. (2013). Examining mental simulations of uncertain events. *Psychonomic Bulletin & Review*, *20*, 391-9. https://doi:10.3758/s13423-012-0348-1

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, *6*, 555-556.

Flumini, A., & Santiago, J. (2013). *Time (also) flies from left to right... if it is needed!* In M. Knauff, M. Pauen, N. Sebanz, & I. Wachmuz (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. *2315-2320*). Austin, TX: Cognitive Science Society.

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*, 1430-1451. https://doi:10.1111/j.1551-6709.2010.01105.x

Gilead, M., Liberman, N., & Maril, A. (2012). Constructing counterreal worlds: The role of abstraction. *European Journal of Social Psychology*, *42*, 391-397. https://doi:10.1002/ejsp.1862

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280-5. https://doi:10.1016/j.tics.2004.04.001

Johnson-Laird, P.N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*, 71-115.

Kaup, B., Yaxley, R.H., Madden, C.J., Zwaan, R.A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology*, *60*, 976-90. https://doi:10.1080/17470210600823512

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 726-736. https://doi:10.1037/a0024955

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Núñez, R.E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*, 401-450. https://doi:10.1207/s15516709cog0000_62

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36-45. https://doi:10.1016/j.jml.2014.04.001

Ouellet, M., Santiago, J., Israeli, Z., & Gabay, S. (2010). Is the future the right time? *Experimental Psychology*, *57*, 308-14. https://doi:10.1027/1618-3169/a000036

Piaget, J. & Inhelder, B. (1985). *De la lógica del niño a la lógica del adolescente*. Paidós Ibérica: Barcelona.

Proctor, R.W., & Cho, Y.S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*, 416-42. https://doi:10.1037/00332909.132.3.416

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510-532. https://doi:0.1037/0033-2909.114.3.510

Ruiz Campillo, J.P. (2014). La lógica del espacio. Un mapa operativo del sistema verbal del español. *Journal of Spanish Language Teaching*, *1*:1, 62-85. https://doi:10.1080/23247797.2014.898518

Santiago, J., & Lakens, D. (2015). Can conceptual congruency effects between number, time, and space be accounted for by polarity correspondence? *Acta Psychologica*, *156*, 179-191. https://doi:10.1016/j.actpsy.2014.09.016

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M.J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*, 512-6. Buenos Aires. https://doi: 10.3758/bf03194099

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research*, *74*, 59-70. https://doi:10.1007/s00426-008-0220-0

Schlemenson, S. (2004). *Subjetividad y lenguaje en la clinica psicopedagógica. Voces presentes y pasadas*. Paidós: Buenos Aires.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.

Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, *30*, 745-757. https://doi:10.1207/s15516709cog0000_67

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*, 515-557. https://doi: 0010-0285/9

Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*, 126-138. https://doi:10.1016/j.cognition.2010.08.001

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, *15*, 426-30. https://doi: 10.3758/pbr.15.2.426

Appendix

Table 1.- Verbal Stimuli used in Experiment 1. English translations are shown within brackets. Note that the English translations often provide less rich information about subject gender and number, and about verb tense and mood than the Spanish original (see discussion in the main text). Note also that the items "Ellos iniciaron" ("they started"), "Nosotros silbamos" ("We whistled"), "Nosotros soñamos" ("We dreamt") and all items with "permanecer" (to remain) were used in Experiment 1, but substituted in Experiment 3 by the third person singular ("Él inició" , "Él silbó", "Él soñó"), and by the verb "sonreír" (to smile: FP: He smiled; FF: He will smile; PP: He had smiled; PF: He would.

Verbs	Factual Past (FP)	Factual Future (FF)	Potential Past (PP)	Potential Future (PF)
Actuar	Ellas actuaron	Ellas actuarán	Él hubiera actuado	Ellas actuarían
Apartarse	Tú te apartaste	Tú te apartarás	Te hubieras apartado	Tú te apartarías
Callarse	Usted se calló	Él se callará	Se hubiera callado	Él se callaría
Cambiar	Ellos cambiaron	Tú cambiarás	Tú hubieras cambiado	Tú cambiarías
Despertar	Ella despertó	Ella despertará	Hubiera despertado	Ella despertaría
Dormir	Nosotros dormimos	Nosotros dormiremos	Se hubiera dormido	Ella se dormiría
Escaparse	Usted se escapó	Usted escapará	Se hubiera escapado	Usted escaparía
Iniciar	Ellos iniciaron	Ellos iniciarán	Él hubiera iniciado	Ellos iniciarían
Llegar	Ellas llegaron	Usted llegará	Tú hubieras llegado	Usted llegaría
Marchar	Ellos marcharon	Ellos marcharán	Yo hubiera marchado	Ellos marcharían
Morir	Ellas murieron	Ellas morirán	Ella hubiera muerto	Ellas morirían
Nadar	Ellas nadaron	Ellas nadarán	Ella hubiera nadado	Ellas nadarían
Ocultarse	Usted se ocultó	Yo me ocultaré	Me hubiera ocultado	Yo me ocultaría
Permanecer	Él permaneció	Yo permaneceré	Hubiera permanecido	Yo permanecería
Retirarse	Tú te retiraste	Tú te retirarás	Me hubiera retirado	Tú te retirarías
Silbar	Nosotros silbamos	Ellos silbarán	Yo hubiera silbado	Ellos silbarían
Soñar	Nosotros soñamos	Nosotros soñaremos	Él hubiera soñado	Él soñaría
Sudar	Ellas sudaron,	Ellas sudarán	Él hubiera sudado	Ellas sudarían
Trabajar	Usted trabajó	Usted trabajará	Él hubiera trabajado	Usted trabajaría
Volar	Ellos volaron	Ellos volarán	Yo hubiera volado	Ellos volarían

Table 2.- Mean latencies in milliseconds and proportion of errors per condition (within brackets). All Time-Response side pairwise contrasts of latencies reached significance in post-hoc analyses of Experiments 1 and 2. All Potentiality-Response side pairwise contrasts reached significance in post-hoc analyses of Experiment 3

Potentiality-Time conditions	Experiment 1: Temporal task		Experiment 2: Temporal task		Experiment 3: Potentiality task	
	Left	Right	Left	Right	Left	Right
Factual Past	1,369	1,443			1,089	993
	(.06)	(.07)			(.02)	(.01)
Factual Future	1,411	1,210			1,261	1,168
	(.04)	(.02)			(.11)	(.06)
Potential Past	1,366	1,491	1,129	1,234	1,115	1,198
	(.04)	(.07)	(.02)	(.08)	(.05)	(.05)
Potential Future	1,582	1,399	1,255	1,043	1,159	1,232
	(.07)	(.11)	(.07)	(.04)	(.06)	(.04)

Taza decorativa de la Monalisa
Monalisa's decorative mug

Ésta fue la **segunda taza** que se obsequió a los voluntarios por su participación en los experimentos e investigaciones del CIBPsi. Entre 2016 y 2018, los investigadores de las líneas y proyectos del centro obsequiaban esta taza a sus voluntarios.

This was the **second decorative mug** gifted to volunteers for their participation in **CIBPsi** experiments and research. Between 2016 and 2018, the researchers of the center's lines and projects gave this mug as a gift to their volunteers.

¿Cómo explorar las diferencias y similitudes en el léxico mental de poblaciones oyentes y sordas, considerando la modalidad como el factor de discriminación?

Roberto Aguirre[1], Jorge Vivas[2], Matías Yerro[2], María Noel Macedo[1], Jordi Villoro Armengol[3] & Alejandro Fojo[1,4]

1. **Centro de Investigación Básica en Psicología (CIBPsi)**. Montevideo, Uruguay
2. Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT CONICET-UNMDP), Mar del Plata, Argentina
3. ESIC Business & Marketing School Barcelona. Barcelona, España
4. Carrera de Tecnólogo en Traducción e Interpretación LSU-Español. Montevideo, Uruguay

Roberto Aguirre https://orcid.org/0000-0001-8386-1380, Jorge Vivas https://orcid.org/0000-0001-8648-2562, Matías Yerro https://orcid.org/0000-0001-9572-6565, María Noel Macedo https://orcid.org/0000-0002-3957-1340, Jordi Villoro Armengolhttps://orcid.org/0000-0003-2145-5795

Correspondence concerning this manuscript should be addressed to Roberto Aguirre, Facultad de Psicología, Tristán Narvaja 1674, Montevideo 11200, Uruguay. Email: **zimmer20uy@gmail.com**

This is a free Spanish version -with the approval of the authors- of the pre-print of the manuscript Aguirre, R. Macedo, M.N. & Villoro Armengol, J. / Vivas, J. (2023, aceptado). *Interdisciplinaria Revista de Psicología y Ciencias Afines, (40)*, 2, 117-131. https://doi.org/10.16888/inter.2023.40.2.7

Abstract

The efforts to understand the relationship of the meaning of a lexical pieces to the object to which it refers, the relationship between the lexical pieces and the concepts, and the understanding of the meanings shared between two speakers who use the same lexical pieces, have constituted a major problem for the semantic memory models. According to contemporary literature, perceptual-motor, linguistic, and social information has different weights in the formation of concepts, whether concrete or abstract, stored in the aforementioned memory. Regardless of the models developed so far, it is interesting to note that semantic knowledge is represented by various ways of relating the concepts and the types of relationships between them. In this context, studies in sign languages and comparative studies between spoken and sign languages are scarce in the world. Thus, little is known about the effect of linguistic modality on the semantic networks. After all, the theory on semantic networks and norms for the production of features has been grounded on theories of language and its processing adjusted to spoken languages. As the incorporation of the sign languages and the deaf populations has shown in other psycholinguistic and linguistic topics, the importance of including these languages and populations, and comparatives with spoken languages, might increase the explanatory power of the theory to account for the universal and contextual aspects of language and its processing. In this effort, there is a latent risk: the linguistic modality can be only a vehicle for more well-known or studied cross-modal variables (e.g., age of acquisition and first exposition, functional distribution of language, size of the available lexicon, etc.). If it is considered that languages are not stored together, but similar processes can occur in them, it is essential to find out what may be a singular feature of each modality (spoken vs. sign) that might ground differentiated processes. Considering the high iconicity of the sign languages, the possibility of a high concreteness of the lexical pieces in them, and the differences in the material subtract (sound-internal body vs. space-externe body) of these language modalities as distinctive features, the article suggests repeated free association tasks as an appropriated approach to avoid the aforementioned risk in the study of the effects of the linguistic modality (sign vs oral) in the semantic memory. This type of task does it possible to collect meanings related to linguistic information and non-linguistic experience because affective and experiential information is accessible by doing the task in different repetitions. The approach and the tool are exemplified by a comparative study between deaf signing and hearing populations. The partial findings of this study also serve to focus on the expected effects of the difference in iconicity and the level of concreteness / abstractness of the lexical pieces of each linguistic modality.

Keywords: Conceptual domains, Semantic memory, Mental lexicon, Spoken languages, Sign languages

Introducción

¿De qué manera la modalidad lingüística (lengua oral vs. lengua de señas) podría modular el perfil de conceptos concretos como el espacio y abstractos como el tiempo? En las primeras investigaciones, algunos autores sugirieron un mayor pensamiento concreto en los niños sordos que en los oyentes (Myklebust y Brutton, 1953). Otros argumentaron similitudes entre ellos (Fürth, 1971). Para Zwiebel y Merten (1985), la percepción tiene un mayor peso cognitivo en la conceptualización de las personas sordas en comparación con la población oyente y la memoria tiene un mayor peso en el pensamiento abstracto de las personas sordas.

Para dar cuenta de la agrupación de conceptos en conjuntos más o menos flexibles, anclados en la estructura de la experiencia para proporcionar significado, Langacker (1987) sugirió la noción de dominio. Tal y como lo describen Radden y Dirven (2007), el dominio conceptual es el campo general al que pertenece una categoría o marco en una situación determinada. Por ejemplo, una sartén pertenece al dominio "cocina" cuando se utiliza para freír un huevo, pero al dominio "lucha" cuando se utiliza como arma en una disputa doméstica.

Memoria semántica

Como sintetizan McRae y Jones (2013), la memoria semántica tiene que ver con nuestro conocimiento sobre el mundo y en ella se inscriben nuestras representaciones internas sobre las cosas, sus etiquetas verbales y sus propiedades. El estudio de la semántica tiene una gran tradición con algunos enfoques complementarios a la hora de describir el significado del lenguaje (Andrews, et al. 2014). Estos enfoques pueden clasificarse (Meteyard et al., 2012) según el contraste entre propuestas simbólicas y de simulación. Otros autores (Deyne et al., 2017) los clasificaron según sus fundamentos disciplinares (Psicología, Lingüística e Informática). ¿Qué es relevante en estas clasificaciones? El primer enfoque se organiza por el creciente papel semántico de la información sensoriomotora en las perspectivas más simbólicas (por ejemplo, Quillian, 1968; Collins y Loftus, 1975; Levelt, 1989; Landauer y Dumais, 1997) hacia las más simuladoras (por ejemplo, Glenberg y Kaschak, 2003; Zwaan, 2004).

Deyne et al. (2017) argumentaron que la asociación de palabras es una herramienta única y útil porque es una expresión del pensamiento. En estas tareas, se eliminan las exigencias de la sintaxis, la morfología y la pragmática. Entonces, los autores pretenden construir redes semánticas a gran escala a partir de tareas de asociación de palabras. De este modo, coinciden con Collins y Loftus (1975) en que una representación en red del léxico mental proporciona una descripción psicológicamente plausible. A continuación, la estrategia sugerida por Deyne et al. (2017) consiste en utilizar la red empírica de asociación de palabras como herramienta para acercarse a una red semántica latente.

Un reto importante en la formulación de modelos de memoria semántica es comprender la relación del significado de una pieza léxica con el objeto, la entidad o la experiencia a la que se refiere, la relación entre esa pieza y los conceptos, y la comprensión de los significados compartidos entre dos hablantes que utilizan las mismas piezas léxicas. Una forma de entender estos aspectos es considerar la proximidad entre piezas léxicas como expresión de la proximidad de los conceptos y los dominios a los que pertenecen. Así, la evocación de un concepto implica la propagación del pensamiento hacia otro concepto al que está vinculado lógica (relaciones de pertenencia, coordinación, etc.) o analógicamente (metafórico, metonímico, introspección, etc.). El léxico mental ofrece una ventana a la configuración de los dominios conceptuales y a la forma de la memoria, al menos parcialmente. En un léxico mental determinado, los distintos dominios pueden tener diversos grados de relación o estar configurados de forma diferente, debido a mecanismos de proyección conceptual como las analogías y las metáforas.

La organización de los conceptos

Borghi et al. (2017) coinciden en que todos los conceptos son flexibles, es decir, dependen del contexto. Sin embargo, los conceptos concretos son más estables a lo largo del tiempo y están más moldeados por las experiencias cotidianas, las situaciones y la cultura (Barsalou, 1987). Los conceptos concretos, como SILLA, tienen límites físicos con sentidos que cualquiera puede identificar. Estas características priman disposiciones más amplias y prolongadas en el procesamiento de los conceptos y sus características. Por el contrario, los conceptos abstractos están más alejados de referentes físicamente identificables. Si bien tanto los conceptos concretos como los abstractos se asocian con elementos del contexto, estos últimos parecen implicar más eventos y acciones, y por lo tanto evocan información contextual más compleja que los conceptos concretos (Borghi y Binkofski, 2014; Connell y Lynott, 2012). Por último, Borghi et al. (2017) sugieren que la información perceptiva y de acción es más relevante para los conceptos concretos, mientras que la información emocional y lingüística se expresa sobre todo con conceptos abstractos. Los matices de concreción/abstracción pueden explicarse en un enfoque de gradiente.

Desde la perspectiva de la concretud/abstracción de los conceptos y los dominios en los que se inscriben, las identificadas como teorías integradoras (*Representational Pluralism*, Dove, 2009; *Language and Situated Simulation*, Barsalou, 2008; *Words as Social Tools*, Borghi y Cimatti, 2009) consideran la información sensoriomotora, lingüística (en el sentido de codificación) y social (uso del lenguaje y entorno cultural y social extralingüístico) como relevante para el procesamiento y caracterización de conceptos tanto concretos como abstractos con diferentes mecanismos (Borghi et al., 2017).

El conocimiento semántico se representa mediante diversas formas de relacionar conceptos. Las relaciones taxonómicas son aquellas en las que los conceptos se organizan jerárquicamente desde el nivel menos inclusivo al más inclusivo o viceversa (por ejemplo, “mamíferos” es una categoría superior que incluye perros, lobos, vacas y gatos o “espacio” es superior a altura y anchura). Los conceptos que pertenecen a un dominio organizado taxonómicamente están relacionados entre sí en virtud de las características que comparten (por ejemplo, los perros, los lobos, las vacas y los caballos son mamíferos por la presencia de glándulas mamarias que en las hembras producen leche para alimentar a sus crías). Sin embargo, las relaciones taxonómicas no parecen exclusivas de las características físicas concretas de las entidades, como las referidas en los ejemplos anteriores. Por ejemplo, “Causa” y “Efecto” son antónimos pero no indican características concretas como las del color o el tamaño de ningún objeto. La organización taxonómica permite una forma rentable de procesar la información, facilitando su recuperación para futuras aplicaciones, posibilitando las analogías y la resolución de problemas, y promoviendo el desarrollo de nuevos conocimientos.

En cambio, la organización temática de un dominio permite relacionar un concepto con otro en función de su co-ocurrencia en un evento o situación. Esta organización incluye asociaciones espaciales y temporales entre agentes, objetos y pacientes de una acción (Borghi y Caramelli, 2003; Estes et al., 2011). Además, permiten la organización contextual de la experiencia, así como el establecimiento de predicciones para situaciones futuras similares. Ambos procesos —comparación e integración— y organizaciones —temática y taxonómica— están disponibles para las mentes humanas (Borghi y Carammelli, 2003; Siegler y Shipley, 1995; García Coni et al., 2019). Ambos tipos de organización corresponden a mecanismos cognitivos diferentes: las relaciones taxonómicas activan predominantemente un proceso de comparación entre objetos, mientras que las relaciones temáticas activan un proceso de integración.

La organización jerárquica y cada vez más inclusiva del conocimiento taxonómico permite la generación de piezas léxicas que se alejan gradualmente de referentes físicamente identificables. Esa organización permite incluir más acontecimientos y acciones en cada pieza léxica y aumentar la relevancia de la información emocional y lingüística para identificar e interpretar dichas piezas léxicas. Por el contrario, los conceptos concretos están cerca de referentes físicamente identificables, incluyen aspectos de eventos/acciones específicos y aumentan la relevancia de la información sensoriomotora. Aquí sugerimos tomar la organización taxonómica de la semántica léxica como índice de una interpretación abstracta y la organización temática como índice de una interpretación concreta de dominios de experiencia.

El estudio del léxico mental

En una tarea de asociación de palabras, las respuestas sintagmáticas son palabras que siguen al estímulo en una secuencia sintáctica (por ejemplo, FRÍO-fuera) o palabras que comparten una relación temática con el estímulo (por ejemplo, FRÍO-suéter, FRÍO-invierno). En cambio, las respuestas paradigmáticas son palabras de la misma clase de palabras (o paradigma) que el estímulo (por ejemplo, FRÍO-calor; Sheng et al., 2006). Las relaciones sintagmáticas pueden derivarse de experiencias perceptivas y conceptuales tangibles. Las respuestas paradigmáticas representan relaciones lingüísticas más abstractas y taxonómicas. En el modo de asociación repetida, el acceso a las relaciones semánticas, en particular las paradigmáticas, se vuelve progresivamente más difícil a medida que la activación semántica atraviesa la red (Elbers y van Loon-Vervoorn, 1998; Sheng y McGregor, 2010). Además, al hacer la tarea en distintas oportunidades, este tipo de tarea permite acercarse a significados relativos a la información lingüística y a la experiencia no lingüística porque la información afectiva y experiencial es accesible al hacer la tarea (De Deyne et al., 2021).

Los estudios comparativos del léxico mental entre la lengua oral y la lengua de señas son escasos. Marschark et al. (2004) estudiaron la organización y el uso del léxico mental en hablantes sordos y oyentes utilizando una tarea de asociación libre de respuesta única. Sus resultados indicaron una similitud general en la organización del conocimiento léxico en ambos grupos, con asociaciones más fuertes entre nombres de categorías y ejemplares para los hablantes oyentes. Mann et al. (2016) realizaron una tarea de asociación libre repetida con niños sordos de la American Sign Language (ASL) y niños ingleses oyentes. Sus resultados mostraron patrones similares en las respuestas de sujetos sordos y sujetos oyentes monolingües en inglés. Para los autores, estos resultados sugieren que el desarrollo del lenguaje en las lenguas de señas y orales se rige por mecanismos de aprendizaje similares, basados en el desarrollo de redes semánticas. Nótese que los niños del estudio de Mann et al. (2016) eran bilingües y esta condición podría influir en sus resultados. Dong et al. (2005) indican que los bilingües de dos lenguas orales tienden a integrar las diferencias conceptuales entre equivalentes de traducción. Sin embargo, los autores encuentran que los bilingües muestran una tendencia a mantener el sistema conceptual de la L1 para las palabras de la L1 y a adoptar el sistema conceptual de la L2 con las palabras de la L2. Además, los autores descubrieron que, en el aprendizaje muy temprano de una palabra de L2, los aprendices de L2 son más dependientes de su L1: aprender una L2 es un proceso de integración de diferencias conceptuales en dos lenguas. Entonces, si una población específica de sordos señantes permanece en una fase temprana de aprendizaje de una lengua oral como el español, cabe esperar que la integración conceptual antes mencionada esté lejos de producirse.

Diferencias y similitudes entre modalidades del lenguaje

Para caracterizar la noción de modalidad lingüística (limitada en nuestro caso al contraste señas vs. oral) en sus diferentes aspectos, a continuación discutiremos algunas diferencias y similitudes semióticas, lingüísticas, psicolingüísticas y sociolingüísticas entre las lenguas de señas y las lenguas orales que pueden discriminar entre el léxico mental de señantes y oyentes. No es nuestro objetivo ser exhaustivos en cada aspecto, sino sólo destacar en cada uno de ellos algunos puntos que entendemos de interés para el léxico mental y su procesamiento.

Aspectos semánticos Dentro de los aspectos semióticos, aunque existe un número sustancial de elementos léxicos arbitrarios en las lenguas de señas, la literatura (Klima y Bellugi, 1979; Taub, 2001; Wilcox, 2004; Cuxac y Sallandre, 2007. Para una revisión, Perniss et al., 2010) considera una mayor omnipresencia de elementos léxicos icónicos en las lenguas de señas que en las orales. Wilcox (2004, p. 123) describe la iconicidad como una relación de semejanza entre dos interpretaciones: una de las escenas del mundo real y otra de la forma de los objetos de referencia. La literatura reconoce algunos tipos y grados de iconicidad en las lenguas de señas (es decir, pantomímica y perceptual-motora) con enfoques alternativos (por ejemplo, Ortega et al., 2017; Gentner y Ratterman, 1991). La mencionada semejanza entre una forma (i.e., pieza léxica) y su significado es producto de los procesos cognitivos del individuo que realiza la comparación. La iconicidad podría considerarse un caso de mapeo de estructuras (Gentner, 1983; Gentner y Markman, 1997) porque los ítems icónicos implican un proceso de comparación entre una representación semántica (por ejemplo, el significado pasado) y una representación visual (por ejemplo, mover la mano sobre el hombro de adelante hacia atrás) o acústica (por ejemplo, usar la palabra atrás) de una forma lingüística (Emmorey, 2014).

A diferencia de las lenguas orales, la iconicidad subléxica de las lenguas de señas tiene su potencial en que el polo fonológico de las señas está formado por componentes del cuerpo humano espacialmente expuestos a los interlocutores (manos, torso, cara y cabeza). Esta iconicidad es relativa a los referentes que implican la motricidad asociada a acciones, emociones o la percepción de objetos concretos, y su eventual uso como base para dotar de significado a piezas léxicas mediante algunos mecanismos de mapeo como las metáforas o las metonimias. Además, aunque la iconicidad no está ausente en el lenguaje oral (Dingemanse et al., 2015; Perniss et al., 2010), se basa principalmente en componentes internos del cuerpo humano, no expuestos a los interlocutores (es decir, laringe, lengua, alvéolos, etc.). Además, la imaginabilidad, es decir, la facilidad para imaginar el significado de una palabra (en parte debido a su iconicidad), es una propiedad de las palabras/ señas que se correlaciona con la abstracción (Kousta et al., 2011).

Como sugieren Rodríguez-Cuadrado et al. (aceptado), la iconicidad se basa en la concreción. Los referentes de conceptos concretos son accesibles como información sensoriomotora (es decir, vistos, oídos) y es posible interactuar con ellos (véase Borghi et al. (2017). Los hablantes pueden utilizar la apariencia física de estos referentes para guiar la formación de ítems icónicos. Compartimos el enfoque teórico (Lakoff y Johnson, 1980) que sostiene que la mente humana se fundamenta en conceptos concretos e incluye mecanismos regulares como la metáfora conceptual y la metonimia. Para este enfoque, los ítems de conceptos abstractos pueden convertirse en icónicos por la apariencia física de esos referentes. En lo que respecta al vocabulario de las lenguas de señas, este enfoque permite distinguir entre señas icónicas y arbitrarias para conceptos abstractos (Taub, 2001; Wilcox, 2000).

Una diferencia fundamental es que las señas representan mejor icónicamente los conceptos que denotan y tienen su materia simbólica en el espacio (Taub, 2001). Esto se debe al elevado potencial de acceso semántico basado en el sustrato espacial de los señas y a la analogía de éstas con la información sensoriomotora relevante. A pesar de la mencionada iconicidad omnipresente en las lenguas de señas, Hohenberger y Leuninger (2012, p. 718) señalan que no hay diferencia en el procesamiento de señas icónicas y no icónicas entre las lenguas orales y de seña. Por ejemplo, la iconicidad no desempeña ningún papel en la recuperación léxica. Sin embargo, Ortega y Morgan (2015), estudiando la sensibilización de las señas icónicas en el léxico mental de adultos oyentes, sugirieron que los oyentes no señantes procesan las señas icónicas a la manera de los gestos. Los señantes, en cambio, utilizan un mecanismo diferente para procesar las señas icónicas. Autores como Sehyr y Emmorey (2019) descubrieron que la percepción del mapeo entre forma y significado en ASL depende del conocimiento lingüístico y de la tarea.

Aspectos lingüísticos Dentro de los aspectos lingüísticos, Johnston y Schembri (1999, p. 126) distinguen entre señas y lexemas. Para estos autores, estos últimos son piezas totalmente convencionalizadas y pueden almacenarse como léxico de una lengua. La diferencia entre una seña lexicalizada y una con clasificador es que esta última responde al dominio discursivo. Es decir, por ejemplo, un verbo con clasificador como "caminar" no está lexicalizada en la Lengua de Señas Uruguaya (LSU). Esto último permite al señante modificar esta seña hasta el punto de adaptar la predicación del verbo CAMINAR al interés discursivo. El señante puede indicar con la seña caminar hacia adelante o doblar la esquina a la izquierda o a la derecha. Por lo tanto, no se trata de una seña estandarizada.

Schembri (2003, p. 18) muestra una serie de dificultades para encontrar una descripción de los clasificadores que pueda ser transversal entre las lenguas habladas y signadas. Lo relevante del fenómeno de los clasificadores para este estudio es su presencia en ambas modalidades lingüísticas, su fuerte relación con el discurso, las dificultades y la falta de consenso para encontrar definiciones que satisfagan lo que ocurre en ambas modalidades y el carácter cerrado de estas piezas.

Zwitserlood, 2012 (p.158) describió los clasificadores como "... morfemas sin significado específico que se expresan mediante configuraciones particulares del articulador manual y que representan entidades denotando características salientes…". El estatus de los clasificadores como elementos léxicos es controvertido y existen diferentes enfoques al respecto (véase Emmorey, 2003).

Además, los estudios sobre los errores en la producción de la *Deutschebärdesprache* (DGS) (Leuninger et al., 2007) sugieren que la bipartición del léxico mental en un componente de lema y otro de lexema parece ser válida para ambas modalidades lingüísticas (oral vs de señas). Por ejemplo, en español y LSU, la seña "Mesa de luz" está constituida por el lema mesa y el lexema luz. Por último, factores como la familiaridad, la vecindad fonológica y la frecuencia de las piezas léxicas son factores igualmente relevantes con importantes similitudes en el procesamiento de las lenguas orales y de señas (Carreiras et al., 2008).

Aspectos psicolingüísticos Dentro de los aspectos psicolingüísticos, la literatura actual sobre la organización y el desarrollo léxico de los sordos tiene tres líneas principales de discusión relevantes para una perspectiva comparativa como la de nuestro manuscrito: el debate sobre (i) el papel de la experiencia temprana en la seña para las primeras etapas del desarrollo léxico, (ii) el peso de la iconicidad en el vocabulario temprano, y (iii) la ventaja de la seña en el desarrollo léxico temprano de los sordos sobre los oyentes.

La privación lingüística se relaciona con la primera línea de discusión porque este tipo de privación se entiende como la falta de estímulos lingüísticos necesarios para los procesos de adquisición del lenguaje. La exposición temprana a una primera lengua —en nuestro caso, una lengua de señas, el LSU, y una lengua oral, la española— se considera un predictor de futuros resultados lingüísticos. Por ejemplo, Hall (2017) sugiere que los niños sordos que aprenden a señar más tarde en la vida tienen más probabilidades de procesar las señas no como input lingüístico sino visual. Esta situación contrasta con los niños expuestos desde el nacimiento al lenguaje de señas. Éstos procesan las señas en la misma región del cerebro en la que los oyentes procesan el lenguaje oral. Además, Caselli et al. (2020) descubrieron que la privación del lenguaje tiene efectos adversos sobre el procesamiento fonológico, la velocidad de reconocimiento de las señas y la precisión. En un ámbito más general, estos autores descubrieron que la forma organiza los léxicos de las lenguas orales y de señas, que el reconocimiento del léxico se produce a través de la competencia basada en la forma, y que los mapeos forma-significado no impulsan el acceso léxico incluso cuando la iconicidad es omnipresente en el léxico.

En cuanto al vocabulario temprano, Meier et al. (2008) informaron que, en sus primeras señas, la iconicidad no es un factor importante en el rendimiento fonológico de los niños sordos. En lugar de la iconicidad, las categorías semánticas parecen ser un factor relevante en la organización del vocabulario temprano de los niños expuestos a las seña (Chen Pichler, 2012, p. 659).

Por último, la ventaja de las señas en el desarrollo léxico temprano fundamenta la suposición de que los parámetros de las señas se desarrollan antes que los de las lenguas orales. En consecuencia, los niños señantes pueden producir piezas léxicas antes que los oyentes y acelerar las primeras etapas del desarrollo del lenguaje. Algunos estudios apoyan estas afirmaciones (Bonvillian et al., 1990; Meier et al., 2008), y otros no (Meier & Newport, 1990).

También existen importantes diferencias de grupo en lo que respecta al aprendizaje formal de la lengua. A pesar de las diferentes experiencias de aprendizaje léxico formal e informal de los niños sordos (Marschark & Wauters, 2008), muchos estudios sobre ASL y otras lenguas de señas han sugerido tendencias de desarrollo similares a las reportadas para sus contrapartes orales (ASL: Novogrodsky et al., 2014a; Novogrodsky et al., 2014b, *British Sign Language*: Mann & Marshall, 2012; Marshall et al., 2013; Mason et al., 2010, y *Lingua dei segni italiana*: Tomasuolo et al., 2010). Por último, las lenguas de señas generalmente carecen de una forma escrita estandarizada (Meir, 2012), lo que deja a las personas sordas sin este recurso para mejorar sus experiencias de aprendizaje de vocabulario (Goldin-Meadow y Mayberry, 2001).

Aspectos sociolingüísticos Dentro de los aspectos sociolingüísticos, el reconocimiento legal de la lengua de señas es un factor importante para su desarrollo, ya que fomenta su enseñanza y uso, al menos en espacios públicos o gubernamentales. Por otro lado, el rápido desarrollo tecnológico, la perspectiva médica de la sordera y la disminución del número de personas sordas en todo el mundo ponen en peligro las lenguas de señas. Recibir un implante coclear a una edad muy temprana suele ir acompañado de una escasa o nula exposición temprana de los niños sordos a la lengua de señas (Schermer, 2012, p. 904) y una distancia identitaria respecto a la comunidad sorda.

El mismo autor (2012, p. 903) indica que factores como la inclusión de la lengua de señas en el sistema educativo, los programas bilingües con la lengua oral y la formación de intérpretes en los niveles secundario y terciario han ayudado al desarrollo de nuevas señas. Por ejemplo, en asignaturas académicas (escolares) y técnicas.

En contextos asimétricos de contacto lingüístico intermodal, la lengua oral principal afecta al desarrollo de una lengua de señas minoritaria y sus espacios de uso pueden ser muy limitados (por ejemplo, en la asociación de sordos, en la familia si los padres o hermanos hablan con señas). En estos contextos, los mismos señantes considerarían inadecuado utilizar las lenguas de señas en algunas situaciones de devaluación de las mismas. Entonces, el hablante de la lengua de señas se vería obligado a dominar la forma escrita de la lengua oral mayoritaria, con la pérdida de rasgos de la lengua de señas, que sería minoritaria (Adam, 2012, p.842).

Esta situación podría inhibir la acuñación de nuevas señas para los sujetos de esos contextos y el interés de la comunidad sorda por generar un uso más normalizado de algunas señas. Los efectos del contacto lingüístico incluyen una serie de fenómenos como préstamos, interferencia, convergencia, habla extranjera, cambio de código, cambio, desgaste, declive o muerte de la lengua (Thomason, 2001).

Contra los efectos adversos de la asimetría en el contacto intermodal de las lenguas, como indica Schermer (2012, p. 904), existen herramientas para ampliar el vocabulario, como la creación de bases de datos disponibles con un léxico ampliado sistemáticamente (por ejemplo, el léxico de la Nederlandse Gebarentaal [NGT]) o la acuñación de nuevas señas con un equipo de señantes sordos nativos, lingüistas (sordos) y personas que tengan los conocimientos de contenido necesarios (por ejemplo, los Arbeitsgruppe Fachgebärden de la Universidad de Hamburgo). Las nuevas piezas léxicas se acuñan por medios externos (préstamos de otras lenguas) e internos (creación o derivación de nuevas palabras a partir de combinaciones sonoras permitidas, compuestos y procesos morfológicos).

Además, los préstamos lingüísticos intermodales pueden repercutir en el léxico de otra lengua. McKee et al. (2007) describieron la importación de piezas léxicas orales a las lenguas de signos. Esta importación se produce a través del deletreo con los dedos, la pronunciación con la boca, las formaciones de señas inicializadas y la traducción de préstamos. Las formas extranjeras son otra alternativa. Éstas combinan elementos estructurales de dos lenguas, a veces a través de la pronunciación y las señas inicializadas. Adam (2012, p. 847) indica que los individuos bilingües (por ejemplo LSU-español o viceversa) son fundamentales para introducir nuevos usos y acuñaciones de una segunda lengua en la comunidad. Sin embargo, Plaza-Poust & Weinmeister (2008, en Plaza-Poust, 2012) demostraron que el préstamo léxico y estructural entre el alemán y la *Deutschgebärdesprache* (DGS) se produce en etapas de desarrollo específicas para ambas lenguas y disminuye a medida que los hablantes progresan y adquieren fluidez. Genesse, (2002, en Plaza-Poust, 2012) demostró que las variaciones individuales en las lenguas de señas mostraban patrones similares a los encontrados en estudios comparativos entre lenguas orales.

El bilingüismo intermodal es un conjunto de situaciones en las que más de una segunda lengua intermodal es utilizada por todos o un amplio conjunto de miembros de una comunidad (bilingüismo) o sólo un grupo selecto de miembros de esta comunidad es bilingüe (bilingüismo) de forma intermodal. Como indica Adam (2012, p. 847), el préstamo léxico "... se produce cuando los hablantes en contacto con otra lengua más dominante (es decir, una lengua oral como el español) perciben una laguna o una necesidad de referencia a nuevos conceptos extranjeros en su primera lengua (por ejemplo, la LSU)..." Como resultado, el léxico de la lengua no dominante se amplía o nuevas piezas léxicas sustituyen a las ya existentes.

A manera de ejemplo: el caso de los conceptos de espacio y tiempo en el léxico mental

El espacio y el tiempo son ámbitos de experiencia universales para los seres humanos. Son candidatos a ocupar un lugar destacado en cualquier lengua y cultura. La literatura cognitiva (por ejemplo, Lakoff y Johnson, 1980) considera el espacio como un dominio concreto de la experiencia en el sentido de que el organismo humano tiene acceso perceptivo-motor a información específica del dominio (por ejemplo, la vista y el tacto nos dan información sobre lo alta o baja que es una persona). Por el contrario, el tiempo se considera un dominio abstracto en la medida en que no existe tal acceso sensoriomotor a información específica del dominio (es decir, el color naranja de la línea del horizonte y la posición baja del sol son información espacial -no temporal- que sin embargo indica un intervalo de tiempo del día).

Esta descripción de lo concreto y lo abstracto se ha evidenciado en poblaciones de diferentes culturas y con medios verbales, no verbales y gestuales. Para una revisión, véase Núñez y Cooperrider (2013), Bender y Beller (2014), y Callizo-Romero et al. (2022). Sin embargo, la representación del espacio y el tiempo en las lenguas orales y de señas presenta similitudes y diferencias cuyo efecto sobre la conceptualización y el léxico mental es todavía una cuestión abierta.

A diferencia de las lenguas orales, en las lenguas de señas el espacio es la sustancia en la que tiene lugar la producción lingüística. Wilcox y Martínez (2020) sugieren que la conceptualización del espacio en las lenguas de señas se manifiesta semántica y fonológicamente. Esta conceptualización da valor semántico y fonológico al espacio que rodea al señante (espacio señante) y a otros lugares (es decir, el de los objetos a los que el señante señala, la ubicación de esos objetos y el movimiento de esos objetos en el espacio) en una estructura simbólica compleja. Esta diferencia con las lenguas orales puede sugerir diferencias en la modalidad lingüística importantes para la conceptualización del espacio y la memoria espacial (para la relación entre el lenguaje espacial y la memoria espacial en la población sorda, véase Karadöller et al., 2022).

En cuanto a cómo se expresa el tiempo en las lenguas de señas, Pereiro y Soneira (2004: 65-66) sugieren que el tiempo se codifica mediante el uso de ítems léxicos, elementos gramaticales o líneas de tiempo. Los autores mencionan tres expresiones de línea de tiempo: una perpendicular al cuerpo del señante, otra localizada delante del cuerpo del señante y la última etiquetada como línea de crecimiento. Las lenguas de señas estudiadas hasta ahora (Emmorey, 2001) presentan un amplio repertorio de alternativas de línea de tiempo.

En cambio, en las lenguas orales, sólo algunos elementos léxicos espaciales se utilizan con valor temporal. Por ejemplo, izquierda y derecha no tienen valor temporal en ninguna lengua oral (Haspelmath, 1997), pero sí lo tienen atrás, adelante, arriba y abajo (por ejemplo, expresiones españolas como *El pasado quedó atrás*). Así pues, el espacio físico que rodea al hablante no tiene valor semántico ni fonológico.

Estos aspectos han de considerase cuando nos preguntamos, por ejemplo, si las personas sordas y oyentes tienen la misma organización conceptual y el mismo perfil de interpretación (concreto frente a abstracto) de los dominios del tiempo y el espacio. Las redes semánticas muestran cómo la memoria describe la organización de los hechos declarativos y el conocimiento en la mente (véase Collins y Loftus, 1975). Entonces, como procedimiento, las redes semánticas pueden decirnos sobre los procesos de memoria implicados en la categorización conceptual de las poblaciones sorda y oyente.

Por su parte, el procedimiento de asociación libre repetida aportaría evidencias sobre las similitudes y diferencias con el que se establecería —por parte de los hablantes— el léxico mental espacio-temporal, considerando que la modalidad lingüística podría explicar la variabilidad del constructo. Entre sus virtudes, este paradigma permite captar todo tipo de respuestas (sintagmáticas vs. paradigmáticas), todo tipo de relaciones semánticas (taxonomía vs. entidades vs. situación vs. introspección) y todo tipo de piezas léxicas (abiertas vs. cerradas).

Con los elementos de la revisión expuesta en este artículo, habría razones para hipotetizar que si las diferencias de modalidad lingüística tienen efectos sobre la categorización conceptual, entonces podría haber sesgos entre tipos de categorización por modalidad consistentes con alternativas en el uso de información sensoriomotora, lingüística y social. Debido a que la iconicidad se basa en la concreción (en la que la información sensoriomotora se vuelve más relevante) y la iconicidad es omnipresente en las lenguas de señas, los señantes sordos podrían preferir relaciones semánticas más sensibles a la información sensoriomotora e interpretaciones concretas que los oyentes. Por ejemplo, las preferencias podrían adoptar la forma de un sesgo hacia interpretaciones concretas de los dominios de espacio y tiempo para la población de señantes (principalmente respuestas sintagmáticas y relaciones semánticas de entidad y situación), pero no para la población oyente. Previsiblemente, el adecuado control de variables sociolingüísticas, psicolingüísitcas y de desarrollo entre poblaciones de señantes y oyentes, que pudiesen empañar a la modalidad como variable explicado de la organización del léxico mental, es una complicación de mayor magnitud y posiblemente insalvable. La población de oyentes hijos de sordos señantes (Child of Deaf -CODA por su abreviatura en inglés) es una buena alternativa para zanjar la dificultad antes mencionada.

Acknowledgements

This research was supported by I+D grant of the Comisión Sectorial de Investigación Científica of the Universidad de la República (I+D CSIC 2014) to RA. Portions of this research were presented at the International Online Conference of Psycholinguistic and Neurolinguistic Research: Methods, Materials, and Approaches in 2020, Annual Conference of the Cognitive Science Society, I Internation Congress of the SUCCC, XVIII Reunión Nacional y VII Encuentro Internacional de la Asociación Argentina de Ciencias del Comportamiento in 2021 and in Concepts in Action. Representation, Learning, and Application (CARLA), in 2022. We thank Rocío Goñi for her enthusiastic collecting of data with the hearing population.

Author contributions

RA and JV conceived the study. MNM, RA, JV, AF, MY contributed to the documental review. RA, JV, and JVA contributed to the writing of the article.

Disclosures

The authors declare no conflict of interests.

Referencias

Adam, R.. (2012). Language contact and borrowing. In: R. Pfau, M. Steinbach, & W Bencie (Eds.), *Sign Language: An International Handbook*, (pp. 841–861). Walter de Gruyter.

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science, 6*(3), 359–370. https://doi.org/10.1111/tops.12096

Barsalou, L. W. (1987). The Instability of Graded Structure: Implications for the nature of concepts. In: U. Neisser (Ed.), *Concepts and Conceptual Development*, (pp. 101–140). Cambridge University Press

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bender, A., & Beller, S. (2014). Mapping spatial frames of reference onto time: A review of theoretical accounts and empirical findings. *Cognition, 132*(3), 342–382. https://doi.org/10.1016/j.cognition.2014.03.016

Borghi, A. M., & Caramelli, N. (2003). Situation bounded conceptual organization in children: From action to spatial relations. *Cognitive Development, 18*(1), 49–60. https://doi.org/10.1016/S0885-2014(02)00161-2

Borghi, A. M., & Cimatti, F. (2009). Words as tools and the problem of abstract word meanings. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, 31).

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts* (Vol. 2). Springer.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). *The challenge of abstract concepts. Psychological Bulletin, 143*(3), 263, https://doi.org/10.1037/bul0000089

Callizo-Romero, C., Tutnjević, S., Pandza, M., Ouellet, M., Kranjec, A., Ilić, S., ... & Santiago, J. (2020). Temporal focus and time spatialization across cultures. *Psychonomic Bulletin & Review, 27*(6), 1247–1258. https://doi.org/10.3758/s13423-020-01760-5

Carreiras, M., Gutiérrez-Sigut, E., Baquero, S., & Corina, D. (2008). Lexical processing in Spanish Sign language (LSE). *Journal of Memory and Language, 58*(1), 100–122. https://doi.org/10.1016/j.jml.2007.05.004

Caselli, N. K., Hall, W. C., & Henner, J. (2020). American Sign Language interpreters in public schools: An illusion of inclusion that perpetuates language deprivation. *Maternal and Child Health Journal, 24*(11), 1323–1329. https://doi.org/10.1007/s10995-020-02975-7

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. Psychological Review, 82(6), 407. https://doi.org/10.1037/0033-295X.82.6.407

Connell, L., & Lynott, D. (2012). The strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125*(3), 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Cuxac, C. & Sallandre, M.A. (2007) Iconicity and arbitrariness in French Sign Language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity. In E. Pizzuto, P. Pietrandrea, & R. Simone (Eds.), *Verbal and signed languages: Comparing structures, concepts and methodologies*, (pp. 1–33). Mouton de Gruyter.

Chen Pichler, D. (2012). Acquisition. In: R. Pfau, M. Steinbach, & W. Bencie (Eds.), *Sign Language: An International Handbook* (pp. 647–686). Walter de Gruyter.

De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2017). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Big data in Cognitive Science* (pp. 174–202). Routledge/Taylor & Francis Group.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Dong, Y., & Li, P. (2015). The cognitive science of bilingualism. *Language and Linguistics Compass, 9*(1), 1–13. https://doi.org/10.1111/lnc3.12099

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition, 110*, 412–431. https://doi.org/10.1016/j.cognition.2008.11.016

Elbers L. & Van Loon-Vervoorn, A. (1998). Acquiring the lexicon. In: S. Gillis S, A. De Houwer A. (Eds.), *The acquisition of Dutch*. Amsterdam: John Benjamins.

Emmorey, K. (2001). *Language, cognition, and the brain: Insights from sign language research*. Psychology Press.

Emmorey, K. (Ed.). (2003). *Perspectives on classifier constructions in sign languages*. Psychology Press.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations. In: K. D. Federmeier (Ed.), *Psychology of Learning and Motivation*, (Vol. 54, pp. 249–294). Academic Press.

Fürth, H. G. (1971). Linguistic deficiency and thinking: Research with deaf subjects. 1964-1969. *Psychological Bulletin & Review, 76*, 58–72. https://doi: 10.1037/h0031495. PMID: 4934101.

García Coni, A., Ison, M. y Vivas, J. (2019). Conceptual Flexibility in Schoolchildren: switching between taxonomic and thematic relations. *Cognitive Development 52*, 1–13. https://doi.org/10.1016/j.cogdev.2019.100827

Gentner, D. & Ratterman, M.J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development*, (pp. 225–277). Cambridge University Press.

Glenberg, A. M., Kaschak, M. P. (2003). The body's contribution to language. In: B. H. Ross (Ed.), *The Psychology of Learning and Motivation, 43* (pp. 93–126). Academic Press.

Goldin-Meadow, S., & Mayberry, R. I. (2001). How do profoundly deaf children learn to read? *Learning Disabilities Research & Practice, 16*(4), 222–229. https://doi.org/10.1111/0938-8982.00022

Hall, W. C. (2017). What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. Maternal and Child Health Journal, 21(5), 961–965. https://doi.org/10.1007/s10995-017-2287-y

Haspelmath, M. (1997). *From space to time*. Lincom.

Hohenberger, A. & Leuninger, H. (2012). Psycholinguistics and neurolinguistics. In: R. Pfau, M. Steinbach, & W Bencie (Eds.). *Sign Language: An International Handbook*, (pp. 711–738). Berlin: Walter de Gruyter.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241–254. https://doi.org/10.1007/BF02289588

Karadöller, D. Z., Sümer, B., Ünal, E. et al. (2022). Late sign language exposure does not modulate the relation between spatial language and spatial memory in deaf children and adults. *Memory and Cognition*. https://doi.org/10.3758/s13421-022-01281-7

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14. https://doi.org/10.1037/a0021446

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195–208. https://doi.org/10.1016/S0364-0213(80)80017-6

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites* (Vol. 1). Stanford University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Leuninger, H., Hohenberger, A., & Waleschkowski, E. (2007). Sign language: typology vs. modality. In: T. Carson Schütze & V. S. Ferreira (Eds.), *The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop*, (vol. 53, pp. 317–345). MIT Working Papers in Linguistics.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. London: MIT.

Mann, W., & Marshall, C. (2012). Investigating deaf children's vocabulary knowledge in British Sign Language. *Language Learning, 62*(4), 1024–1051. https://doi.org/10.1111/j.1467-9922.2011.00670.x

Mann, W., Sheng, L., & Morgan, G. (2016). Lexical-semantic organization in bilingually developing deaf children with ASL-dominant language exposure: Evidence from a repeated meaning association task. *Language Learning, 66*(4), 872–899. https://doi.org/10.1111/lang.12169

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(4), 329–358. https://doi.org/10.1080/09528130110100252

Marshall, C., Rowley, K., Mason, K., Herman, R., & Morgan, G. (2013). Lexical organisation in deaf children who use British Sign Language: Evidence from a semantic fluency task. *Journal of Child Language*, *40*(1), 193–220. https://doi: 10.1017/S0305000912000116

Marschark, M., Convertino, C., McEvoy, C., & Masteller, A. (2004). Organization and use of the mental lexicon by deaf and hearing individuals. *American Annals of the Deaf*, *149*(1), 51–61. https://www.jstor.org/stable/26234642

Marschark, M., & Wauters, L. (2008). Language comprehension and learning by deaf students. In M. Marschark & P. C. Hauser (Eds.), *Perspectives on deafness. Deaf cognition: Foundations and outcomes*, (p. 309–350). Oxford University Press.

Mason, K., Rowley, K., Marshall, C. R., Atkinson, J. R., Herman, R., Woll, B., & Morgan, G. (2010). Identifying specific language impairment in deaf children acquiring British Sign Language: Implications for theory and practice. *British Journal of Developmental Psychology*, *28*(1), 33–49. https://doi: 10.1017/S0305000912000116

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

McRae, K., & Jones, M. (2013). 14 Semantic Memory. In: D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology,* (pp. 206–219). Oxford University Press.

Meier, R. (2002). Why Different, Why the Same? Explaining Effects and Non-effects of Modality Upon Linguistic Structure in Sign and Speech. In: R. Meier, Richard, K. Cormier, D. Quinto-Pozos, (Eds.), *Modality and Structure in Signed and Spoken Languages*, (pp. 1–25). Cambridge University Press.

Meir, I. (2012). Morphology. II. Word class and word formation. In: R. Pfau, M. Steinbach, & W Bencie (Eds.), *Sign Language: An International Handbook,* (pp. 77–111). Walter de Gruyter.

Morgan, G., & Woll, B. (Eds.). (2002). *Directions in sign language acquisition* (Vol. 2). John Benjamins Publishing.

Myklebust, H., & Brutton, M. (1953). A study of visual perception in deaf children. *Acta Oto-Laryngologica, Supplementum*, *105*:1–126.

Novogrodsky, R., Caldwell-Harris, C., Fish, S., & Hoffmeister, R. J. (2014a). The development of antonym knowledge in American Sign Language (ASL) and its relationship to reading comprehension in English. *Language Learning*, *64*(4), 749–770. https://doi.org/10.1111/lang.12078

Novogrodsky, R., Fish, S., & Hoffmeister, R. (2014b). The Acquisition of Synonyms in American Sign Language (ASL) Toward a Further Understanding of the Components of ASL Vocabulary Knowledge. *Sign Language Studies*, *14*(2), 225–249. https://www.jstor.org/stable/26190871

Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, *17*(5), 220–229. https://doi.org/10.1016/j.tics.2013.03.008

Ortega, G., & Morgan, G. (2015). Phonological development in hearing learners of a sign language: The influence of phonological parameters, sign complexity, and iconicity. *Language Learning*, *65*(3), 660–688. https://doi.org/10.1111/lang.12123

Ortega, G., Sümer B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. Developmental Psychology, 53(1), 89–99. https://doi.org/10.1037/dev0000161

Parks, E., & Williams, H. (2013). *The Uruguayan Deaf Community. Journal of Language Survey Reports*, *21*. Texas: Sil International.

Pereiro, C. C., & Soneira, A. F. (2004). The expression of time in Spanish Sign Language (LSE). *Sign Language & Linguistics*, *7*(1), 63–82. https://doi.org/10.1075/sll.7.1.06cab

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227. https://doi.org/10.3389/fpsyg.2010.0022

Plaza Pust, P. (2012). Morphology. 5. Word class and word formation. In: R. Pfau, M. Steinbach, & W. Bencie (Eds.), *Sign Language: An International Handbook*, (pp. 949–979). Walter de Gruyter.

Quillian, M. R. (1967). Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*, 410–430. https://doi.org/10.1002/bs.3830120511

Radden, G. & Dirven, R. (2007). *Cognitive English Grammar* (Vol. 2). John Benjamins Publishing.

Rodríguez-Cuadrado, S., Ojedo, F., Vicente-Conesa, F., Romero-Rivas, C., Sampedro, M. A., & Santiago, J. (accepted). Sign iconicity helps learning new words for abstract concepts in a foreign language. *Second Language Research*.

Sheng, L., McGregor, K. K., & Marian, V. (2006). Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language and Hearing Research*, *49*(3): 572–587. https://doi.org/10.1044/1092-4388(2006/041)

Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *53*(1), 146-159. https://doi.org/10.1044/1092-4388(2009/08-0160

Siegler, R. S. & Shipley, C. (1995). Variation, selection, and cognitive change. In: T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*, (pp. 31–76). Hillsdale: Erlbaum.

Smith, A. J. (1976). On the effectiveness of set associative page mapping and its application to main memory management. In: *Proceedings of the 2nd International Conference on Software Engineering* (pp. 286–292).

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.

Thomason, S. G. (2001). *Language contact*. Edinburgh University Press.

Tomasuolo, E., Fellini, L., Di Renzo, A., & Volterra, V. (2010). Assessing lexical production in deaf signing children with the Boston naming test. *Language, Interaction and Acquisition*, *1*(1), 110–128. https://doi.org/10.1075/lia.1.1.07tom

Wilcox, P. P. (2000). *Metaphor in American Sign Language*. Gallaudet University Press.

Wilcox S. & Martínez, R. (2020). The Conceptualization of Space: Places in Signed Language Discourse. *Frontiers in Psychology.* https://doi: 10.3389/fpsyg.2020.01406

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Zwaan, R. (2004). Time in Language, Situation Models and Mental Simulation. *Language Learning*, *58*(s1): 13–26. https://doi.org/10.1111/j.1467-9922.2008.00458.x

Zwiebel, A., & Mertens, D. M. (1985). A comparison of intellectual structure in deaf and hearing children. *American Annals of the Deaf*, *130*(1), 27–31. https://doi.org/10.1353/aad.2012.0896

Zwitserlood, I. (2012). Classifiers. In: R. Pfau, M. Steinbach, & W Bencie (Eds.), *Sign Language: An International Handbook,* (pp. 158–185). Walter de Gruyter.

Sesión de entrenamiento para el uso de equipos de Estimulación Transcraneal por Corriente Directa (tDCS)
Training session for using Transcranial Direct Current Stimulation (tDCS) instrumentos

En la foto, la doctora Lorena Channes, de la Universitat Autònoma de Barcelona, y la doctora Ana Pires, investigadora principal del **CIBPsi**, durante una sesión del Neurostimation Workshop: Transcranial Direct Current Stimulation (tDCS). (11/2013 - 11/2013).

In the picture, Dr. Lorena Channes, from the Universitat Autònoma de Barcelona, and Dr. Ana Pires, main researcher of **CIBPsi**, during a session of the Neurostimation Workshop: Transcranial Direct Current Stimulation (tDCS). (11/2013 - 11/2013).

Neural processing of iterated prisoner's dilemma outcomes indicates next-round choice and speed to reciprocate cooperation

Francisco Cervantes Constantino[1,], Santiago Garat[1], Eliana Nicolaisen-Sobesky[1], Valentina Paz[1], Eduardo Martínez-Montes[2], Dominique Kessel[3], Álvaro Cabana[1], and Victoria B Gradin[1]

Francisco Cervantes Constantino https://orcid.org/0000-0002-7393-3579, Eliana Nicolaisen-Sobesky https://orcid.org/0000-0002-4806-2863, Valentina Paz https://orcid.org/0000-0002-2401-5472, Eduardo Martínez-Montes https://orcid.org/0000-0002-5852-964X, Dominique Kessel https://orcid.org/0000-0001-8275-9923, Álvaro Cabana https://orcid.org/0000-0002-8637-290X, and Victoria B Gradin https://orcid.org/0000-0001-6599-3523

1 **Centro de Investigación Básica en Psicología (CIBPsi).** Facultad de Psicología. Universidad de la República, Uruguay
2 Neuroinformatics Department, Cuban Center for Neuroscience, Cuba
3 Facultad de Psicología, Universidad Autónoma de Madrid, Madrid

Correspondence concerning this manuscript should be addressed to Francisco Cervantes. Centro de Investigación Básica en Psicología. Facultad de Psicología, Universidad de la República. Dr. Tristán Narvaja 1674. Montevideo 11200, Uruguay. Telephone: +598 24008555 ext 285. Fax: +598 24008640. E-mail: **francisco.cervantesconstantino@gmail.com**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Cervantes Constantino, F., Garat, S., Nicolaisen-Sobesky, E., Paz, V., Martínez-Montes, E., Kessel, D., Cabana, Á., & Gradin, V. B. (2021). Neural processing of iterated prisoner's dilemma outcomes indicates next-round choice and speed to reciprocate cooperation. *Social Neuroscience*, *16*(2): 103-120. Available at **https://doi.org/10.1080/17470919.2020.1859410**

Abstract

The iterated prisoner's dilemma (iPD) game is a well established model for examining how people cooperate with each other, yet the neural processes that unfold after its distinct outcomes have been only partly described. Recent theoretical models emphasize the ubiquity of intuitive cooperation, raising questions on the neural timelines involved. We studied the outcome feedback stage of iPD rounds with electroencephalography (EEG) methods. Results showed that neural signals associated to this stage also relate future choice in an outcome-dependent manner: (i) after zero-gain 'sucker's payoffs' (unreciprocated cooperation), a participant's decision thereafter relates to changes to the feedback-related negativity (FRN); (ii) after one-sided non-cooperation (participant gain), by the P3; (iii) after mutual cooperation, by late frontal delta-band modulations. Critically, faster choices to reciprocate cooperation were predicted, on a single-trial basis, by P3 and frontal delta modulations at the immediately preceding trial. Delta band signaling is considered in relation to homeostatic regulation processing in the literature. The findings relate the relatively early outcome feedback stage to subsequent decisional processes in the iPD, providing a first neural account of the brief timelines implied in heuristic modes of cooperation.

Keywords: Social heuristics, Prisoner's dilemma, Delta, P3, FRN

Introduction

Organized social life is based on balancing self and group benefit-seeking, a hallmark of cooperation[1]. Social scientists often model sustained cooperative interactions based on the iterated prisoner's dilemma (iPD)[2,3]. In this iconic game, two agents make simultaneous and independent decisions on whether to cooperate or defect with each other, and each agent receives a payoff depending jointly on both decisions. One-sided outcomes lead to imbalanced payoffs that may trigger reports of anger or betrayal in the cooperator and elation (but also guilt) in the defector; balanced mutual cooperation usually induces positive outcomes such as bonding or trust[4,5]. The iPD outcomes imply varied social contexts that allow to recreate relevant scenarios from which to investigate the neural dynamics of cooperation. Through iPD outcomes' analysis it is also possible to explore how different social contexts relate to future decisions on cooperation.

Cooperative cognition and its neural bases have been addressed in relation to feedback processing at the iPD, typically via functional magnetic resonance imaging[4,6–14] and in some cases including populations of clinical interest[5,15–21]. High temporal resolution electroencephalography (EEG) methods have also been applied[22–24], with two previous studies directly addressing the processing of the game outcomes[25,26]. These two studies identified two relevant event related potential (ERP) components, the feedback related negativity (FRN) and the P3 (or P3b), both of which are often featured in the feedback processing literature[27]. The FRN is a deflection elicited shortly after feedback (200 to 350 ms) and represents the earliest index in reward outcome evaluation; it usually shows a more negative amplitude following negative versus positive outcomes[27–33]. When social feedback is involved, the FRN may show sensitivity to social comparisons[34], fairness[35–38] and cooperation[25,26], and may comprise the affective impact of an event[28,39]. In *value-based* decision-making studies, the P3 (or P3b) is a mid-latency (300 to 600 ms) centroparietal deflection showing sensitivity to reward magnitude, valence and probability or risk[27,30,40–48]. In social studies it has similarly shown modulation by social comparisons[34,49], fairness[37,49,50] and cooperation[26].

ERP analyses are popular tools for inferring neural dynamics and timing of economic decision-making, and so are spectrotemporal methods[25–27,51]. However, to our knowledge, no study has investigated yet the late (e.g. > 600 ms) ERP signals elicited by the iPD outcomes, neither the associated frequency band activity. Feedback ERPs often include the late positive potential (LPP, cf.[52]) which is involved in emotional processing[27,53–56]. Spectrotemporal activity involved in feedback processing (e.g.[57]) has not been addressed in the game, except for derivate measures in hyperscanning protocols[22–24]. The relatively late window (> 500 ms) represents an intermediate phase between early neural updating by the outcome and next round decision-making behavior. Therefore, an investigation of feedback processing related to decision making may need to address these later processes, which also potentially mediate integration of the slow dynamics sometimes involved in decision-making[27]. It is also important to consider how the differential processing of outcome types reflect upon the distinct social and emotional contexts implied. Recent advances in the neurobiology of social behavior suggest the processing of such contexts operate according to the organization of homeostatic systems that maintain social connection[149]. Homeostasis here refers to adaptive processes that aim towards equilibrium of emotional and motivational states; in the case of social information, these may include two principal drivers in social interaction: the aversive state of isolation and the hedonic value of social reward[149.]

The current study had two main objectives. First, we set to comprehensively characterize EEG signals from the feedback processing stage of the iPD, including an analysis of FRN and P3 but also of (previously unexplored) late ERP signals and corresponding spectrotemporal activity. As a second objective, we investigated how the social context imposed by each outcome type relates to next decision. At this point, deciding whether to cooperate relies on potentially different assessments about the co-player, e.g. whether they are interested in a sustained, retaliatory, or deceptive interaction. This leads to situations where, even after experiencing the same outcome (e.g. CD's on separate occasions), a player may choose to cooperate or not thereafter. For this we investigated, for each outcome type, whether the same neural signals demonstrating sensitivity to outcomes were also predictive of a participant's subsequent choice. Of particular interest was the potential timing relationship between relevant neural activity after outcome presentation and effective choice at the next round. Reaction times are a central variable in the 'social heuristics hypothesis', a dual-process theory[58,59] application to cooperation. On the basis of the time taken to decide, faster decisions have been interpreted as indication of cooperative intuition, whereas slower decision-making may reflect as'second thoughts' of a more self-interested nature[60,61]. Findings that faster reaction times typically reflect cooperation have led to suggest a cooperative heuristic 'default' in decision-making[62,63]. The approach has been critiqued however[150] by relatively long decision times in behavioral studies, which makes it difficult to determine what relevant processes are automatic. Hence, we employed the present EEG feedback signal and timing analyses as a means to address candidate neural correlates of decision-making that are automatic (short latency) and fast (short reaction times) when facing the prospect of repeated cooperative interactions.

Materials and methods

Subjects Thirty-one volunteers (16 female; mean age 22.3 ± 2.9 SD) with no history of neurological or psychiatric disorders participated in the present study. All reported normal or corrected-to-normal visual acuity, and provided written informed consent for their participation. All experiments were performed in accordance with the World Medical Association's Declaration of Helsinki guidelines. Approval of the experimental procedures was obtained by the Faculty of Psychology at Universidad de la República.

Experimental setup Presentation and response time logging were performed with PsychoPy[64] software, delivered over a cathode ray tube monitor (E. Systems, Inc., CA) with 40 cm size, 83 dpi resolution, and 60 Hz refresh rate. EEG recordings were performed using a BioSemi ActiveTwo 64-channel system (BioSemi, The Netherlands) with 10/20 layout, at 2048 Hz digitization rate with CMS/DRL (ground) and two mastoid reference electrodes. A 5th order cascaded integrator-comb low-pass filter with -3 dB point at 410 Hz was applied online, after which signals were decimated to 256 Hz. Online high-pass response was fully direct current (DC) coupled. Full experimental sessions lasted ~2.5 h.

Iterated prisoner's dilemma task Each participant was introduced to a same-sex person, presented as their co-player for the study but in fact a confederate, i.e. an associated researcher with no prior knowledge of the participant. The confederate and the participant were then escorted to different rooms. The participant was instructed that, on each trial, both players would have to make a simultaneous and independent decision regarding whether to cooperate or not cooperate with each other. On each round, and depending on their decisions, they would both receive points. Mutual cooperation ('CC') resulted in both players earning 2 points and mutual non-cooperation ('DD') resulted in a payoff of 1 point for each player, while unbalanced outcomes ('CD' and 'DC') earned 0 points for the player that cooperated and 3 for the player that did not. Participants were told that points accumulated over time, and were instructed to maximize earnings. They did a practice test and waited for the experimenter to supposedly provide similar instructions to the confederate.

During EEG recordings, participants played the iPD over a 14.4° by 6.2° visual interface. In each round, the payoff matrix was displayed (Figure 1A) and participants were required to decide whether to 'Cooperate' ('C') or 'Not cooperate' ('D') by pressing one of two buttons (left/right). After a fixation, the "joint" final outcome was highlighted as feedback, e.g. a 'CD' outcome means the participant cooperated but the co-player did not (i.e. a 'sucker's payoff'). Decision-making by the confederate was simulated by a probabilistic 'tit-for-tat' style algorithm. In the first round, the algorithm cooperates and in subsequent rounds it reciprocates the player's choice with 80% probability; after three consecutive rounds of identical, mutually-balanced outcomes, it switches to the alternative option (i.e. after three CC outcomes, the algorithm switches to D). Three pauses were offered every 50 rounds, completing the session with 200 rounds. Afterwards, participants rated their emotional response (happiness, anger, sadness, betrayal, and guilt) to each of the

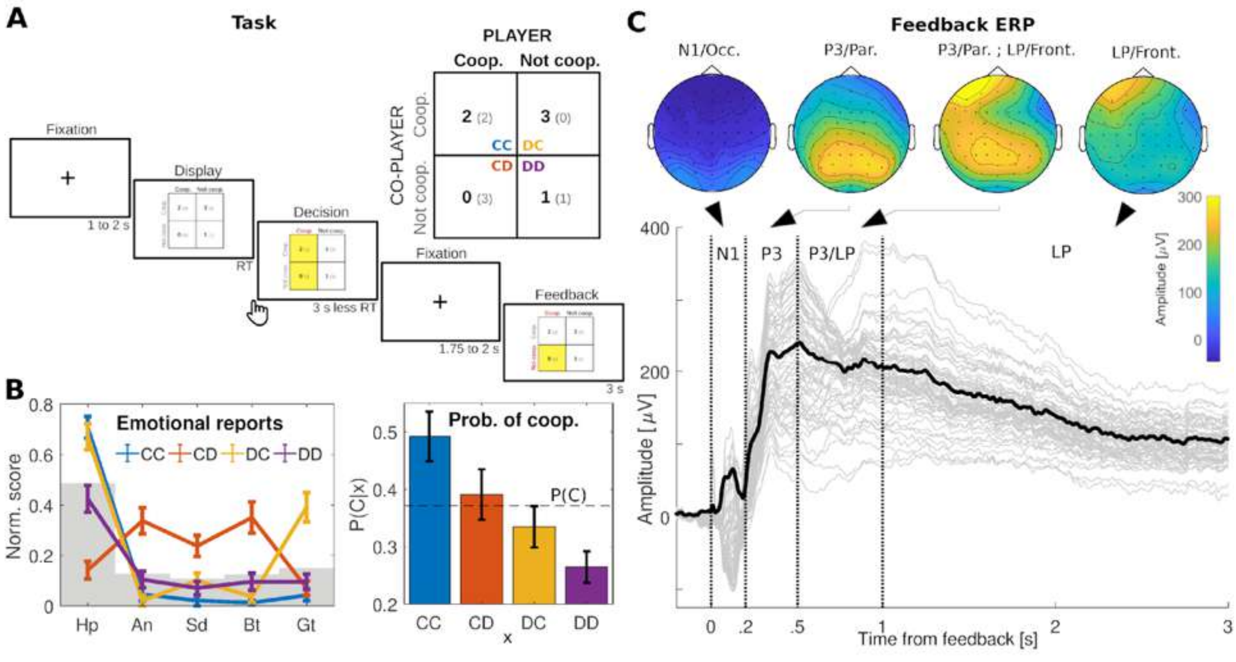


Figure 1. iPD presentation and associated emotional, behavioral and feedback signals. (A) Visual display of the task. The game matrix (inset) assigns earnings to both players as a joint function of their decisions to cooperate with each other. Resulting combinations apportion different amounts to player vs. co-player earnings (indicated in large vs small fonts respectively). CC: mutual cooperation; DD: mutual non-cooperation; CD: one-sided cooperation/'sucker's payoff'; DC: one-sided non-cooperation). At each trial, the matrix is displayed after a fixation period, prompting for the participant's decision. Given his/her choice, the two possible outcomes are briefly highlighted. After a second wait period, the final outcome is indicated. (B) Left: across participants (N=30), each outcome type leads to a distinct emotional profile (Hp: happiness; An: anger; Sd: sadness; Bt: betrayal; Gt: guilt). Grey level indicates outcome report averages by emotion. Right: Grand average of next-round reaction times at the iPD are outcome-dependent. Responses after mutual cooperation are on average faster than after one-sided non-cooperation. Error bars indicate ±1 SEM. See Supplemental Data for emotional and behavioral analyses. (C) In the iPD, feedback signals begin with occipital N1 activity peaking around 100 ms, followed by a fast-rising parietal P3 component which vanishes until about 1 s. Transition into a frontal, left-lateralized late potential (LP) begins at about 500 ms, and is sustained until the end of the trial. Topographies indicate time averages over the selected windows. Black line indicates root-mean-square over channel data (grey).

four outcomes on 8-level Likert scales in order to address recalled emotional impact of the game setup. Participants were debriefed and one subject reported disbelief of the cover story being excluded from the analyses. Thus, the final sample for analyses included 30 subjects. Participants received a cinema ticket in appreciation for their time. In one subject's EEG recording, data transfer was incomplete for the first half of the session and the remainder was included in the analyses.

Data analysis

EEG preprocessing EEG data were referenced offline to the mastoid average and DC offset was removed. A fourth-order 0.1 – 30 Hz Butterworth filter was applied in the forward and reverse direction. Trials were epoched -0.25 to 3 s re to outcome presentation. To remove signal artifact, single-trial and -channel data were rejected by a variance-based criterion[65] across all channels and epochs (confidence coefficient = 4), resulting in removal of 1.75% of channel-epoch timeseries across the experiment (subject range 0.1% - 13.9%).

Feedback signals' spatial topographies were obtained after averaging epochs across all outcomes and participants. In this grand average (Figure 1C), a sequence of N1, P3 and a frontal left-lateralized late potential ('LP') was identified, approximately corresponding to interval windows 0 – 0.2 s, 0.2 – 0.5 s, and 1 – 2.5 s, respectively. To prevent experimenter bias in sensor selection[27,66], a joint decorrelation spatial filtering procedure[67,68] was applied. The resulting set of three (non-orthogonal) linear combinations maps the original 64-channel dataset into 3 feedback components (Supplementary Figure 1A,B) that, separately, optimize detection of evoked activity within their respective time intervals.

To address the feedback-related negativity (FRN) component, the spatial filtering procedure was applied to the P3 dataset (the 0.2 – 0.5 s interval after outcome onset), which contains the classic FRN window[34,40,69]. Due to relative amplitude difference with respect to the P3 in the ERP grand average, spatial filtering was applied to extract reproducible activity representing the FRN in terms of a difference between two conditions in repeated trials[68]. We estimated the subspace of the P3 dataset which maximized the difference between specific mutual cooperation (CC) and unreciprocated cooperation (CD) outcomes. As seen in emotional data, each condition represented the most and least emotionally rewarding outcomes in the iPD from the player perspective, respectively. CC and CD trial waveforms were projected onto the resulting spatial filter.

Feedback signal statistical analysis Data from participants attaining ≥ 10 trials per outcome type were included in the analysis (n=26). N1, FRN, P3, and LP feedback signals were window-averaged according to the corresponding epoch in which the underlying evoked component was observed to be present, as indicated above. General feedback components (i.e. N1, P3, and LP) were analyzed by the separate decisions that the player and the co-player made leading to the presented outcome. As with behavioral data, averages were submitted to a 2-way repeated measures Analysis of Variance (ANOVA), with Player and Co-player as within-subject factors. FRN differences between CC and CD outcomes were analyzed by a Student's t-test.

Spectrotemporal analysis Evoked power from iPD feedback conditions and components was obtained from spectrotemporal correlograms over a log-spaced frequency step range of 1–25 Hz (200 spectral bins with power-5 spacing), via the continuous Morlet wavelet transform, log-transformed to dB-scale and referenced to average baseline values. This corresponds to the event-related spectral perturbation[70,71] estimate of the ERP. For each feedback signal, corresponding time-frequency maps were analyzed within windows as indicated above. Spectrotemporal data were transformed to spectrotemporal t-map difference contrasts between outcome condition pairs, and resulting t-clusters above an a priori t=2 threshold were submitted to nonparametric statistical testing[72,73] with N=1024 resamplings. To avoid systematic baseline differences by player's foreknowledge of his/her executed choice, contrasts only involved conditions with a common participant action ('CC' versus 'CD', 'DC' versus 'DD'). After significant clusters were identified, spectrotemporal measures were estimated by double integration of spectrotemporal power across the time and frequency boundaries given by the cluster. Spectral domain integration was performed in linear scale. All statistical analyses were performed with MATLAB® (The Mathworks, Natick, MA) except where indicated.

Next-decision analyses

Feedback timeseries and spectrotemporal cluster data In this stage of the study we focused on feedback timeseries and/or spectrotemporal cluster results that reflected modulations by outcome processing. FRN, P3, LP and LP-delta feedback signals (see Results) were tested for the hypothesis that they may relate to the participant's decision at the next round. For each signal (e.g. FRN) and outcome type (e.g. CD), data were partitioned by whether at the next round the player would cooperate or not (e.g. 'C|CD' versus 'D|CD', respectively) in order to reveal strategic social decision-making differences stemming as early as during outcome processing. As before, data from participants attaining ≥ 10 trials per condition (n=16 for C,D|CC; n=18 for C,D|CD; n=21 for C,D|DC) were analyzed. Data were tested for normality by Shapiro-Wilk tests implemented with SPSS Statistics 24 (IBM Corporation, Armonk NY), and submitted to Student's t, or alternatively to Wilcoxon rank-sum testing as applicable. After these analyses, to discard the possibility that any observed effects may be due to signal-to-noise differences from imbalanced sample sizes (e.g. more CC trials leading to cooperation than not), additional non-parametric randomization tests were performed[74]. For each contrast, e. g. 'C|CC' versus 'D|CC' trials, subject data were randomly shuffled with resampled trial groups maintaining the original imbalance sizes (i.e. resampled C|CC trials with a same size as the empirical sample). In each experiment resampling, shuffling was done independently per subject, and the testing randomization distribution consisted of 2^{15} experiment resamplings. Significance of empirical effect sizes was assessed by one-sided testing of the hypothesis of no difference between next-choice conditions.

EEG and reaction time correlations Feedback signals found to be sensitive to next-choice conditions (P3 and LP-cluster, see Results) were examined at the single-trial level for associations to forthcoming reaction times. To test the strength of the linear association between feedback signal and next reaction time, single trial data (n=26 subjects) were submitted to repeated measures correlation analysis[75] accounting for within-subject non-independence of trial-based measures. Correlation coefficient confidence intervals of the effect size were obtained by 10^5 random resamplings within-subjects with replacement, and implemented with RStudio, version 1.2.1335 (RStudio, Inc., Boston, MA).

Results

Emotional and behavioral results To first address whether iPD outcomes triggered emotions and behaviors as expected, a 2-way repeated measures ANOVA on emotional ratings identified significant main effects for emotion type ($F(4,116)=68.1$; $p<0.001$; $\eta_p^2=0.701$), outcome type ($F(3,87)=9.8$; $p<0.001$; $\eta_p^2=0.252$), and a significant emotion by outcome interaction ($F(12,348)=29.6$; $p<0.001$; $\eta_p^2=0.505$). Each outcome type was associated with specific emotional reactions (Figure 1B), consistent with previous work[5,15]. CC outcomes were associated with the emotion of happiness; CD with anger, sadness and betrayal; DC with satisfaction and guilt; and DD with intermediate levels of all emotions.

Behaviorally, reaction times following each outcome type were analyzed. A 2 x 2 factorial repeated measures ANOVA with Player and Co-player choices as independent variables showed a significant main effect of Player choice on subsequent reaction time ($F(1,28)=5.51$; $p=0.026$; $\eta_p^2=0.164$) (Figure 1C), and no significant effect by Co-player choice ($F(1,28)=1.68$; $p=0.21$; $\eta_p^2=0.06$). The main effect was qualified by a significant interaction between Player and Co-player choices ($F(1,28)=4.32$; p=0.047; $\eta_p^2=0.13$). Post-hoc comparisons indicated that reaction times were on average 67 ms faster (CI: 17 to 118 ms; Cohen's d=0.345) after CC than DC outcomes ($t(29)=2.74$; $p=0.010$). By contrast, CD versus DD reaction times did not significantly differ ($t(28)=0.06$; $p=0.95$). When all trials were pooled and analyzed for the time to choose to cooperate or not, no significant differences were found between the time taken to cooperate (959 ± 203 ms; mean ± SD) and to not cooperate (921 ± 199 ms) ($t(29)=1.26$; p=0.22). Additional analyses showed that the same-sex dyad interaction did not lead to significant differences in reaction time between female and male participants (Supplementary data).

EEG signals of iPD outcome processing The iPD outcomes elicit a sequence of N1, P3 and a left-lateralized frontal late ('LP') potential (Figure 1C). A 2 x 2 factorial repeated measures ANOVA of the N1 window average amplitude with Player and Co-player choices as independent variables, showed no significant main effect of Player or Co-player choice nor an interaction (Figure 2A,B). For P3, a significant main effect of Player choice ($F(1,25)=9.74$; $p=0.005$; $\eta_p^2=0.28$) and a significant main effect of Co-player choice ($F(1,25)=11.94$; $p=0.002$; $\eta_p^2=0.32$) were found, but no significant interaction. Player choice and Co-player choice to cooperate each led to greater P3 window average amplitudes than to not cooperate

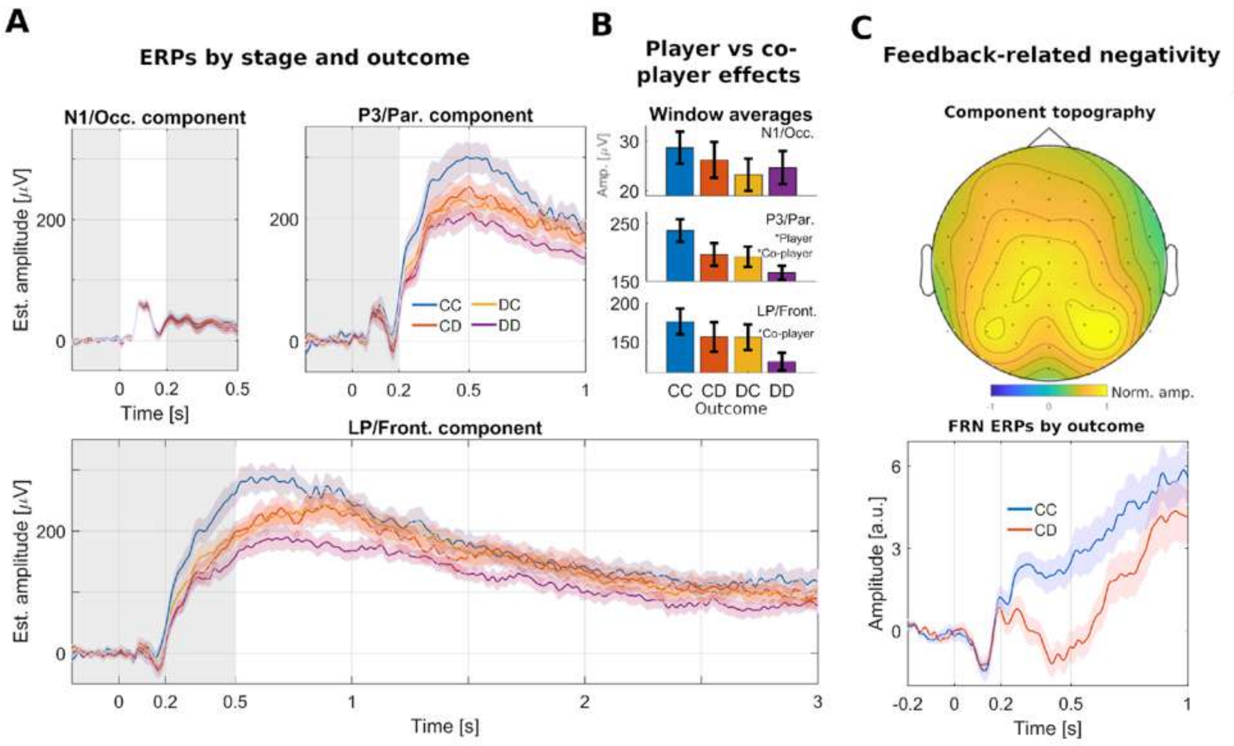


Figure 2. Feedback signals involved in iPD feedback processing. (A) Top left: The visual N1 (0-200 ms; non-shaded region) shows no effects of outcome processing related to players' choices. Top right: The P3 is the first general feedback signal where effects of player and co-player choices are observed. Bottom: The LP shows an effect of co-player's choice only. Curve shades indicate ±1 SEM. (B) Summary of window averages and effects found for all general feedback signals and outcomes. (C) The outcome-specific FRN signal was estimated after reproducible scalp activity which differentiates between CC versus CD outcomes across participants. Top: resulting centro-parietal topography of the FRN. Bottom: FRN timeseries at the 0.2 to 0.5 s window shows outcome sensitivity to co-player's choice given cooperation by the participant.

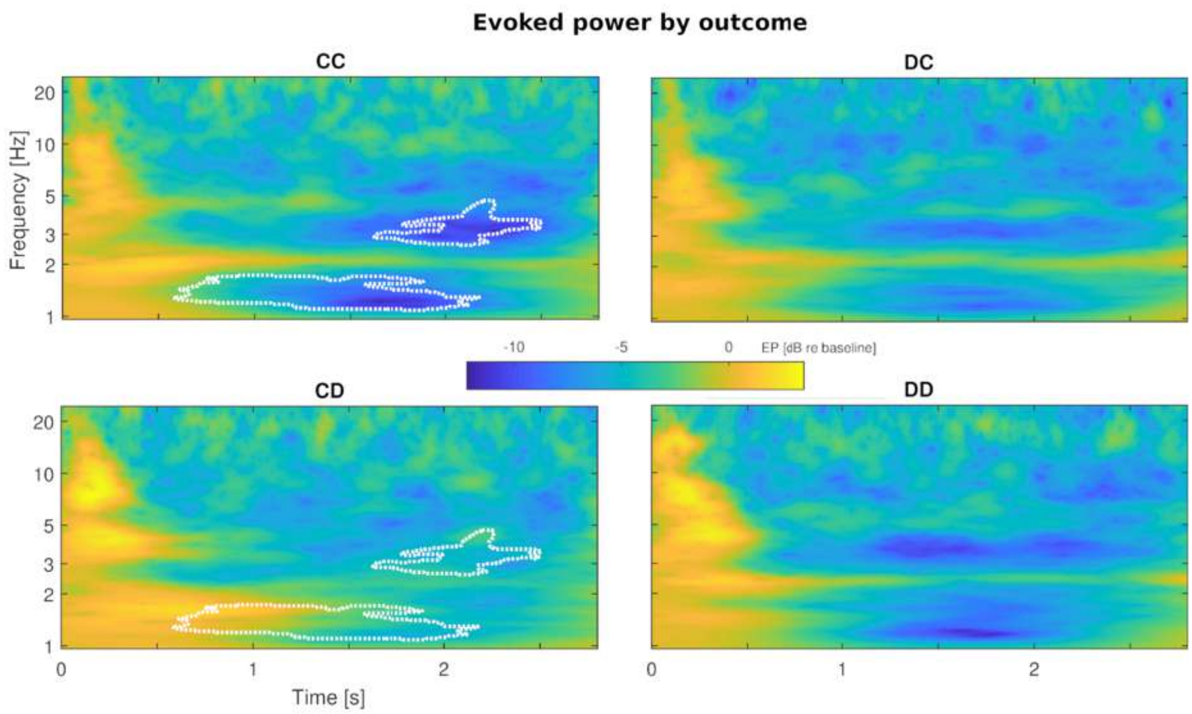


Figure 3. Time-frequency decomposition of the LP feedback signal. (A) Spectrotemporal wavelet correlograms per outcome reveal differential slow-wave (<5 Hz, delta) activity in the 0.5 – 2.5 s interval post feedback onset. Cluster analyses for CC-CD contrasts show significant differences in two time-frequency regions, slow (δ_{low}, 0.5-2 s) and fast (δ_{high}, 1.5-2.5 approx.) delta-band rhythmic activity (p=0.041). Cluster boundaries are superimposed. No clusters were found for DC-DD contrast. (B) Cluster spectral profiles are predominantly contained in the delta band. (C) Effect sizes, defined as participants' evoked power integrated over relevant cluster region boundaries, show significant effects of cooperation by the co-player, indicating relatively greater suppression for CC outcomes.

(Figure 2A,B). LP revealed a significant main effect of Co-player choice ($F(1,25)=4.68$; $p=0.040$; $\eta_p^2=0.16$), but not of Player choice and no significant interaction. As before, cooperation led to greater LP window averages (Figure 2B). In addition, the outcome-specific FRN signal, which showed a centroparietal scalp topography (Figure 2C), revealed a significant difference in its window average for CC (2.0 ± 2.1 ad.u.; mean ± SD) versus CD (-0.2 ± 2.1 ad.u.) outcomes ($t(25)=5.55$; $p<0.001$; Cohen's $d=1.09$). Spectrotemporal evoked power of the FRN, P3 and LP was separately addressed, to test for modulations of these feedback signals not directly reflected on their waveforms. Of these signals, only the LP was significantly modulated by outcome type in evoked spectrotemporal analyses. Slow-wave activity clusters within the delta band (termed 'LP-delta'), showed a significant difference for the CC versus CD mean spectrotemporal power, in the 0.6–2.2 s (low-delta, $p=0.002$), and 1.6–2.5 s (high-delta, $p=0.034$), respectively. In both lower and higher bands, the LP-delta clusters reflected mean evoked power decrease for mutual versus unreciprocated cooperation outcomes (Figure 3). No statistically significant clusters were found in the DC versus DD contrast.

EEG outcome signals indicate future choice and speed of cooperative reciprocation at the iPD Given their role in outcome processing, FRN, P3 and LP waveforms, plus the LP-delta clusters, were further addressed for the hypothesis that they may be early indicators of participants' *next choice* to cooperate. For these analyses, next-trial choices were separated by their leading outcome types (e.g., cooperation given mutual cooperation 'C|CC', or non-coopetation given mutual cooperation 'D|CC'). If a given feedback signal was found to be indicative of next choice, we were then also interested in whether its trial-by-trial amplitude would also relate to the speed to which a given choice was made. In such case, the time to reach a certain choice (e.g. C|CC) was compared against the amplitude of the relevant feedback signal in the prior trial. For this, repeated measures Pearson correlation coefficients[75] were computed to test the relationship between reaction time and single-trial signal amplitudes.

Mutual cooperation (CC) Following mutual cooperation, no significant modulations in the feedback FRN, P3 or LP waveforms were found, in relation to next-trial choice. For LP-delta, C|CC trials showed reduced activity compared to D|CC trials (Figure 4A, bottom), within the high-delta region determined earlier. For the spectrotemporal cluster shown by non-parametric testing (half-maximum 3.6-4.3 Hz, 1.87-2.23 s; $p=0.034$), LP-delta was relatively more attenuated in C|CC compared to D|CC ($t(15)=2.78$; $p=0.014$; $d=0.70$) (Figure 4B). For the low-delta region, no significant modulation by next choice was found. Single-trial LP-delta activity at CC outcomes was further examined in relation to next reaction time, separately for the C|CC and D|CC choices. The data revealed a significant relationship between single-trial LP-delta activity at CC and next trial reaction time to reciprocate (i.e., C|CC), where relative LP-delta decreases were correlated with faster cooperation ($\rho_{rm}(624)=0.084$, CI [-0.001, 0.161], $p=0.035$; Figure 4C). No significant correlation was found between single-trial LP-delta and reaction time at D|CC choices.

Unreciprocated cooperation / sucker's payoff (CD) At the CD feedback stage, FRN signals were significantly modulated by the choice to cooperate at the next trial ($t(17)=3.53$; $p=0.003$; Cohen's $d=0.83$), where D|CD trials had lower FRN amplitudes than C|CD on average (Figure 4A, top). There was limited evidence that the P3 and LP may be significantly modulated by subsequent choice (P3: $t(17)=2.09$; $p=0.052$; $d=0.49$; LP: $t(17)=1.88$; $p=0.078$; $d=0.44$). No LP-delta power cluster differences were significant. At the trial level, FRN activity at CD outcomes was further examined in relation to speed of choice, separately for the C|CC and D|CC cases, but no significant correlations were found.

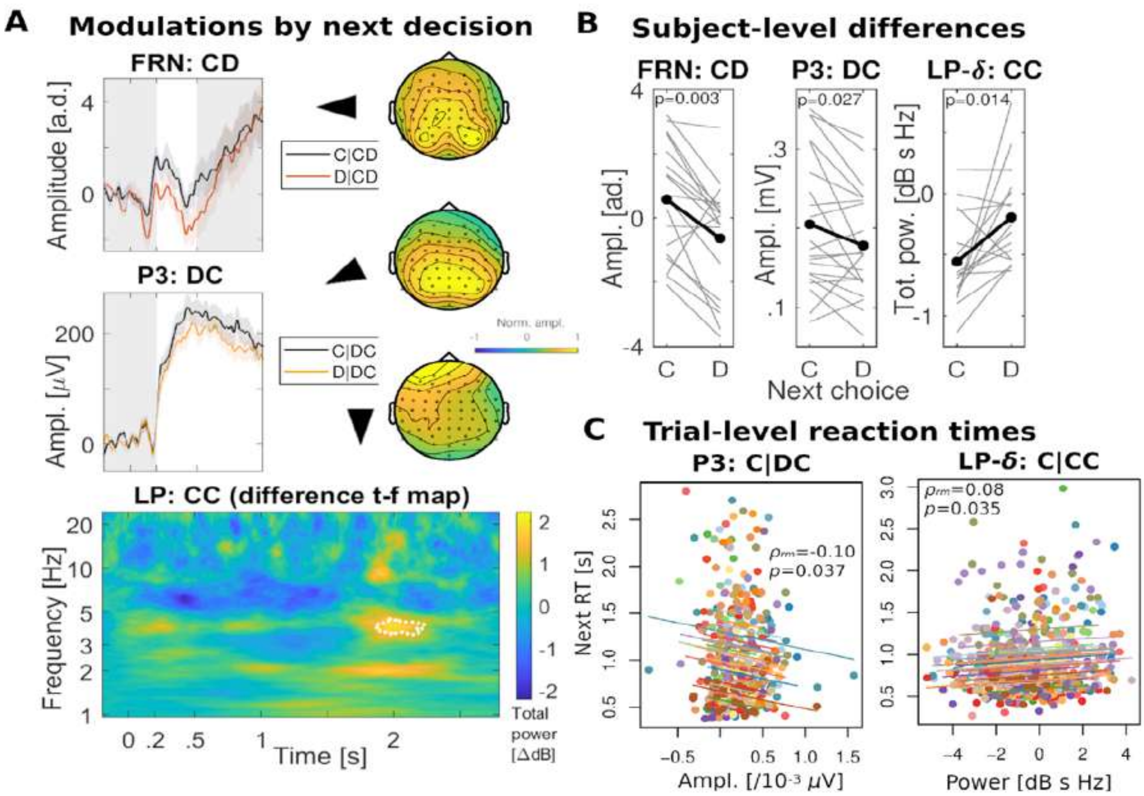


Figure 4. Future round decision-making at the iPD, indicated by scalp activity related to current feedback processing in an outcome-contingent manner. (A) Top: At CD outcomes (sucker's payoff), centro-parietal FRN relates subsequent round decisions. Non-cooperation was indicated by lower-amplitude FRN signals. Middle: For DC (one-sided non-cooperation), higher P3 amplitudes indicate forthcoming cooperation. Bottom: At CC (mutual cooperation), LP- slow-wave activity also relates next round choice. The difference map in the total power correlogram (D|CC minus C|CC) is shown, with the significant cluster boundary superimposed. Topographies for each component are displayed. (B) FRN, P3 and LP-delta subject-level data (in grey) indicate next round decision-making contingent upon CD, DC and CC outcome types, respectively. Grand average values are shown in black. (C) In the specific decision *to reciprocate co-player's cooperation* (i.e. C|CC,DC), trial by trial variations in relevant feedback signals (P3 amplitude for DC, LP-delta suppressions for CC) relate how fast individuals will respond at the next trial.

One-sided non-cooperation (DC) One-sample *t*-tests showed significant mean waveform differences reflecting greater P3 amplitudes at the feedback stage for C|DC than D|DC ($t(20)=2.39$; $p=0.027$; Cohen's $d=0.52$; Figure 4A,B). Trial P3 activity at DC outcomes was further examined in relation to speed of choice, separately for the C|DC and D|DC cases. Single-trial P3 amplitudes involving C|DC choices correlated negatively with the subsequent reaction time, ($\rho_{rm}(422)=-0.101$, CI [-0.196, -0.001], $p=0.037$) (Figure 4C). No significant correlation was found between single-trial P3 and reaction time at D|DC choices.

Mutual defection (DD) No significant mean waveform differences were observed for P3 nor LP signals relating subsequent choice after a DD outcome.

Randomization tests were performed to verify that findings were not a result of systematic bias explainable by sample size imbalances across conditions. Single trial data were randomly shuffled across next decision conditions, per participant, feedback signal, and outcome type (FRN:CD; P3:DC; LP-delta:CC). The observed effects at the FRN ($p=0.018$), P3 ($p=0.026$) and LP-delta ($p=0.002$) regarding modulations by next choice, were found to be significant.

Discussion

This study examined neural processing at the iPD outcome stage in order to provide a more comprehensive picture of relevant feedback signals, while their ability to indicate subsequent decision-making was further investigated. In line with previous studies[25,26], FRN and P3 sensitivity to cooperative choice feedback was observed. In addition, inspection of the previously unaddressed late latency period revealed a frontally-distributed component ('LP'), of which delta-band activity ('LP-delta') was also indicative of cooperative choice feedback. Crucially, the feedback signals were, under specific circumstances, also indicative of choice at the next trial: (i) at unreciprocated cooperation (CD), down-modulations of the FRN related to forthcoming non-cooperation; (ii) at one-sided defection (DC), greater P3 amplitude was associated with player's next cooperation; (iii) at mutual cooperation (CC), LP-delta power reductions similarly related to cooperation at the next round. Single trial P3 activity (at DC rounds) and, separately, LP-delta deactivation (CC) predicted shortening of the time *to cooperate* at the next round. Both scenarios reflect reciprocal reaction, given that the co-player had cooperated. We next discuss the role of each feedback component, and interpret them in relation to relevant social judgments that follow after each associated outcome type. The timing findings may provide a neurally informed account of fast-mode or intuitive cooperative decision-making.

FRN. As expected, the data showed differential processing between mutual and unreciprocated cooperation for the FRN, with decreases for outcomes of unreciprocated cooperation (CD) relative to mutual cooperation (CC). This is consistent with findings from studies using reward, as well as social tasks, where negative stimuli are associated with incremented deflections of the FRN. This component is considered the earliest index in reward outcome evaluation[27,76,77], with sources including the anterior cingulate cortex (ACC)[28,69,76], posterior cingulate cortex (PCC)[29,76,78,79] and basal ganglia[31]. The observed centroparietal distribution appears consistent with PCC activation (cf. [77,79,80]) a region with frontal connectivity required for cognitive and behavioral control[80,81].

Feedback-related activity of the PCC during economic games can persist at subsequent decision-making stages, and predict behavior and strategy change therein[69,78,80,82–84] regarding future action by a co-player[85–87]. Accordingly, in analyses of next decision we found that FRN downmodulations by unreciprocated cooperation (CD) were more prominent if, on the next trial, the participant played non-cooperation. FRN modulations are thought to be underlain by dopaminergic firing pauses[77] and interpreted as a reward prediction error[32,88] or event surprise or saliency[33,78,89,90]. Findings of equal-payoff CD outcomes leading to alternative choices, may raise the question of whether some CD events are experienced as more of a loss than others, either in social and/or emotional terms.

P3. The data featured the P3 as a feedback signal that appeared across outcome types in the iPD, but with relative increases for cooperative choices. These results are in line with findings of P3 increments for desired or positive outcomes at social dilemmas such as the chicken[35,38] and ultimatum games[36,37], however, they contrast with mixed results for the dictator[91] and modified versions of the ultimatum[92] and prisoner's dilemma[25,26]. In a one-shot analysis of feedback signals at the PD game, cooperative choice by a co-player did not show P3 amplitude changes[25]; in an iterated and stochastic version of the game, P3 increases were shown for co-player non-cooperation outcomes[26]. The relative inconsistency across these findings may suggest that magnitude changes in the P3 are sensitive to the relevant payoff and decision-making structure of a task. General frameworks from the perceptual and value-based decision making literature posit P3 change to reflect the accumulation dynamics of relevant feedback evidence, for communication over decision-making networks[93–100]. The P3 is also linked to evidence integration[101–103], event categorization[94], and active working memory dynamics[104–108]. In this line, the variety of effects across studies may be consistent with "temporal filtering" by dedicated attentional and memory resources for outcomes requiring long-term selection for adaptive choice[105,109,110], e. g. increased likelihood of sustaining benefits. Here, participants were set to maximize points as a goal, under a stochastic but stable tit-for-tat strategy (Supplementary Figure XX), and the temporal filtering model would predict that (compared with other outcomes) mutual cooperation events across the session deserve careful realization, since they directly increase the prospects of accumulating reward. This interpretation is consistent with the pattern we observed in our findings. In a naturalistic setting where participants may be continually asked on their belief if an opponent will cooperate[26] (and therefore such assessment acquires especial relevance for performance) one may expect an alternative pattern, where failures to predict cooperation entail maximal P3 change - again consistent with P3 as a feedback signal mediating adaptive choice.

We also found, in next decision analyses, that P3 amplitudes after DC outcomes were more prominent if on the next trial the participant played cooperation. DC outcomes here represent a social context after which decisions require estimation of the co-player's likelihood to reciprocate after having suffered a loss. In the temporal filtering model perspective, the results may suggest an adaptive dissociation between the processing of certain DC events versus others. All DC scenarios provide with the potential to engage on sustained mutual cooperation and, to that extent, the observed relative differences may indicate dedicated attentional or memory resources where the player chooses to engage cooperatively.

The P3 may integrate contributions from social processing systems, including theory of mind, working memory and planning processes, that are subserved across parietal networks[111–113]. A potential interpretation of the findings, in line with the present results for FRN at CD outcomes, is that P3 change by next decision reflects differential surprise or outcome saliency computations across DC events. For the FRN, this surprise was referenced to saliency experienced by the player herself; a possibility is that in DC outcomes, saliency experimented from the co-player-perspective may be estimated[91]. Finally, the P3 feedback latencies observed for DC scenarios could raise questions on conditions for access to conscious choice[114,115].

LP/LP-delta. We observed a late frontal evoked signal ('LP') that was sensitive to co-player choice. In terms of latency, this component appears consistent with the late positive potential ('LPP'), a feedback signal related to emotional processing. However, on a spatial basis, the LPP typically follows a posterior scalp distribution associated with visual cortical areas[27,54] (see [56] for a more frontal distribution). The LPP often displays a positive bias for negative stimuli or losses in value-based decision making studies[27,54–56], although in emotion studies evidence of negative bias at late latency windows is mixed (cf. [53,116,117]) with converging evidence suggesting arousal indexing from motivational systems[118,119]. The observed LP here, nevertheless, may show overlap with studies that relate slow wave responses of extended duration in relation to sustained attentive processing[53], and to the orienting response in a motivationally salient context[120,121].

Waveform signals with such long-term dynamics as the observed LP may motivate spectrotemporal analyses (e.g.[57]) given their potential to be driven by slow-wave components; indeed, the results showed a so far unobserved suppression of LP-delta band activity for mutual versus unreciprocated cooperation. A central role of delta-band modulations has been proposed in terms of homeostatic and autonomic processing of reward and emotion cortical systems[122,123]. Delta decrements are considered to reflect more relaxed (or less anxiogenic) states[124], positive options or feedback[125,126], while increases may be elicited by arousal[57,123,127] or during highly focused states[128] (but see[129]). Arguably, in the iPD, CC outcomes induce less cognitively-demanding states because they imply a default frame for both players (e.g.[130]). Our findings of delta band suppression during mutual versus unreciprocated cooperation appears consistent with more relaxed states induced by the reciprocal and advantageous outcome.

Moreover, we observed that frontal LP-delta suppressions after mutual cooperation were more pronounced as players later chose to maintain cooperation. As our findings relate to a frontal scalp distribution, the delta-band sources may, in principle, involve areas such as medial and ventral frontal cortex[131–134] known to modulate delta activity[135]. Such frontal areas signal upcoming decisions following feedback presentations[136], and are specialized for planning and control (e.g.[137,138]). They additionally involve systems that represent game option values[139–143] and/or socio-emotional perspective taking[144,145]. Since delta activity changes may be related to self-adjusting mechanisms of homeostasis, it is an open question whether the observed effects in next decision analyses reflect distinct representations of value in homeostatic terms[126], or anticipated cooperation risk/belief (e.g.[26]). Future studies may aim to clarify whether such deactivations represent value signals or updating processes[146].

Caveats and limitations We identify at least three limitations that apply to the present findings, which relate to the interaction setting, the feedback signal separation approach, and the predictability of next choice from the EEG during gameplay. In addition, two potential caveats are considered in the present data interpretation, with regards the functional significance of delta and the involvement of intuitive cooperation.

First, behavior relates to same-sex simulated social interactions without participants seeing the confederate during the game. Different game strategies may be preferred when the opponent is a machine versus a person, whether the latter is visible, and their status (e.g. sex, age, familiarity, etc.). We adopted a fixed interaction algorithm to emulate the confederate, which may favor a tit-for-tat cooperative strategy over the course of the interaction. Possibly, access to social cues during face-to-face gameplay may lead up to substantial differences in the behavior and processing of the social interaction, even under a similar artificial response algorithm. As before, the present results may only reflect social cognition on the basis of expectations set by recent outcome and action processing, without the ability to develop or to violate longer-term partnerships.

Second, feedback signals were investigated by whether they are common to (i.e. appear regardless of) every outcome type of the iPD, in the data-driven joint decorrelation method. The FRN component was included by the same method but extracted under a different criterion, namely that it distinguishes between socially hedonic and aversive CC and CD outcomes respectively. This indicates that the reported components may not be the only primary EEG signals relevant to decision making, e.g. components specific to individual outcome types. Also, the method implies that the results do not represent traditional channel sensors, but optimal linear combinations (spatial filters) therefrom. In the present study, these combinations were estimated across all subjects. Thus, for separate studies using the same approach, a large enough number of subject data may need to be collected before estimating an individual component, especially if using an alternative cap system setting.

Third, we demonstrate that single trial EEG data may be predictive of RT at the next round under specific circumstances, namely that the participant cooperates *and* that they are returning cooperation. We do not provide evidence however of the ability to predict next choice to cooperate or not, on a single trial EEG basis. The results do show that *on the average* EEG signals are modulated by such choices. This separates the conclusion that neural signaling during last round processing may facilitate subsequent cooperative behavior, from that where it facilitates subsequent cooperative choice (see below).

With respect to interpreting the frontal delta results, it is worth noting that they are based on the spectrotemporal profile of the dominant left-lateralized frontal spatial component at the late window. Nevertheless, the results may include non-local rhythmic activ. To our knowledge, both spatial and spectral signals have not been previously reported in relation to the iPD. Therefore our preliminary interpretation of these neural signatures may remain limited by reverse inference specific to the present task. It will be important for future studies to address how is such activity specific to different cooperative and other prosocial behavior tasks.

Finally, it has been noted that intuitive modes of cooperation may arise primarily under emotional induction[151]. Participants did not continually report on their experienced emotions by the round, but only did so retrospectively at the end of the session, for each outcome type. Also, effective times to next-round decision involved intervals of a few seconds. At the level of the choice to cooperate or not, the present behavioral paradigm may be insufficient to specify whether intuitive, deliberative, or a combination of both modes of cooperation are generally involved. We propose that the likelihood of intuitive cooperation increases as it is faster *and* it is undertaken in a more automatic form, with earlier indices of the final decision. In the present case, the action *to reciprocate cooperation,* a positive action from the player's perspective, met both criteria for reduced decision times (on the order of a few hundred of milliseconds), and early neural signaling. Future iPD studies may assess whether it is particular social actions, rather that isolated cooperative choices, which increase the likelihood for intuitive cooperation as they entail specific emotions as drivers.

Conclusion

A comprehensive description of neural signals involved in the iterated prisoner's dilemma is crucial to identify the neural basis of cooperative strategies. We extend previous ERP literature to include long latency but also slow-wave activity among the feedback processes that index present but also future cooperative choices. Each outcome type may entice alternative predictive strategies according to the socially learned situation at present146,148, which was reflected in findings that feedback signals encoding cooperative choices also relate future cooperation, in an outcome-dependent manner. For the specific action of reciprocating mutual cooperation, future expedited executions were associated to delta-band activity that has been associated to homeostatic processing in the literature. The results support the involvement of cognitive systems underlying fast decision-making59,147, relating them to the decision to cooperate in a social dilemma60–62. We suggest a direct involvement of such neural signaling with social heuristics deployment.

Acknowledgements

This work was supported by *Proyectos I+D* from the Comisión Sectorial de Investigación Científica (UdelaR), and *Fondo Santiago Achúgar Díaz* 2018 (Uruguay) to V.B.G. We thank support from the Agencia Nacional de Investigación e Innovación (Uruguay) to F.C.C., and from *Ayudas a la Atracción de Talento Investigador* and *Proyectos I+D para jóvenes investigadores*, Comunidad de Madrid and Universidad Autónoma de Madrid (Spain) [2017-T2/SOC-5569; SI1-PJI-2019-00011]. We also thank funding support from ANII and Programa de Desarrollo de las Ciencias Básicas to A.C. and V.B.G. We also would like to thank the anonymous reviewers of this manuscript for their comments and suggestions.

Data availability statement

The data that support the findings of this study are openly available in Open Science Framework at http://doi.org/10.17605/OSF.IO/G8Z7Y

Disclosure of interest

The authors report no conflict of interest

References

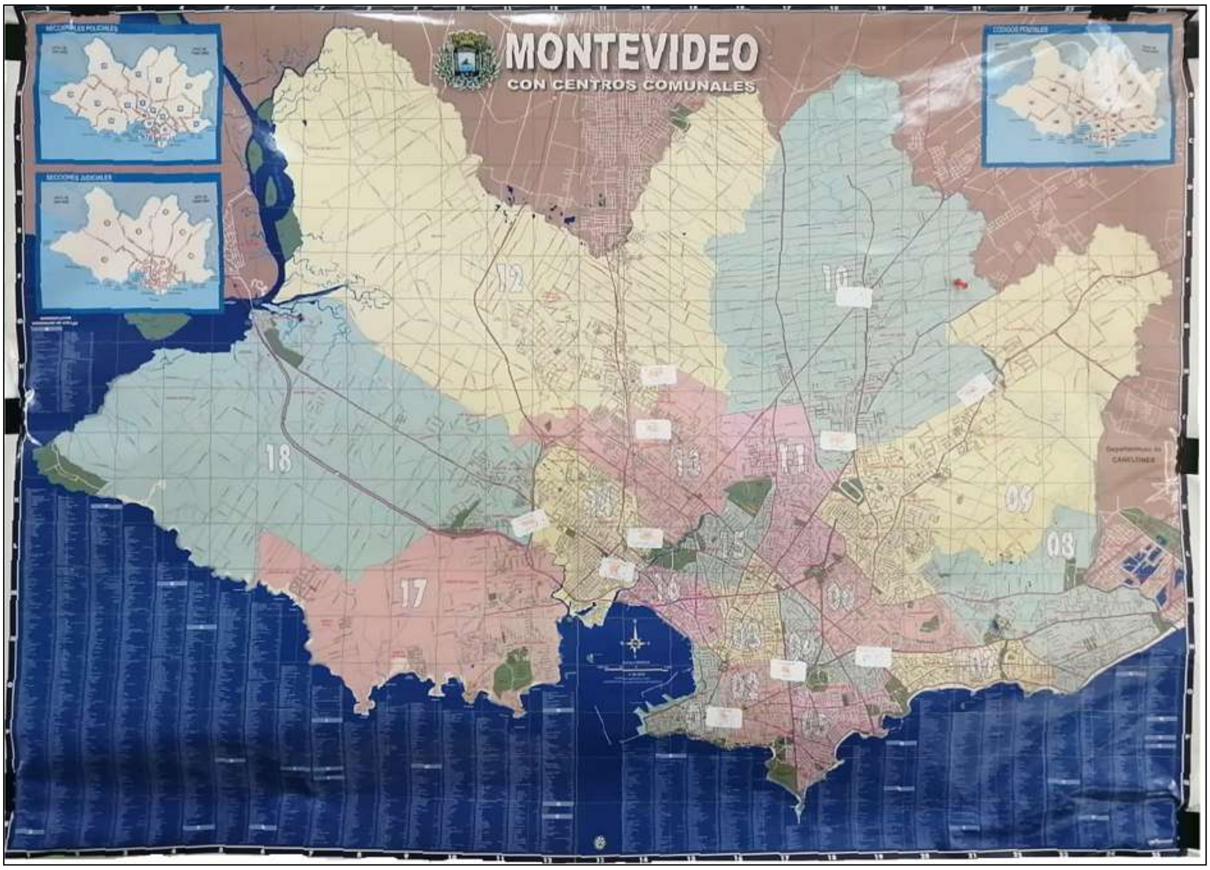
1. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* 211, 1390–1396 (1981).
2. Flood, M. M. Some Experimental Games. *Manag. Sci.* 5, 5–26 (1958).
3. Rapoport, A. & Chammah, A. M. *Prisoner's dilemma: a study in conflict and cooperation*. (University of Michigan Press, 1965).
4. Rilling, J. K. *et al.* The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266 (2008).
5. Gradin, V. B. *et al.* Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. *Psychol. Med.* 46, 1289–1300 (2016).
6. Rilling, J. K. *et al.* A Neural Basis for Social Cooperation. *Neuron* 35, 395–405 (2002).
7. Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J. & Frith, C. D. Brain responses to the acquired moral status of faces. *Neuron* 41, 653–662 (2004).
8. Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* 22, 1694–1703 (2004).
9. Kircher, T. *et al.* Online mentalising investigated with functional MRI. *Neurosci. Lett.* 454, 176–181 (2009).
10. Krach, S. *et al.* Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neurosci.* 10, 9 (2009).
11. Suzuki, S., Niki, K., Fujisaki, S. & Akiyama, E. Neural basis of conditional cooperation. *Soc. Cogn. Affect. Neurosci.* 6, 338–347 (2011).

12. Ramsøy, T. Z., Skov, M., Macoveanu, J., Siebner, H. R. & Fosgaard, T. R. Empathy as a neuropsychological heuristic in social decision-making. *Soc. Neurosci.* 10, 179–191 (2015).
13. Etzel, J. A., Valchev, N., Gazzola, V. & Keysers, C. Is Brain Activity during Action Observation Modulated by the Perceived Fairness of the Actor? *Plos One* 11, e0145350 (2016).
14. Lambert, B., Declerck, C. H., Emonds, G. & Boone, C. Trust as commodity: social value orientation affects the neural substrates of learning to cooperate. *Soc. Cogn. Affect. Neurosci.* 12, 609–617 (2017).
15. Rilling, J. K. *et al.* Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biol. Psychiatry* 61, 1260–1271 (2007).
16. McClure-Tone, E. B. *et al.* Preliminary Findings: Neural Responses to Feedback Regarding Betrayal and Cooperation in Adolescent Anxiety Disorders. *Dev. Neuropsychol.* 36, 453–472 (2011).
17. Rilling, J. K. *et al.* Sex differences in the neural and behavioral response to intranasal oxytocin and vasopressin during human social interaction. *Psychoneuroendocrinology* 39, 237–248 (2014).
18. Feng, C. *et al.* Oxytocin and vasopressin effects on the neural response to social cooperation are modulated by sex in humans. *Brain Imaging Behav.* 9, 754–764 (2015).
19. Schneider-Hassloff, H., Straube, B., Nuscheler, B., Wemken, G. & Kircher, T. Adult attachment style modulates neural responses in a mentalizing task. *Neuroscience* 303, 462–473 (2015).
20. Chen, X. *et al.* Effects of oxytocin and vasopressin on the neural response to unreciprocated cooperation within brain regions involved in stress and anxiety in men and women. *Brain Imaging Behav.* 10, 581–593 (2016).
21. Zheng, H., Kendrick, K. M. & Yu, R. Fear or greed? Oxytocin regulates inter-individual conflict by enhancing fear in men. *Horm. Behav.* 85, 12–18 (2016).
22. Fallani, F. D. V. *et al.* Defecting or Not Defecting: How to "Read" Human Behavior during Cooperative Games by EEG Measurements. *Plos One* 5, e14187 (2010).
23. Jahng, J., Kralik, J. D., Hwang, D.-U. & Jeong, J. Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game. *Neuroimage* 157, 263–274 (2017).
24. Hu, Y. *et al.* Inter-brain synchrony and cooperation context in interactive decision making. *Biol. Psychol.* 133, 54–62 (2018).
25. Bell, R. *et al.* Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology* 53, 216–228 (2016).
26. Zhang, D., Lin, Y., Jing, Y., Feng, C. & Gu, R. The Dynamics of Belief Updating in Human Cooperation: Findings from inter-brain ERP hyperscanning. *NeuroImage* 198, 1–12 (2019).
27. Glazer, J. E., Kelley, N. J., Pornpattananangkul, N., Mittal, V. A. & Nusslock, R. Beyond the FRN: Broadening the time-course of EEG and ERP components implicated in reward processing. *Int. J. Psychophysiol.* 132, 184–202 (2018).
28. Gehring, W. J. & Willoughby, A. R. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282 (2002).
29. Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J. A., Heslenfeld, D. J. & Holroyd, C. B. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *Eur. J. Neurosci.* 21, 3161–3168 (2005).
30. San Martín, R. Event-related potential studies of outcome processing and feedback-guided learning. *Front. Hum. Neurosci.* 6, 304 (2012).
31. Foti, D., Weinberg, A., Bernat, E. M. & Proudfit, G. H. Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 126, 1338–1347 (2015).
32. Sambrook, T. D. & Goslin, J. A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychol. Bull.* 141, 213–235 (2015).
33. Heydari, S. & Holroyd, C. B. Reward positivity: Reward prediction error or salience prediction error? *Psychophysiology* 53, 1185–1192 (2016).
34. Boksem, M. A. S., Kostermans, E. & de Cremer, D. Failing where others have succeeded: Medial Frontal Negativity tracks failure in a social context. *Psychophysiology* 48, 973–979 (2011).
35. Wang, Y. *et al.* Psychophysiological correlates of interpersonal cooperation and aggression. *Biol. Psychol.* 93, 386–391 (2013).
36. Qu, C., Wang, Y. & Huang, Y. Social exclusion modulates fairness consideration in the ultimatum game: an ERP study. *Front. Hum. Neurosci.* 7, 505 (2013).
37. Riepl, K., Mussel, P., Osinsky, R. & Hewig, J. Influences of State and Trait Affect on Behavior, Feedback-Related Negativity, and P3b in the Ultimatum Game. *PloS One* 11, e0146358 (2016).
38. Wang, Y. *et al.* Social value orientation modulates the FRN and P300 in the chicken game. *Biol. Psychol.* 127, 89–98 (2017).
39. Boksem, M. A. S. & De Cremer, D. Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128 (2010).
40. Yeung, N. & Sanfey, A. G. Independent coding of reward magnitude and valence in the human brain. *J. Neurosci. Off. J. Soc. Neurosci.* 24, 6258–6264 (2004).
41. Hajcak, G., Holroyd, C. B., Moser, J. S. & Simons, R. F. Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170 (2005).
42. Sato, A. *et al.* Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411 (2005).
43. Wu, Y. & Zhou, X. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res.* 1286, 114–122 (2009).
44. Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G. & Daum, I. Brain correlates of risky decision-making. *NeuroImage* 49, 1886–1894 (2010).
45. Zhou, Z., Yu, R. & Zhou, X. To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. *Neuropsychologia* 48, 3606–3613 (2010).
46. Wang, L., Zheng, J., Huang, S. & Sun, H. P300 and Decision Making Under Risk and Ambiguity. *Intell Neurosci.* 2015, 1:1–1:1 (2015).
47. Mei, S., Li, Q., Liu, X. & Zheng, Y. Monetary Incentives Modulate Feedback-related Brain Activity. *Sci. Rep.* 8, 11913 (2018).
48. Severo, M. C., Walentowska, W., Moors, A. & Pourtois, G. Goals matter: Amplification of the motivational significance of the feedback when goal impact is increased. *Brain Cogn.* 128, 56–72 (2018).
49. Wu, Y., Leliveld, M. C. & Zhou, X. Social distance modulates recipient's fairness consideration in the dictator game: An ERP study. *Biol. Psychol.* 88, 253–262 (2011).
50. Xu, Q., Yang, S., Huang, Q., Chen, S. & Li, P. A sense of unfairness reduces charitable giving to a third-party: Evidence from behavioral and electrophysiological data. *Neuropsychologia* 142, 107443 (2020).
51. Ravaja, N., Bente, G., Katsyri, J., Salminen, M. & Takala, T. Virtual Character Facial Expressions Influence Human Brain and Facial EMG Activity in a Decision-Making Game. *Ieee Trans. Affect. Comput.* 9, 285–298 (2018).
52. Massi, B. & Luhmann, C. C. Fairness influences early signatures of reward-related neural processing. *Cogn. Affect. Behav. Neurosci.* 15, 768–775 (2015).
53. Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N. & Lang, P. J. Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol. Psychol.* 52, 95–111 (2000).
54. Hajcak, G., MacNamara, A. & Olvet, D. M. Event-Related Potentials, Emotion, and Emotion Regulation: An Integrative Review. *Dev. Neuropsychol.* 35, 129–155 (2010).
55. Pornpattananangkul, N. & Nusslock, R. Motivated to win: Relationship between anticipatory and outcome reward-related neural activity. *Brain Cogn.* 100, 21–40 (2015).
56. Trimber, E. M. & Luhmann, C. C. Implicit predictions of future rewards and their electrophysiological correlates. *Behav. Brain Res.* 333, 184–191 (2017).
57. Güntekin, B. & Başar, E. A review of brain oscillations in perception of faces and emotional pictures. *Neuropsychologia* 58, 33–51 (2014).
58. Kahneman, D. A perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* 58, 697–720 (2003).
59. Kahneman, D. *Thinking, Fast and Slow*. (Penguin Books, 2012).
60. Rand, D. G. Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. *Psychol. Sci.* 27, 1192–1206 (2016).
61. Jagau, S. & van Veelen, M. A general evolutionary framework for the role of intuition and deliberation in cooperation. *Nat. Hum. Behav.* 1, 0152 (2017).
62. Bear, A. & Rand, D. G. Intuition, deliberation, and the evolution of cooperation. *Proc. Natl. Acad. Sci.* 113, 936–941 (2016).

63. Rand, D. G. *et al.* Social heuristics shape intuitive cooperation. *Nat. Commun.* 5, 1–12 (2014).
64. Peirce, J. W. PsychoPy—Psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13 (2007).
65. Junghöfer, M., Elbert, T., Tucker, D. M. & Rockstroh, B. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology* 37, 523–532 (2000).
66. Cohen, M. X. Comparison of linear spatial filters for identifying oscillatory activity in multichannel data. *J. Neurosci. Methods* 278, 1–12 (2017).
67. de Cheveigné, A. & Simon, J. Z. Denoising based on spatial filtering. *J. Neurosci. Methods* 171, 331–339 (2008).
68. de Cheveigné, A. & Parra, L. C. Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage* 98, 487–505 (2014).
69. Walsh, M. M. & Anderson, J. R. Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci. Biobehav. Rev.* 36, 1870–1884 (2012).
70. Makeig, S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293 (1993).
71. Makeig, S., Debener, S., Onton, J. & Delorme, A. Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210 (2004).
72. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190 (2007).
73. Maris, E. Statistical testing in electrophysiological studies. *Psychophysiology* 49, 549–565 (2012).
74. Edgington, E. & Onghena, P. *Randomization Tests*. (Chapman & Hall/CRC, 2007).
75. Bakdash, J. Z. & Marusich, L. R. Repeated Measures Correlation. *Front. Psychol.* 8, (2017).
76. Hauser, T. U. *et al.* The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *NeuroImage* 84, 159–168 (2014).
77. Gheza, D., Paul, K. & Pourtois, G. Dissociable effects of reward and expectancy during evaluative feedback processing revealed by topographic ERP mapping analysis. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* 132, 213–225 (2018).
78. Cohen, M. X. & Ranganath, C. Reinforcement learning signals predict future decisions. *J. Neurosci. Off. J. Soc. Neurosci.* 27, 371–378 (2007).
79. Liu, X., Hairston, J., Schrier, M. & Fan, J. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 35, 1219–1236 (2011).
80. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* 137, 12–32 (2014).
81. Bzdok, D. *et al.* Subspecialization in the human posterior medial cortex. *NeuroImage* 106, 55–71 (2015).
82. Vogt, B. A. Pain and Emotion Interactions in Subregions of the Cingulate Gyrus. *Nat. Rev. Neurosci.* 6, 533–544 (2005).
83. Hayden, B. Y., Nair, A. C., McCoy, A. N. & Platt, M. L. Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60, 19–25 (2008).
84. Fouragnan, E., Retzler, C., Mullinger, K. & Philiastides, M. G. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nat. Commun.* 6, (2015).
85. Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y. & Platt, M. L. Posterior cingulate cortex: adapting behavior to a changing world. *Trends Cogn. Sci.* 15, 143–151 (2011).
86. Houser, D. & McCabe, K. Chapter 2 - Experimental Economics and Experimental Game Theory. in *Neuroeconomics (Second Edition)* (eds. Glimcher, P. W. & Fehr, E.) 19–34 (Academic Press, 2014).
87. Li, D., Meng, L. & Ma, Q. Who Deserves My Trust? Cue-Elicited Feedback Negativity Tracks Reputation Learning in Repeated Social Interactions. *Front. Hum. Neurosci.* 11, (2017).
88. Warren, C. M. & Holroyd, C. B. The Impact of Deliberative Strategy Dissociates ERP Components Related to Conflict Processing vs. Reinforcement Learning. *Front. Neurosci.* 6, 43 (2012).
89. Talmi, D., Atkinson, R. & El-Deredy, W. The feedback-related negativity signals salience prediction errors, not reward prediction errors. *J. Neurosci. Off. J. Soc. Neurosci.* 33, 8264–8269 (2013).
90. Hird, E. J., El‐Deredy, W., Jones, A. & Talmi, D. Temporal dissociation of salience and prediction error responses to appetitive and aversive taste. *Psychophysiology* 55, e12976 (2018).
91. Sun, L., Tan, P., Cheng, Y., Chen, J. & Qu, C. The effect of altruistic tendency on fairness in third-party punishment. *Front. Psychol.* 6, (2015).
92. Fabre, E. F., Causse, M., Pesciarelli, F. & Cacciari, C. Sex and the money--How gender stereotypes modulate economic decision-making: An ERP study. *Neuropsychologia* 75, 221–232 (2015).
93. Sutton, S., Braren, M., Zubin, J. & John, E. R. Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188 (1965).
94. Kok, A. On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577 (2001).
95. Nieuwenhuis, S., Aston-Jones, G. & Cohen, J. D. Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychol. Bull.* 131, 510–532 (2005).
96. Duncan, C. C. *et al.* Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 120, 1883–1908 (2009).
97. Polich, J. Neuropsychology of P300. in *The Oxford handbook of event-related potential components* 159–188 (Oxford University Press, 2012).
98. Luck, S. J. *An Introduction to the Event-Related Potential Technique*. (MIT Press, 2014).
99. Twomey, D. M., Murphy, P. R., Kelly, S. P. & O'Connell, R. G. The classic P300 encodes a build-to-threshold decision variable. *Eur. J. Neurosci.* 42, 1636–1643 (2015).
100. Warren, C. M., Nieuwenhuis, S. & Donner, T. H. Perceptual choice boosts network stability: effect of neuromodulation? *Trends Cogn. Sci.* 19, 362–364 (2015).
101. McCarthy, G. & Donchin, E. A metric for thought: a comparison of P300 latency and reaction time. *Science* 211, 77–80 (1981).
102. Pisauro, M. A., Fouragnan, E., Retzler, C. & Philiastides, M. G. Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nat. Commun.* 8, 15808 (2017).
103. Wichary, S., Magnuski, M., Oleksy, T. & Brzezicka, A. Neural Signatures of Rational and Heuristic Choice Strategies: A Single Trial ERP Analysis. *Front. Hum. Neurosci.* 11, (2017).
104. Donchin, E. & Coles, M. G. Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 357–427 (1988).
105. Polich, J. Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 118, 2128–2148 (2007).
106. Chase, H. W., Swainson, R., Durham, L., Benham, L. & Cools, R. Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. *J. Cogn. Neurosci.* 23, 936–946 (2010).
107. Feldman, H. & Friston, K. Attention, Uncertainty, and Free-Energy. *Front. Hum. Neurosci.* 4, (2010).
108. Eppinger, B., Walter, M. & Li, S.-C. Electrophysiological correlates reflect the integration of model-based and model-free decision information. *Cogn. Affect. Behav. Neurosci.* 17, 406–421 (2017).
109. Aston-Jones, G. & Cohen, J. D. An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annu. Rev. Neurosci.* 28, 403–450 (2005).
110. Nieuwenhuis, S. Learning, the P3, and the Locus Coeruleus-Norepinephrine System. in *Neural Basis of Motivational and Cognitive Control* (The MIT Press, 2011).
111. Behrens, T. E. J., Hunt, L. T., Woolrich, M. W. & Rushworth, M. F. S. Associative learning of social value. *Nature* 456, 245–249 (2008).
112. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34 (2014).
113. Marek, S. & Dosenbach, N. U. F. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci.* 20, 133–140 (2018).
114. Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci.* 108, 20754–20759 (2011).
115. van den Berg, P. & Wenseleers, T. Uncertainty about social interactions leads to the evolution of social heuristics. *Nat. Commun.* 9, 2151 (2018).
116. Palomba, D., Angrilli, A. & Mini, A. Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *Int. J. Psychophysiol.* 27, 55–67 (1997).
117. Delplanque, S., Lavoie, M. E., Hot, P., Silvert, L. & Sequeira, H. Modulation of cognitive processing by emotional valence studied

through event-related potentials in humans. *Neurosci. Lett.* 356, 1–4 (2004).
118. Schupp, H. *et al.* Brain processes in emotional perception: Motivated attention. *Cogn. Emot.* 18, 593–611 (2004).
119. Hajcak, G. & Olvet, D. M. The persistence of attention to emotion: brain potentials during and after picture presentation. *Emot. Wash. DC* 8, 250–255 (2008).
120. Codispoti, M., Ferrari, V. & Bradley, M. M. Repetitive picture processing: autonomic and cortical correlates. *Brain Res.* 1068, 213–220 (2006).
121. Ferrari, V. *et al.* Novelty and emotion: Pupillary and cortical responses during viewing of natural scenes. *Biol. Psychol.* 113, 75–82 (2016).
122. Knyazev, G. G. Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neurosci. Biobehav. Rev.* 31, 377–395 (2007).
123. Knyazev, G. G. EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci. Biobehav. Rev.* 36, 677–695 (2012).
124. Patron, E., Mennella, R., Messerotti Benvenuti, S. & Thayer, J. F. The frontal cortex is a heart-brake: Reduction in delta oscillations is associated with heart rate deceleration. *NeuroImage* 188, 403–410 (2019).
125. Nácher, V., Ledberg, A., Deco, G. & Romo, R. Coherent delta-band oscillations between cortical areas correlate with decision making. *Proc. Natl. Acad. Sci.* 110, 15085–15090 (2013).
126. van de Vijver, I., van Driel, J., Hillebrand, A. & Cohen, M. X. Interactions between frontal and posterior oscillatory dynamics support adjustment of stimulus processing during reinforcement learning. *NeuroImage* 181, 170–181 (2018).
127. Klados, M. A. *et al.* A framework combining delta Event-Related Oscillations (EROs) and Synchronisation Effects (ERD/ERS) to study emotional processing. *Comput. Intell. Neurosci.* 549419 (2009).
128. Harmony, T. The functional significance of delta oscillations in cognitive processing. *Front. Integr. Neurosci.* 7, (2013).
129. Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M. & Herrmann, C. S. Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* 87, 217–233 (2013).
130. Engel, C. & Rand, D. G. What does "clean" really mean? The implicit framing of decontextualized experiments. *Econ. Lett.* 122, 386–389 (2014).
131. Michel, C. M., Lehmann, D., Henggeler, B. & Brandeis, D. Localization of the sources of EEG delta, theta, alpha and beta frequency bands using the FFT dipole approximation. *Electroencephalogr. Clin. Neurophysiol.* 82, 38–44 (1992).
132. Michel, C. M., Henggeler, B., Brandeis, D. & Lehmann, D. Localization of sources of brain alpha/theta/delta activity and the influence of the mode of spontaneous mentation. *Physiol. Meas.* 14 Suppl 4A, A21-26 (1993).
133. Alper, K. R. *et al.* Correlation of PET and qEEG in normal subjects. *Psychiatry Res.* 146, 271–282 (2006).
134. Mellem, M. S., Wohltjen, S., Gotts, S. J., Ghuman, A. S. & Martin, A. Intrinsic frequency biases and profiles across human cortex. *J. Neurophysiol.* 118, 2853–2864 (2017).
135. Pripfl, J., Tomova, L., Riecansky, I. & Lamm, C. Transcranial Magnetic Stimulation of the Left Dorsolateral Prefrontal Cortex Decreases Cue-induced Nicotine Craving and EEG Delta Power. *Brain Stimulat.* 7, 226–233 (2014).
136. Passecker, J. *et al.* Activity of Prefrontal Neurons Predict Future Choices during Gambling. *Neuron* 101, 152-164.e7 (2019).
137. Dien, J. Looking both ways through time: the Janus model of lateralized cognition. *Brain Cogn.* 67, 292–323 (2008).
138. Tops, M., Quirin, M., Boksem, M. A. S. & Koole, S. L. Large-scale neural networks and the lateralization of motivation and emotion. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* 119, 41–49 (2017).
139. Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704–708 (1999).
140. Arana, F. S. *et al.* Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *J. Neurosci. Off. J. Soc. Neurosci.* 23, 9632–9638 (2003).
141. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223 (2006).
142. Padoa-Schioppa, C. & Assad, J. A. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.* 11, 95–102 (2008).
143. Glimcher, P. Value-Based Decision Making. *Neuroeconomics Decis. Mak. Brain Second Ed.* 373–391 (2013).
144. Ochsner, K. N. *et al.* Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.* 16, 1746–1772 (2004).
145. Bzdok, D. *et al.* Segregation of the human medial prefrontal cortex in social cognition. *Front. Hum. Neurosci.* 7, (2013).
146. Joiner, J., Piva, M., Turrin, C. & Chang, S. W. C. Social learning through prediction error in the brain. *NPJ Sci. Learn.* 2, (2017).
147. Strack, F. & Deutsch, R. The duality of everyday life: Dual-process and dual system models in social psychology. in *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* 891–927 (American Psychological Association, 2015)..
148. Suzuki, S. *et al.* Learning to simulate others' decisions. *Neuron* 74, 1125–1137 (2012).
149. Matthews, G. A. & Tye K. M. Neural mechanisms of social homeostasis. *Ann. N. Y. Acad. Sci.* . 1457, 5–25 (2019).
150. Myrseth, K. O. R. & Wollbrant C. E. Cognitive foundations of cooperation revisited: Commentary on Rand et al., (2012, 2014). *J. Behav. Exp. Econ.* 6, 26–42 (2017).
151. Kvarven, A. *et al*. The intuitive cooperation hypothesis revisited: a meta-analytic examination of effect size and between-study heterogeneity. *Journal of the Economic Science Association.* 6, 26–42 (2020).

Mapa de Montevideo
Chart of Montevideo city



Este mapa de Montevideo se colgó en la oficina de la dirección del **CIBPsi** sobre el mes de septiembre de 2013 para poder visualizar el despliegue territorial que tuvo el primer proyecto de intervención cognitiva en la escuela realizado por varios investigadores del **CIBPsi**. Para ello se utilizaron las primeras 1000 tablets entregadas por el Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea (Plan Ceibal) en las 11 escuelas de Montevideo que están marcadas con papelitos blancos sobre el mapa. La intervención contó con varios equipos de aplicadores que visitaban cada escuela dos veces por semana y proponían una serie de juegos en la tablet especialmente diseñados para la estimulación de las bases cognitivas de la matemática temprana. Los datos de la interacción de cada niño con la tablet llegaban en tiempo real al servidor (ubicado en la Facultad de Psicología) y eran visualizados en un terminal ubicado en la oficina de dirección del **CIBPsi**. El mapa buscaba ayudar a ubicar el origen de los datos que iban llegando en tiempo real.

This map of Montevideo was hung in the **CIBPsi** management office in September 2013 in order to visualize the territorial deployment of the first cognitive intervention project in schools carried out by several **CIBPsi** researchers. For this purpose, the first 1000 tablets delivered by Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea (Plan Ceibal) were used in the 11 schools of Montevideo that are marked with white papers on the map. The intervention involved several teams of applicators who visited each school twice a week and proposed a series of games on the tablet specially designed for the stimulation of the cognitive bases of early mathematics. The data from each child's interaction with the tablet arrived in real time to the server (located at the Faculttad de Psicología) and were displayed on a terminal located in the **CIBPsi** management office. The map was intended to help locate the origin of the data arriving in real time.

Socioeconomic Status Differences in Children's Affective Decision-Making: The Role of Awareness in the Children's Gambling Task

Hernán Delgado[1], Carina Aldecosa[1], Ñeranei Méndez[1], Richard Rodríguez[1], Verónica Nin[1], Sebastián Lipina[2], & Alejandra Carboni3

Hernán Delgado https://orcid.org/0000-0002-5738-1516,Verónica Nin https://orcid.org/0000-0003-3162-1582, Sebastián Lipina https://orcid.org/0000-0001-5939-4073, Alejandra Carboni https://orcid.org/0000-0003-4009-1122

1 **Centro de Investigación Básica en Psicología (CIBPsi)** Facultad de Psicología. Universidad de la República, Uruguay
2. **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República, Uruguay
3 Centro de Educación Médica e Investigaciones Clínicas (CEMIC) "Norberto Quirno", Argentina
4 **Centro de Investigación Básica en Psicología (CIBPsi)** & Instituto de Fundamentos y Métodos en Psicología. Facultad de Psicología. Universidad de la República, Uruguay

Correspondence concerning this article should be addressed to Hernán Delgado, Centro de Investigación Básica en Psicología (CIBPsi) & Instituto de Fundamentos y Métodos en Psicología, Facultad de Psicología, Tristan Narvaja 1674, Montevideo, Uruguay. Contact: **hdelgadovivas@gmail.com**

This is a free version, no exact replication -with the approval of the authors- of the pre-print of the manuscript Delgado, H., Aldecosea, C., Menéndez, Ñ., Rodríguez, R., Nin, V., Lipina, S., & Carboni, A. (2022). Socioeconomic status differences in children's affective decision-making: The role of awareness in the Children's Gambling Task. *Developmental Psychology.*58(9), 1716–1729. **https://doi.org/10.1037/dev0001382**

Abstract

Future-oriented decision-making is an important adaptive behavior. In the present study, we examined whether decision-making varies as a function of socioeconomic status (SES) using the Children's Gambling Task (CGT). We administered the CGT to 227 children (49% female, 48% low SES) between the ages of 5 and 7 years. After completing the CGT, we assessed children's knowledge of the reward/loss contingencies. Data analysis was conducted through multilevel modeling. Fluid intelligence, as measured by the Test of Nonverbal Intelligence, was included as a covariate in the analysis. Overall performance differed between SES groups. Children from middle/high-SES backgrounds learned to choose more from the deck with higher future rewards. In contrast, children in the low-SES group did not act in a full future-oriented manner. No differences were found in the level of explicit understanding of the task reached by the two SES groups. Whereas middle/high-SES children with higher knowledge of the game performed better on the last blocks of the task in comparison with their same-SES peers with no understanding, low-SES children with higher explicit knowledge did not exhibit an improvement in their decision-making strategy in comparison with their same-SES low-awareness counterparts. Fluid intelligence did not predict CGT performance, suggesting that SES differences were not mediated by reasoning capabilities. The finding that children from low-SES families continued exhibiting an immediate reward-oriented strategy despite being aware of deck contingencies fits with (although speculatively) the evolutionary-developmental framework.

Keywords: decision-making; children's gambling task; socioeconomic status; evolutionary psychology; life-history theory

Introduction

Making decisions that will bring long-term benefits is considered an adaptive self-regulatory skill that develops during childhood (Garon & Moore, 2004). Often considered 'affective decision-making', this ability is commonly categorized as a "hot" executive function, involving the integration of emotion and cognition (Zelazo et al., 2010). The notion that future-focused regulatory strategies contribute to social adjustment is well-established in the psychological literature. For example, multiple studies in the field of self-control and executive function —specifically those focused on delayed gratification— have shown that individual differences in the ability to regulate behavior and delay rewards during early stages of development predict important and desirable life outcomes, including academic achievement, socioemotional competence, wealth, and health (Casey et al., 2011; Moffitt et al., 2011). Within this research tradition it is common to find the expressions "shortsightedness", "myopia for the future" or "failure to delay gratification" to describe a here-and-now preference (the preference for immediate rewards over larger future rewards) (Jachimowicz et al., 2017; Haushofer & Fehr, 2014; Mani et al., 2013), a tendency that has contributed to strengthening a view of development in resource-poor contexts based on the notions of deficit and maladaptation. An issue with this deficit-based approach is that it is supported by normative assumptions of human development, underpinned by a research tradition nurtured by middle-class, European American samples (Lerner, 2006). Thus, over time, the adaptive dimension of behavior in resource-rich environmental contexts has tended to overlap with positive and desirable developmental outcomes in terms of health and wellbeing.

Without denying the contributions of deficit-based approaches, it has been claimed, however, that it is necessary to explore eventual mechanisms of adaptation (Frankenhuis & de Weerth, 2013). A key limitation of deficit-based approaches is that sometimes the behaviors that emerge in stressful and poor resource environments can be adaptive and contextually functional in response to ecological challenges (Ellis et al., 2017; Ellis & Del Giudice, 2014, 2019; Mittal et al., 2015; Young et al., 2018), although quite different to those observed in white, Western, middle-class samples. In this sense, the recent integration of evolutionary theory into psychological science in what is known as the life-history (LH) theory, offers an alternative conceptual lens through which to explore findings and interpret the effects of childhood poverty on cognitive development (Ellis et al., 2020; Frankenhuis et al., 2020; Frankenhuis & Nettle, 2020). According to LH strategies, while people growing up in supportive environments develop a tendency to think and act in a future-oriented manner, those who grow up under high-stress conditions develop immediate reward-oriented decision-making strategies (Frankenhuis et al., 2016). However, little research has examined affective decision-making in the context of evolutionary-developmental theory. Furthermore, evolutionary research with samples of children growing up in poverty, and in low- and middle-income countries, is still scarce. The present study, therefore, aimed to address these gaps in the literature.

The evolutionary-developmental framework posits that early adversity or chronic stress does not necessarily impair cognition (Ellis & Del Giudice, 2019; Frankenhuis & Nettle, 2020), but rather an individual's cognition may become developmentally adapted to face those challenges that are relevant in the types of environments in which the person grows up, despite the consequences for health and well-being that these adaptations may entail (Ellis & Del Giudice, 2014). Because those strategies that are successful under certain environmental circumstances may be ineffective in others, it has been proposed that natural selection may favor adaptive developmental plasticity. Specifically, mechanisms that evolved through natural selection increase the capacity of individuals to sense environmental cues and use these to guide the development of the organism's phenotypes in an environment-fitting manner (Ellis & Del Giudice, 2019).

Two key environmental dimensions — harshness and unpredictability — have been proposed to affect adaptive developmental trajectories and life-history strategies (Ellis et al., 2009). Harshness refers to morbidity and mortality caused by factors beyond the control of an individual (e.g., the possibility of dying from a stray bullet wound in the context of a neighborhood gang dispute)

(Ellis et al., 2009). Environmental unpredictability is defined as the rates at which environmental harshness varies over time and space (Ellis et al., 2009). Various factors or life situations can increase environmental unpredictability, such as multiple residential changes, the constant change of household members, or the scarcity and instability of resources. Children growing up in high-stress environments such as poverty are exposed to higher levels of harshness and unpredictability. However, it should be mentioned that poverty is not a one-dimensional circumstance and its nature throughout the world will always be relative (Sheridan & McLaughlin, 2014). Thus, the experiences associated with poverty will differ between Pakistan, France, and Uruguay. Beyond this assertion, stressors linked to both unpredictability and harshness should typically be more frequent in low-SES environments. In fact, SES has been proposed as a marker of early-life environmental harshness (Ellis & Del Giudice, 2019).

Remarkably, it has been suggested that early-life environments characterized by high levels of harshness and unpredictability may favor the development of “fast” strategies (Ellis et al., 2009). A characteristic feature of the cluster belonging to fast strategies is the present orientation (Del Giudice, 2015; Fawcett et al., 2012; Griskevicius et al., 2013; Frankenhuis et al., 2016). At a psychological level, this orientation can be associated with impulsivity, short-term opportunism, and immediate reward-oriented strategies with little regard for long-term consequences (Fenneman & Frankenhuis, 2020). Recent studies based on the life-history framework have found evidence of an association between different forms of environmental harshness and unpredictability and present-oriented behaviors, including poorer scores on inhibition (overriding dominant responses) (Mittal et al., 2015; Fields et al., 2021), enhanced reward-oriented problem-solving (Suor et al., 2017), increased preference for immediate rewards (Humphreys et al., 2015; Sturge-Apple et al., 2016) and risky financial decisions (Griskevicius et al., 2011). It is worth noting that in the field of affective decision-making, the evidence comes mostly from studies that have used the delay gratification task. This task is relatively simple compared to the Children’s Gambling Task (CGT), which has never been used to test hypotheses in the context of LH theory.

The CGT, designed by Kerr and Zelazo (2004), is a well-established experimental paradigm for studying decision-making processes that require a flexible appraisal of the affective significance of stimuli. The CGT requires children to make continuous selections of cards from one of two decks, each of these with different reward to punishment ratios. One of the decks presents high initial rewards, but results in high punishments over time and is therefore disadvantageous in the long run. The other deck has lower initial rewards, but also lower punishments over time, making it advantageous in the long run. A growing body of research using the CGT has shown an improvement in decision-making and future-oriented thinking throughout childhood and adolescence (Crone et al., 2005; Crone & Van der Molen, 2004; Gao et al., 2009; Garon & Moore, 2007a, 2007b; Kerr & Zelazo, 2004). Crucially, many of these studies have been conducted in samples of children from American or European middle/high socioeconomic backgrounds. Although this point may be debatable, the evidence suggests that when the complexity of the CGT is low due to the use of the two-deck version, 5-year-olds exhibit an emerging ability to integrate information about the decks and make decisions in a future-oriented manner (Bunch et al., 2007; Gao et al., 2009; Kerr & Zelazo, 2004; Mata et al., 2013). The complexity of the task is a relevant issue when evaluating decision-making. For example, Garon and Moore (2004) used a four-deck version of the gambling task with 3-, 4-, and 6-years-olds and found no age-related improvements in card selections, with even 6-year-olds performing at chance level.

Concerning the above, several additional studies have investigated the relationship between performance on the gambling task and fluid intelligence in children (Crone & Van der Molen, 2004; Li et al., 2017; Mata et al., 2013) and adult samples (Demaree et al., 2010; Webb et al., 2014). The gambling task requires participants to retain and manipulate the relations between the choices and the outcomes obtained to work out the pattern of reward allocation and maximize long-term gains. Hence, fluid intelligence, that is, the ability to reason and problem solve using novel information (Carpenter et al., 1990), may be related to performance on the task.

To the best of our knowledge, the study by Mata et al. (2013) is the only one that evaluated decision-making in children aged between 3 and 5 years old and from different socioeconomic backgrounds using the CGT. The CGT, as proposed by Kerr and Zelazo (2004), consists of 50 trials divided over 5 blocks (10 trials each). Although they did not find a difference in overall performance between SES groups, they found an effect of SES in the last block of the task. More specifically, the high-SES group showed a preference for the advantageous deck compared to their low-SES counterparts, a result interpreted by the authors as indicating a weaker ability of low-SES children to adjust their performance throughout trials based on information from past results (Mata et al., 2013). It should be noted that regardless of SES, the three-year-olds performance did not differ from chance—a finding consistent with those of a previous study (Kerr & Zelazo, 2004). It has been argued that by the age of 3, children may not have developed the skills necessary to perform well on the CGT (Kerr & Zelazo, 2004). Thus, the inclusion of children at such a young age may have obscured the effect of SES on performance. In addition, Mata et al. (2013) assessed children’s conscious awareness of the reward and loss contingencies at the end of the task. They found that the proportion of

Table 1. Demographic and Socioeconomic Characteristics of Low and Middle/High SES Samples

	Low SES (N = 108)	Middle/high SES (N = 119)
Age in months, M (SD)	74.41 (5.80)	76.07 (7.74)
Sex, n (%)		
Males	61 (56.48)	54 (45.38)
Females	47 (43.52)	65 (54.62)
SLI score, M (SD)	19.96 (6.12)	43.89 (7.30)
Years of schooling of the householder, n (%)		
0–6	44 (40.74)	7 (5.88)
6–12	63 (58.33)	57 (47.9)
12–16	1 (0.93)	23 (19.33)
16	0 (0.00)	24 (20.17)
>16	0 (0.00)	8 (6.72)

Note. SES = Socioeconomic Status; SLI = Socioeconomic Level Index.

Table 2. Model Summary and ANOVA Results for Model G and Model K

Model:	Model G			Model K		
Fixed effects	Block + SES + Age + Block × SES			Block + SES + Aw + Age + Block × SES + Block × Aw + SES × Aw + Block × SES × Aw		
Random effects [a]	Block/Participant			Block/Participant		
	df	F	p	df	F	p
Main effects:						
Block	900	9.47	.00***	892	10.09	.00***
SES	224	7.31	.01*	222	7.96	.00**
Aw	-	-	-	222	8.16	.00**
Age	224	9.01	.00**	222	6.17	.01*
Interaction effects:						
Block × SES	900	3.7	.00**	892	3.90	.00**
Block × Aw	-	-	-	892	3.90	.00**
SES × Aw	-	-	-	222	.00	.96
Block × SES × Aw	-	-	-	892	2.42	.05*

Note. The organization of parameters in the models (as either fixed or random effects) is indicated in the top rows, and the effects (either alone or in an interaction) of each parameter with a fixed effect are evaluated in the bottom rows. Aw = Awareness; SES = Socioeconomic Status. *$p < .05$; **$p < .01$; ***$p < .001$.

[a] The specification Block/Participant means that the effect of Block (i.e., the slope) varies across children.

high-SES children who reached an explicit level of knowledge about the task was higher than that of children who reached that same level of knowledge about the task in the low-SES group. Interestingly, although it was not statistically evaluated, the main difference in the level of awareness according to the SES seems to have been located in the 3-year-old group. Among the 4- and 5-year-olds, the awareness level did not appear to differ, which is striking considering that these groups contributed the most to the observed Block—SES interaction. These observations could be suggestive of SES group differences in the use of this explicit understanding (when acquired) to guide decision-making. In other words, children from low-SES backgrounds might have continued selecting more cards from the risky deck despite having consciously understood the task contingencies. Additional reports with samples composed predominantly of children from middle-class families have shown that awareness of the reward and loss contingencies is associated with CGT performance (Andrews & Moussaumai, 2015; Garon et al., 2015; Garon & Moore, 2007b). Unfortunately, the association between awareness and performance was not evaluated by Mata et al. (2013).

The main goal of this study was to explore whether SES might differentially shape children's affective decision-making strategies towards an environment-fitting orientation that would benefit them within their local ecological conditions. More specifically, we assessed children aged 5 to 7 years old using a two-deck version of the CGT that was designed to achieve a balance between the complexity of the task (to maintain adequate levels of uncertainty) and the possibilities of performing the task at levels beyond chance. We hypothesized that children of middle/high-SES would show a preference for the advantageous deck (future-oriented behavior), particularly in the later blocks, whereas their lower SES peers would maintain relatively elevated and stable levels of risk-seeking strategies (present-oriented behavior). In addition, we explored the relationship between awareness and task performance. Although we did not predict a difference in performance on the awareness test according to SES, we expected to find a differential impact of this factor on the development of the participants' preferences. Specifically, we expected that a higher level of awareness would be associated with improved performance in the middle/high-SES group. However, we did not anticipate that a greater understanding of deck contingencies would promote the development of a preference for safe choices in the low-SES group.

Method

Participants The sample comprised 227 children (112 girls), aged between 5 and 7 years (mean age=75.28 months, SD=6.93). One hundred and nineteen children (52%) were from a middle/high SES background and 108 (48%) were of low SES, according to the Socioeconomic Level Index (SLI) described later in this section. All participants were recruited from nine public urban schools in Montevideo, Uruguay. Schools in Uruguay are classified by the National Administration of Public Education according to a continuous variable (Sociocultural Level, SCL). The SCL is a composite that is constructed from various measures that assess social, cultural, and economic school dimensions (e.g., the SES of the general school population, maternal education for each child in the school) (ANEP, 2016). Although the SCL score of each school is not publicly available, it is possible to access the distribution of the schools by quintiles according to the SCL. Schools located in the lowest quintile serve the poorest population and are generally located in vulnerable neighborhoods, while schools in the highest quintile serve mainly children from households located in wealthy urban neighborhoods. To obtain a representative sample, all participants were recruited from schools either in the first quintile or in the 5th quintile of the SCL distribution.

Procedure Written informed consent was obtained from parents/caregivers of the children. The study was approved by the Research Ethics Committee of the School of Psychology (Universidad de la República) (Protocol Number: 191175, Project Title: Desarrollo de la autorregulación en la primera infancia: efectos interactivos entre la sensibilidad biológica al contexto y el nivel socioeconómico). All methods were carried out in accordance with the Declaration of Helsinki (World Medical Association, 2014). An experimenter tested all the children individually in a quiet room at their schools. The application of the CGT lasted between 20 and 35 minutes on average. Parents or caregivers were required to complete the SLI questionnaire through telephone interviews.

Measures

Socioeconomic Status SES was determined using the Socioeconomic Level Index (SLI), developed by the Uruguayan Economic Research Center (Perera & Cazulo, 2016). The SLI is a widely used measure in

Table 3. Contrasts From Model K Comparing Proportion of Advantageous Choices Minus Proportion of Disadvantageous choices in Each of the Blocks Against the Score of the First Block in Each SES/Awareness Group

Group	Contrast	b	SE	p
Middle/high-SES – Level 2	Block 2–Block 1	.04	.04	1.00
	Block 3–Block 1	.18	.04	.00***
	Block 4–Block 1	.28	.04	.00***
	Block 5–Block 1	.31	.04	.00***
Middle/high-SES – Level 1	Block 2–Block 1	.03	.04	1.00
	Block 3–Block 1	.06	.05	.92
	Block 4–Block 1	.07	.05	.80
	Block 5–Block 1	.03	.06	1.00
Low-SES – Level 2	Block 2–Block 1	.03	.04	1.00
	Block 3–Block 1	.10	.04	.05*
	Block 4–Block 1	.06	.04	.71
	Block 5–Block 1	.09	.05	.19
Low-SES – Level 1	Block 2–Block 1	−.10	.04	.12
	Block 3–Block 1	−.01	.05	1.00
	Block 4–Block 1	−.04	.06	1.00
	Block 5–Block 1	−.03	.06	1.00

Note. SES = Socioeconomic Status. **p* < .05; ***p* < .01; ****p* < .001.

social research and marketing studies in Uruguay since it allows for classifying the population according to their consumption capacity. It is composed of 12 items that measure several sociodemographic and socioeconomic factors: the place of residence, family composition, type of health coverage (public or private) of the family members, number of income earners in the households, and educational level of the household head, presence of elements of comfort, and frequency of maid help in the home. The questionnaire is scored on a scale ranging from 0 to 100. Based on this scale, an index is constructed that allows for ranking the households according to three economic classes: High (51-100), Middle (32-50), and Low (0-31). In Uruguay, children from the highest economic strata usually attend private educational centers. Since the study sample consisted of children who attended public schools, only 19 children belonged to the High-class economic category. For this reason, two SES groups were created from the three socioeconomic levels: low and middle/high. The behavioral data of 12 children were not included in the analysis due to the inability to locate their parents to complete the questionnaire.

Children's Gambling Task We used a computerized two-deck version of the CGT (programmed using PsychoPy; (Peirce, 2007)) adapted from the manual version of Kerr and Zelazo's version (2004). The task was administered using a computer (monitor and keyboard). The CGT included two decks of 60 cards each. Based on Mata et al. (2013), a box simulating the glass cylinder that holds the earnings was placed on the center of the screen and between the decks. However, to orient the children's attention towards gains and losses, a plastic recipient for the points (small tokens) was placed next to the keyboard. The back of one deck was covered with black and white vertical stripes, while the back of the other deck was covered with horizontal stripes. The front of the cards was divided into two halves. In the upper half, gains were reported through the use of happy yellow faces, while in the lower half, angry red faces represented losses. Depending on the earnings/loss ratio, one deck was advantageous over trials while the other was disadvantageous (the pattern of lines on the back of the cards and the position of the decks on the screen were randomized). The number of earnings remained constant across cards in both decks, but the number of losses was variable between cards and decks. Specifically, the advantageous deck always provided gains of one point, accompanied by either no loss or the loss of one point. The disadvantageous deck always provided an earning of two points together with losses of 0, 4, 5, or 6 points. Therefore, consistently selecting from the advantageous deck yielded a net gain in the long-run, whereas selecting from the disadvantageous deck yielded a net loss. The order of the cards in each deck was fixed, as described in Kerr and Zelazo (2004).

Beyond global similarities, the version of the CGT used in the present study differed from the two cited versions (Kerr & Zelazo, 2004; Mata et al., 2013) in the following two ways. First, to increase the power to detect group differences, our version included 60 test trials instead of 50. This feature is important since the differences in task performance tend to become more marked in the last two blocks. Second, the rewards were modified to avoid the use of sweets. Instead, the CGT version used in this study involved the child trying to win as many points as possible (represented by happy faces). The child was told that if he or she accumulated enough points (an amount never specified), the points could be exchanged for a prize from a table containing a variety of prizes varying in size and desirability (e.g., a bubble blower, a magnetic drawing board, or a fidget spinner). However, if they got a small number of points, they could only exchange the points for an unattractive prize (e.g., a pencil or an eraser). This method was developed based on previous research (Wilson et al., 2009) to ensure that children could identify at least one award that was attractive enough to motivate them to earn the most points in the task. Thus, before starting the task, children were led to the prize table and encouraged to select the prize they were going to try to obtain in the game.

The task was administered according to the instructions outlined in Kerr and Zelazo (2004) and Mata et al. (2013). Thus, once the desired prize was selected by the child, the task began with six demonstration trials. In this phase the experimenter selected three cards consecutively from one of the decks, followed by three consecutive cards from the other deck. Each time a card was selected, the experimenter announced the number of points earned and then deposited the points in the plastic container. In addition, the points were added to the cylinder presented between the decks on the computer screen. Similarly, when points were lost, they were removed from the plastic container and returned to the experimenter's box (a process also observed in the virtual container).

Table 4. Contrasts From Model K Comparing Proportion of Advantageous Choices Minus Proportion of Disadvantageous choices in Blocks 1 to 5 Between SES/Awareness Groups

Block	Contrast	b	SE	p
Block 1	Middle/high-SES – Level 2 — Middle/high-SES – Level 1	.01	.05	1.00
	Low-SES– Level 2 — Low-SES – Level 1	−.02	.05	1.00
	Middle/high-SES – Level 2 — Low-SES – Level 2	.05	.04	1.00
	Middle/high-SES – Level 1 — Low-SES – Level 1	−.02	.05	1.00
Block 2	Middle/high-SES – Level 2 — Middle/high-SES – Level 1	.02	.04	1.00
	Low-SES– Level 2 — Low-SES – Level 1	.11	.05	.11
	Middle/high-SES – Level 2 — Low-SES – Level 2	.02	.04	1.00
	Middle/high-SES – Level 1 — Low-SES – Level 1	.11	.05	.11
Block 3	Middle/high-SES – Level 2 — Middle/high-SES – Level 1	.13	.05	.07
	Low-SES– Level 2 — Low-SES – Level 1	.08	.05	.74
	Middle/high-SES – Level 2 — Low-SES – Level 2	.09	.05	.26
	Middle/high-SES – Level 1 — Low-SES – Level 1	.04	.06	1.00
Block 4	Middle/high-SES – Level 2 — Middle/high-SES – Level 1	.22	.06	.00**
	Low-SES– Level 2 — Low-SES – Level 1	.07	.06	1.00
	Middle/high-SES – Level 2 — Low-SES – Level 2	.23	.05	.00***
	Middle/high-SES – Level 1 — Low-SES – Level 1	.08	.06	1.00
Block 5	Middle/high-SES – Level 2 — Middle/high-SES – Level 1	.29	.06	.00***
	Low-SES– Level 2 — Low-SES – Level 1	.10	.07	0.98
	Middle/high-SES – Level 2 — Low-SES – Level 2	.23	.06	.00***
	Middle/high-SES – Level 1 — Low-SES – Level 1	.04	.07	1.00

Note. SES = Socioeconomic Status. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

After the demonstration trials, children were given 15 happy faces to start with. The test trials then started and selections were made across 60 test trials, which were divided into five blocks of twelve card choices (see Fig. S1 in the online supplemental materials for an example of a trial sequence). The test trials were administered in the same way as the demonstration trials. A bonus stage was programmed at the end of the task for those children who did not obtain many points. This stage ensured that all the children achieved enough points to select the prize of their choice.

Awareness Test Based on Garon and Moore (2004), immediately after completion of the task, children were given four awareness questions and received one point for each correct answer. Thus, the total awareness score could range from 0 to 4. The first two questions were related to the advantageous deck. The experimenter asked the children, "Now that we've finished the game, which deck was the best to pick from?" After this, children were asked, "Why do you think this was the best to pick from?" If children chose the advantageous deck for the first question, they would be awarded a point. If children were able to answer the second question indicating that the proportion of happy faces was higher for the advantageous deck, they were awarded two points. Similarly, the last two questions were related to the disadvantageous deck. Children were asked, "Which deck was the worst to pick from?" and then ‘‘Why was this deck the worst to pick from?’’ Again, if children initially chose the disadvantage deck, they were awarded a point. If they could also indicate that there were more losses in this deck, they were awarded two points. According to a previous report (Garon et al., 2006) children were divided into two awareness levels. The Awareness Level 1 group was composed of children who scored 0 to 1 on the awareness test. These children showed no understanding of which deck was advantageous/ disadvantageous or why. Children in Awareness Level 2 were those who scored 2 to 4 on the awareness test. Children in this showed, at the end of the task, a minimum and sufficient degree of understanding to discern which deck was better.

Test of Nonverbal Intelligence A computerized version of the Test of Nonverbal Intelligence (TONI) (Brown et al., 1990) was administered as a measure of participants' nonverbal cognitive abilities (fluid intelligence). The test is designed to be language-free. In this task, children are shown a sequence of abstract figures and must choose an image that completes each puzzle. Patterns become progressively more difficult. In this study, the total number of correctly completed puzzles was the outcome variable. Evidence on the psychometric properties (e.g., test-retest reliability and construct validity) has been reported for the TONI (Brown et al., 1990; Ritter et al., 2011). Because 10 children did not attend school on testing days, the TONI was not applied to them.

Statistical analyses

To test the study hypotheses, we conducted two-level multilevel analyses using the nlme package (Pinheiro et al., 2013) on the R programming platform (Team, 2013). Multilevel modeling is advantageous as it allows for the estimation of fixed effects while simultaneously accounting for the clustered or hierarchical structure of data, namely, within-cluster relations (the random effect).

Sample size determination in studies with data that are hierarchically structured is complex. Recently, Arend and Schäfer (2019) conducted a series of simulation-based multilevel power analyses and derived the minimum number of groups and observations per group (*K/n*) to detect a particular effect (small, medium, or large) with a power of .80 and an alpha of .05. They established that for medium values of intraclass correlation (ICC), 30 groups (with 5 observations per group) should be sufficient to detect a medium-sized level-1 effect. The simulations also show that sample sizes of 125/5 are the minimum required to detect a medium-sized level-2 effect, and 200/9 for cross-level interaction effects. In our model, there were 5 proportion scores per participant (see below) with the participant as a grouping variable. Thus, we were confident that our sample size (*K/n*=227/5) was sufficient to detect conceptually meaningful effects.

In this study, independent variables included Block (a categorical variable with 5 levels by dividing the 60 card selections into five instances of 12 cards each (Kerr & Zelazo, 2004)) as the within-subject variable (Level 1), and Sex, Age in months (centered by calculating the difference in relation to the youngest child in the sample), SES (low-SES vs. middle/high-SES), and Awareness (Level 1 vs. Level 2) as the between-subject variables (Level 2). Treating SES as a categorical variable is a common practice in the studies that assess the consequences of child poverty (see, for example, Sturge-Apple et al., 2016; Chen et al., 2010; Last et al., 2018). Also, the categorization of continuous predictor variables (e.g., SES, age, and awareness) is usual in the studies using gambling tasks (see, for example, Kerr & Zelazo, 2004; Crone & Van der Molen, 2004; Mata et al., 2013; Garon et al., 2006). This approach facilitates group comparisons and data visualization. A sensitivity analysis of coding decisions was repeated treating SES and Awareness as continuous variables. As in Kerr & Zelazo (2004), the primary dependent measure was whether children made an advantageous or disadvantageous choice on each trial. Proportion scores were used to analyze performance across the 5 blocks of 12 trials each. The proportion score of each block was calculated by subtracting the proportion of disadvantageous choices from the proportion of advantageous choices, which yielded difference scores ranging from -1 to 1. A net score above zero implied that children were selecting cards advantageously, and a net score below zero implied a disadvantageous selection.

This work was not pre-registered. The anonymous data set (with proportion scores) has been published on Figshare (Delgado et al., 2019-2020). The raw data that support the findings of this study and the R script will be available from the corresponding author upon reasonable request.

Results

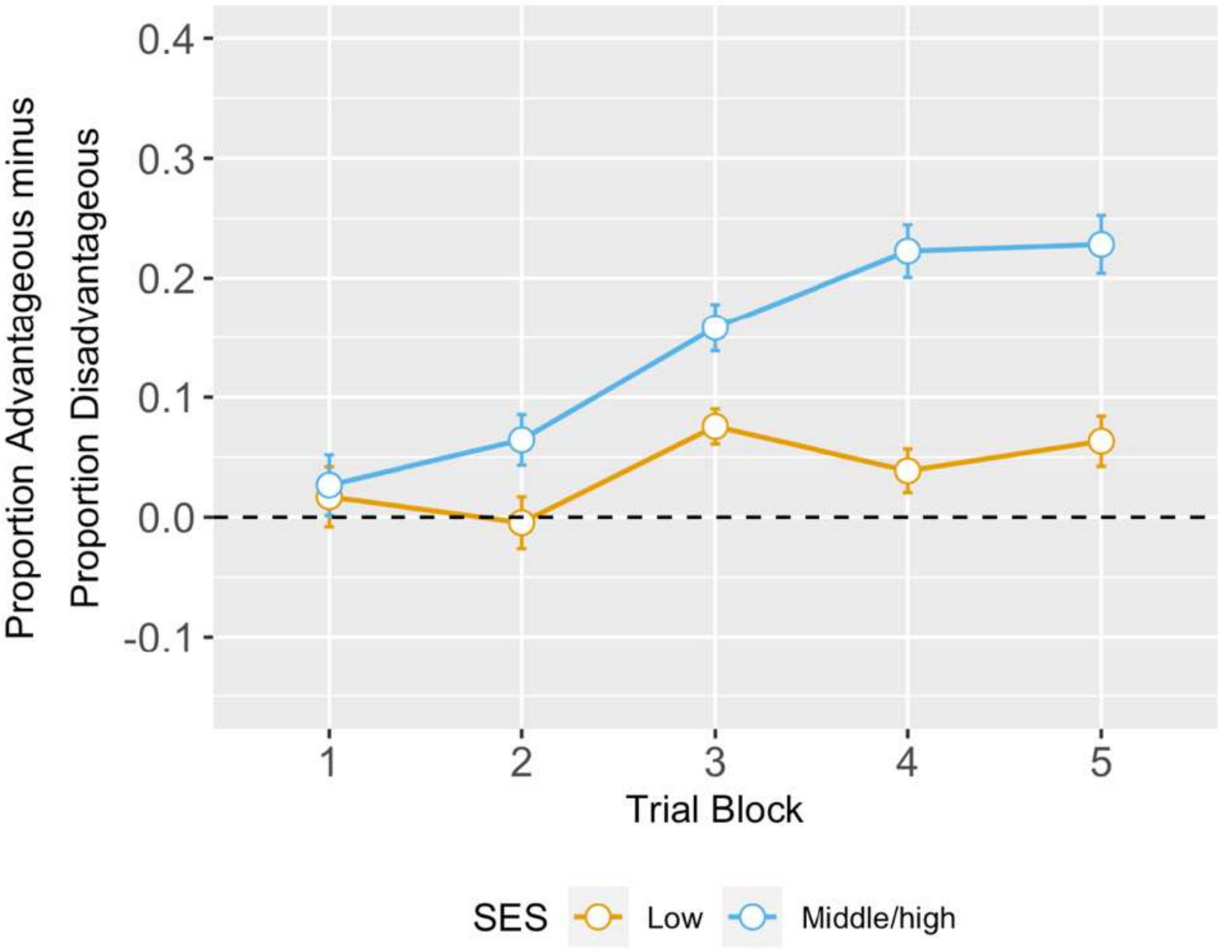
SES Differences in performance on the Children's Gambling Task
The analysis followed a series of successive steps that increased the complexity of the models to select the most parsimonious, that is, the one that considered both the structure of the data collection and our hypothesis. As a first step, we created an unconditional means model (Model A). The main reason to fit this model is that it allowed us to calculate the ICC coefficient (ρ). The ρ value observed in Model A indicated that 38% of the variability in CGT score was between subjects and 62% was within subjects. Therefore, the ICC supported the notion of raising the initial question of whether children with different SES and levels of awareness would show different trajectories of performance. Then, as mentioned previously, we progressively added complexity in terms of fixed and random effects. Because we wanted to develop the most parsimonious model, we tested whether adding more complexity to the model improved the model fit above and beyond the existing terms in the model. Thus, at each step, we compared the fit of the model using AIC and BIC indices and tested the change in the −2log-likelihood to aid decisions about including specific terms. An unconditional growth model (model containing repeated measures) was then fitted and compared with the unconditional means model to assess whether the repeated measures improved the prediction of the dependent variable. Comparing the unconditional growth Model B to the unconditional means Model A greatly improved the model fit, $\chi^2(4) = 59.34$, $p < .001$. This suggests that in the overall sample there was a significant effect of Block, representing an increase in the proportion of advantageous choices as the experimental task progressed. Next, covariates (Age, Sex) were added but only kept if they significantly improved the model fit. While the addition of Age (Model C) significantly improved the fit, $\chi^2(1) = 13.10$, $p < .01$, indicating that being older was associated with better performance, the inclusion of Sex (Model D) did not, $\chi^2(1) =0.01$, $p = .90$. Therefore, only Age was maintained as a covariate in the model. Then, with the aim of allowing participants to have not only different mean scores but also different learning rates (reflected by the development of the decision pattern as the task progressed), a random slope was added to the model (Model E). Adding random slopes to the model resulted in further improvement in the model's fit, $\chi^2(14) = 168.76$, $p < .0001$, which indicates significant variability in the effect of Block across participants. We then incorporated SES (Model F) and its interaction with Block (Model G). These terms were incorporated to test our main hypothesis, that is, that SES influences task performance beyond the effect of Age. The inclusion of SES (Model F) resulted in a better fitting model, $\chi^2(1) = 5.11$, $p < .05$. This indicates that SES accounts for part of the variance in the overall score obtained on the task. Adding the interaction between SES and Block (Model G) also improved the model fit, $\chi^2(4) = 14.48$, $p < .01$, suggesting that SES is associated with performance as blocks progress and children accumulate experience with the results of their choices (Table 2). Contrasts were used to explore the sources of this interaction (with Bonferroni-type adjustment). The first contrasts compared the proportion score of the first block against the proportion score at each of the subsequent levels of the block variable across SES groups. This analysis revealed that scores increased across blocks only for the middle/high-SES group (see Figure 1). More specifically, compared with the first block, children in this group made more advantageous choices in blocks 3 ($b = 0.13$, $p < .001$), 4 ($b = 0.20$, $p < .001$) and 5 ($b = 0.20$, $p < .001$). No comparison yielded significant results for the low-SES group. The second contrast looked for differences between SES groups when conducting separate comparisons for each task block. As can be seen in Figure 1, children from the middle/high-SES group made more advantageous choices than those from the low-SES group on blocks 3 ($b = 0.07$, $p < .05$), 4 ($b = 0.18$, $p < .001$) and 5 ($b = 0.16$, $p < .01$) of the task. To better understand the relation between SES and performance, we graphed the number of advantageous and disadvantageous choices for each SES group (see Fig. S2 in the online supplemental materials). These findings suggest that the middle/high-SES group developed a preference for the advantageous deck during the task. Table S1 provides a summary of the models and ANOVA results for models A to G.

Sensitivity to "Punishment" To assess whether performance differences according to SES were related to the processing of punishment outcomes, a different approach was adopted. This analysis assessed whether the children switched from one deck to the other after receiving punishment. Switch percentage was computed for disadvantageous and advantageous decks separately, according to Crone et al. (2004). To calculate the switch percentage following punishment, we divided the total number of response switches following punishment between the total number of punishments (n switch following punishment/ [n switch following punishment + n stay following punishment]).

A SES × Deck analysis of variance (ANOVA) was conducted on the percentage of switches after punishment. This analysis revealed a main effect of Deck, $F(1, 450) = 28.21$, $p < .001$, indicating that participants switched responses more often following punishment from the disadvantageous deck (84% of the time) than following punishment from the advantageous deck (73% of the time). Further, there was a two-way interaction between SES and Deck, $F(1, 450) = 12.69$, $p < .001$. Tukey's HSD post hoc tests were carried out for pairwise comparisons. For the disadvantageous deck, the SES groups did not differ in the switch percentage after punishment ($p = .20$). However, a difference was found between SES groups in the advantageous deck ($p < .001$): the middle/high-SES group exhibited a significantly lower percentage of switches after punishment (67%) than the low-SES group (78%).

Awareness of the Task An analysis was conducted to assess whether there was a difference between the SES groups in terms of awareness of the game. As mentioned previously, the awareness test was scored out of 4, with 0 indicating no awareness of what was occurring in the task and a score of 4 indicating a full understanding of deck contingencies. The children were then divided according to two awareness levels. Among the middle/high-SES group, 73 of 119 children showed full explicit understanding of deck contingencies, while among the low-SES group, 66 of 108 children exhibited a higher awareness level. The rate of awareness of the game was examined using a 2 (low-SES vs. middle/high-SES) × 2 (Level 1 of awareness vs. Level 2 of awareness) chi-square. The results showed no significant association between SES and Awareness, $\chi^2(1) = 0.001$, $p = 0.97$.

Examining Awareness Differences in CGT Performance according to SES To determine whether awareness is related to performance, we added Awareness (as a variable with two levels; Model H) to Model G. The inclusion of Awareness led to an improvement of the model fit, $\chi^2(1) = 4.64$, $p < .05$, suggesting that awareness influences the overall score obtained. Adding the interaction between Awareness and Block (Model I) also resulted in measurable improvements in goodness of fit, $\chi^2(4) = 15.15$, $p < .01$, indicating that awareness influences the choice



Note. SES = Socioeconomic Status

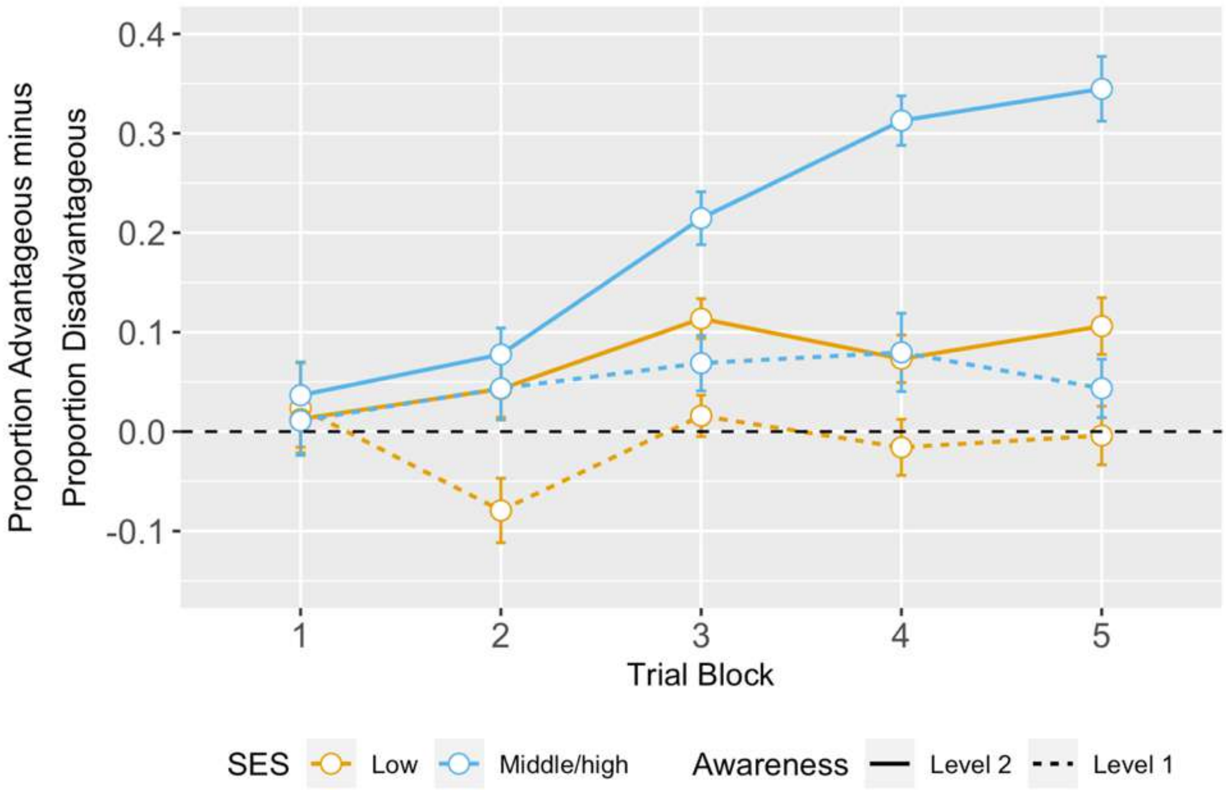
Figure 1. Mean (and Standard Error) Proportion of Advantageous Choices Minus Proportion of Disadvantageous Choices as a Function of SES and Block of 12 Trials

pattern throughout the blocks. In contrast, the inclusion of the interaction between Awareness and SES (Model J) did not have an impact on model fit, $\chi^2(1) = 0.001$, $p = .97$, suggesting no differences in the way in which awareness modulates overall decision-making between the SES groups. Finally, the interaction term of SES × Awareness × Block was added (Model K) to determine whether the effect of Awareness on task performance differed between the SES groups across blocks. Adding this term significantly improved the fit of the model, $\chi^2(4) = 9.64$, $p < .05$ (Table 2). This indicates that the Block × Awareness interaction described previously was different between SES groups (see Figure 2 and Figure S3 in the online supplemental materials). Regarding the random effects, results from our final model showed significant variance in intercepts across children, SD=0.23 (95% CI: 0.05, 1.14), $\chi^2(1) = 226.30$, $p < .0001$. In addition, the slopes varied across children, with the higher variance observed in block 5, SD=0.34 (95% CI: 0.08, 1.43), $\chi^2(14) = 168.76$, $p < .0001$. Contrasts were conducted to explore the three-way interaction observed. The first contrast showed that, compared to the first block, proportion scores in the Level 2 awareness groups increased in blocks 3 ($b = 0.18$, $p < .001$), 4 ($b = 0.28$, $p < .001$) and 5 ($b = 0.31$, $p < .001$) for the middle/high-SES group, and increased in block 3 ($b = 0.10$, $p < .05$) for the low-SES group (Table 3). These findings show that as awareness of gain/loss contingencies increase, performance on the task improves, but only in the middle/high-SES group. In Level 1 awareness groups, no significant difference was observed in the comparison between blocks. Contrasts were then carried out to investigate the pattern of decisions in the five blocks among the four SES/Awareness groups. In blocks 4 and 5, children from the middle/high-SES group who reported awareness of the game chose more cards from the advantageous deck than their same-SES peers who showed no understanding of the task (that is, Level 1 awareness) (block 4, $b = 0.22$, $p < .01$; block 5, $b = 0.29$, $p < .001$) (Table 4). Interestingly, the effect of Awareness was different in the low-SES group. In contrast with their middle-high peers, children from low-SES backgrounds performed similarly on all blocks of the task, independently of their level of awareness. Finally, post-hoc analyses revealed no between-group differences for all scores in individuals who did not show acquisition of explicit understanding. Among those who have a correct understanding, children from the middle/high-SES group chose more cards from the advantageous deck in blocks 4 ($b = 0.23$, $p < .001$) and 5 ($b = 0.23$, $p < .001$) than their low-SES peers (Table 4). Table S2 in the online Supplementary materials provides a detailed summary of Models H to K and their respective ANOVA results. Results from sensitivity analyses regarding SES and Awareness as continuous variables remained unchanged and are presented in Supplementary materials (see Table S3).

Fluid intelligence and CGT To analyze the contribution of fluid intelligence to CGT performance, we added four terms—TONI's score, TONI's score × SES, TONI's score × Block, TONI's score × SES × Block—to Model K. There were no main effects of TONI on CGT performance $F(1, 210) = 0.02$, $p = .88$, and no significant interactions between TONI and SES, $F(1, 218) = 2.8$, $p = .10$, TONI and Block, $F(4, 844) = 1.54$, $p = .19$, or a three-way interaction between the variables, $F(4, 844) = 0.13$, $p = .96$.

Discussion

The present study investigated whether affective decision-making is associated with SES in children aged 5-7 years old. As expected, we



Note. SES = Socioeconomic Status

Figure 2. Mean (and Standard Error) Proportion of Advantageous Choices Minus Proportion of Disadvantageous Choices as a Function of SES, Awareness, and Block of 12 Trials

observed a main effect of SES on overall CGT performance, with middle/high-SES children outperforming their low-SES peers in the number of points achieved in the whole task. In particular, our results showed that children from middle/high-SES backgrounds made more advantageous choices than their low-SES peers in the final blocks of the task. The middle/high-SES children's pattern of decision-making was characterized by an increasing preference for the advantageous deck. In contrast, although their performance improved over trials, children in the low-SES group did not act in a full future-oriented or long-term manner. A second important finding was that greater awareness of which decks were 'good' and 'bad' was a correlate of future-oriented decision-making only for the middle/high-SES group. Specifically, while the middle/high-SES children within the Level 2 awareness group developed a preference for safe choices and performed higher on the CGT than their same-SES peers in the Level 1 group, the performance of low-SES children in the Level 2 group did not differ from that of their same-SES peers in the Level 1 group.

Our findings regarding SES are in line with those of the only existing study by Mata et al. (2013). However, although the latter authors reported a Block × SES interaction, they did not report a main effect of SES. This discrepancy could be due to Mata et al.'s (2013) inclusion of three age groups (i.e., 3, 4, and 5 years of age). The fact that the mean scores of the three-year-old group did not differ from chance (regardless of SES) might have prevented the observation of a main effect of SES. In addition, the version of the CGT employed by Mata and others consisted of 50 trials, while our task consisted of 60, which increased the likelihood of detecting differences. The present results are also in consonance with others who found an association between SES and delay of gratification (Evans, 2003; Evans & English, 2002; Sturge-Apple et al., 2016; Watts et al., 2018). However, some debate over these results prevails, because, for example, neither Noble et al. (2005, 2007) nor Farah et al. (2006) found a significant association between SES and affective decision-making.

The CGT requires the ability to inhibit and reverse prior learning of a reward contingency. For instance, at the beginning of the task, there is a short period in which a preference for the high (and immediately) rewarding deck is reinforced (the disadvantageous deck appears to be advantageous because the first two cards contain only rewards). Because losses do not occur immediately, a predominant response is established in favor of the disadvantageous deck. It has been hypothesized that decision-making requires the inhibition of this dominant response once it is identified that the contingencies of the decks have changed, and, therefore, problems with reverse learning underlie performance on gambling tasks. The evidence supporting this hypothesis comes from studies conducted in adult patients with frontal lobe damage (e.g., Fellows & Farah, 2005). In our study, children in the low-SES group switched decks after receiving punishment with similar frequency to children with middle/high-SES. Interestingly, instead of staying in the safe deck following a choice switch, children in the low-SES group switched back to making risky decisions. Thus, reversal learning *per se* did not underlie the higher number of risky choices made by the low-SES group. Although this finding is inconsistent with the previously cited study (Fellows & Farah, 2005), it is congruent with other research indicating that there is no relation between gambling task performance and reversal learning (Crone et al., 2003, 2004).

There remains a debate about the specific contribution of implicit (emotion-based processes) and explicit (cognitive processes) learning towards successful performance on gambling tasks (Bechara et al., 1997, 2005; Demaree et al., 2010; Maia & McClelland, 2004). It has been suggested that gambling tasks are cognitively penetrable and that inductive reasoning ability (fluid intelligence) is involved in advantageous decision-making (Maia & McClelland, 2004). Although the results from adult samples suggest that fluid intelligence is associated with future-oriented decision-making in gambling tasks (Demaree et al., 2010; Webb et al., 2014), the evidence obtained in children is inconclusive (Crone & Van der Molen, 2004; Li et al., 2017; Mata et al., 2013). Regarding the present study, it should be noted that previous research has shed light on correlations between low-SES and children's performance on batteries of executive function tasks (Lawson et al., 2018), including fluid intelligence (Lipina et al., 2013). Thus, there is good reason to hypothesize that the association between SES and gambling task performance observed in the present study could be explained by asymmetries between both groups in terms of cognitive skills. However, replicating previous results (Crone & Van der Molen, 2004; Mata et al., 2013), our analyses revealed that fluid intelligence, indexed by accuracy on the TONI, was not associated with task performance. Moreover, explicit reasoning skills have been proposed to improve performance on gambling tasks by facilitating conscious access to the relative goodness and badness of the decks (Maia & McClelland, 2004). Indeed, using various strategies, different reports have shown that awareness of the reward and loss contingencies is associated with CGT performance (Andrews & Moussaumai, 2015; Garon et al., 2015; Garon & Moore, 2007b). Consistent with the lack of an association between TONI scores and CGT performance, an important finding of our study was the absence of differences in the level of awareness of deck contingencies according to SES. Taken together, these findings suggest that SES differences in CGT performance do not appear to be mediated by reasoning capabilities.

Associations between poorer performance on self-regulatory tasks have been initially interpreted as reflecting a deficit with potential long-term effects (e.g., Stevens et al., 2009). In addition, poorer performance on self-control tasks has also been associated with long-term impacts on physical health, substance dependence, personal finances, and criminal offending outcomes (Moffitt et al., 2011; Richmond-Rakerd et al., 2021). In this sense, the finding that the low-SES group performed worse on the CGT can also be interpreted as an impairment of low-SES children in 'real-life' decision-making and as a predictor of future maladaptive social functioning (Casey et al., 2011). Implicit in this deficit-based interpretation are two assumptions that are necessary to discuss. The first is that chronic stress associated with poverty could have impaired cognitive function (Mani et al., 2013; Vohs, 2013). As discussed above**,** based on our findings this hypothesis is challenged by the fact that cognitive skills presumed to affect game performance do not appear to be playing a major role in the differential decision-making profiles observed between the SES groups. Second, the associations between early levels of performance and long-term impacts do not consider the multifactorial and interdependent nature of developmental trajectories, as Relational Developmental Systems meta-theories suggest (Lerner, 2018). The most significant result to challenge the deficit interpretation, however, may be the result of the awareness effect on task performance. As described, there was a significant main effect of Awareness on overall task performance, with Level 2 outperforming Level 1. Despite this overall effect, the consequences of being aware of deck contingencies differed markedly between the SES groups. More specifically, children with middle/high-SES in Level 2 performed better on the last two blocks of the task in comparison with their peers in Level 1. This finding suggests that the middle/high-SES children who consciously understood the task contingencies used their knowledge about deck contingencies to improve their decisions. In contrast, while awareness barely improved overall performance in the low-SES group, no differences were found between Level 2 and Level 1 groups across the blocks. These results suggest that low-SES children in Level 2 persisted in making risky choices despite understanding that greater rewards were accompanied by greater punishments.

Our findings are thought-provoking: Why would children who were aware, to some degree, about the gain/loss schedule continue selecting the disadvantageous deck? We can speculate that the current results could be interpreted from an evolutionary LH framework, which proposes that organisms make resource allocation trade-offs during their life course that result from the combination of responses to environmental harshness and unpredictability (Ellis et al., 2009). These ecological factors, which include the socio-cultural developmental contexts, would shape cognitive development by adaptively adjusting LH individual strategies on the slow-fast continuum to match environmental conditions (Ellis & Del Giudice, 2019). Hypothetically, under both harsher and more unpredictable environmental conditions, children may preferentially shift cognition to fast, immediate, reward-oriented decision-making LH strategies (Frankenhuis et al., 2016). Although the identification of measures to evaluate these environmental dimensions constitutes an important theoretical and methodological challenge for the evolutionary-developmental psychology field (Young et al., 2020), there is consensus that lower levels of SES are associated with high levels of harshness and unpredictability (Ellis & Del Giudice, 2019). Therefore, present-oriented strategies would be context-appropriate for individuals living in low-SES ecologies, in which the resources and opportunities are scarce and the future unpredictable. In such conditions, shifting to a here-and-now orientation could be beneficial to understand that waiting to obtain benefits in the long-term instead of taking advantage of immediate opportunities can come at a very high cost if the delayed benefit becomes unavailable due to the occurrence of major and unpredictable environmental changes (Fawcett et al., 2012). Conversely, children who grew up in resource-rich environments would respond to their ecological cues by shifting to a less risky and more long-term oriented behavior, consistent with the development of a slow life history strategy (Ellis et al., 2017). The findings of Sturge-Apple and colleagues (2016) showing that children from low-SES backgrounds and with high vagal tone exhibited a greater preference for immediate reward than their same-SES peers with low vagal tone fit well with this hypothesis. They argued that vagal tone is a biological marker of context-dependent developmentally functional self-regulation. Thus, in a local context where the possibility of future rewards is uncertain, taking the immediate reward can make a lot of sense for children who have maximized their adjustment, even though society as a whole considers this behavior to be maladaptive (Sturge-Apple et al., 2016). More recently, Duran and Grissmer (2020) showed that children from low-income families who exhibited a proclivity toward immediate gratification on a choice delay gratification task had better classroom self-regulation and more-desirable classroom behaviors than their low-income peers who chose to delay gratification. This finding suggests that the preference for immediate rewards over larger future rewards could be an indicator of adaptive development in low-SES ecologies (Duran & Grissmer, 2020). It is worth mentioning that the evolutionary framework does not deny that growing up in stressful conditions can undermine cognitive abilities in the long term. Indeed, the emergence of adaptive strategies during development that allows for coping with present burdens may lead to substantial costs in the future (Ellis & Del Giudice, 2014, 2019).

The result regarding the differential use of explicit knowledge in the SES groups could eventually be interpreted from a deficit perspective. That is, even considering that the gain/loss schedule has been equally understood during the game by both SES groups, differences associated with the ability to keep this information online and use it advantageously to guide decision-making could explain the differences observed. In this regard, it has been proposed that the adult version of the CGT, the Iowa Gambling Task (IGT) (Bechara et al., 1994), can be subdivided into two different stages that measure different types of decision-making (under ambiguity vs. under risk) (Brand et al., 2006). This distinction could be interpreted as closely related to the shift from Level 1 (decisions under ambiguity) to Level 2 (decisions under risk) of awareness in our study. Some studies notably suggest that when explicit understanding about the contingencies of the task is acquired, executive functions highly influence the ability of individuals to make advantageous decisions (Van Duijvenvoorde et al., 2012; Brand et al., 2007). For example, Van Duijvenvoorde et al. (2012) administered two different versions of the IGT to a sample composed of individuals aged between 7 and 29 years. In a standard version of the IGT, no information was provided to the participants regarding the choice properties (*non-informed* condition). In the alternative version (*informed* condition), the gains, losses, and probabilities of all options were explicitly presented. As the authors note, these two conditions allow for distinguishing between ambiguous and risky decision-making scenarios. Notably, Van Duijvenvoorde and coworkers showed that the youngest individuals (between 7 and 11 years old) of the informed group outperformed their peers of the non-informed group in terms of the final number of selections made from advantageous decks. This result suggests that decision-making in the IGT draws (to some extent) upon working memory. Although we cannot completely rule out the influence of working memory on our results, it should be pointed out that the requirements to track and maintain the gains and losses in our study (with a two-deck version) were less demanding than those in Van Duijvenvoorde et al.'s (2012) study (with a four-deck version). In support of this claim, using the two-deck version of the CGT and based on relational complexity theory, Bunch et al. (2007) demonstrated that 42% of 4-year-olds and 83% of 5-year-olds show mastery in the processing of the ternary relations linking three arguments (deck, magnitude of gain, magnitude of loss) required for success on the task. Moreover, several studies have shown that when the complexity of the task is low, 4-year-olds exhibit sufficient ability to succeed, being capable of performing above chance and at a similar level to 5-year-olds (Bunch et al., 2007; Gao et al., 2009; Kerr & Zelazo, 2004; Mata et al., 2013). For example, using Kerr and Zelazo's (2004) two-deck version of the CGT, Mata et al. (2013) and Gao et al. (2009) showed that performance improved significantly between 3 and 4 years but not between 4 and 5 years.

Given that our sample was composed of older children (mean age 6.32 ± 0.6 years), working memory asymmetries are unlikely to have been responsible for the difference in performance. The weakness of this deficit-based interpretation is exacerbated by the absence of an association between fluid intelligence and task performance.

Limitations and future directions Some limitations of the present study should be noted. First, awareness was measured at the end of the task. Due to the correlational nature of our design, it is impossible to determine the actual role that this played during the task. In other words, knowing the "badness" or "goodness" of the decks at the end of the task does not guarantee that this knowledge has guided choices during the game. The assessment of explicit knowledge is problematic. For instance, using a two-deck version of the task similar to the one used here, Garon and Moore (2007b) found that: (a) A high percentage of the 4-year-olds who performed the task exhibited some knowledge of the game when they were evaluated through an awareness test administered in the middle of the task (at the end of the 40th trial); (b) This awareness improved in a second test applied at the end of the task (after the 60th trial); (c) explicit knowledge during the task was associated with task performance. Garon and Moore (2007b) suggested that the simple act of asking the children questions about the characteristics of the decks in terms of gain and losses may have led them to reflect, allowing them to access and use their conscious knowledge to make favorable choices. Although they confirmed their hypothesis in a second experiment (Garon & Moore, 2007b), they failed to find the "scaffolding effect" in a later study (Garon et al., 2015). Future studies could manipulate the time at which the awareness test is given. Assessing the level of awareness in the middle of the task would allow us to determine, for example, whether children from low-SES backgrounds who exhibit explicit knowledge of deck contingencies continue showing risk-taking behavior. In a similar vein, future research on SES differences in CGT performance should include an informed condition so that more direct comparisons can be made regarding the role of explicit knowledge in task performance. Regarding the present study, the significant main effect of awareness on overall task performance suggests that this information was available to some extent and contributed to decision-making during the game.

It is worth noting one final caveat to the interpretation of our findings. Our sample included almost no children from the more economically privileged sectors of Uruguayan society. This point is interesting since the differences observed between the SES groups could have been even greater had we compared affective decision-making in populations at the extremes of the socioeconomic gradient.

The CGT has inherited the theoretical-methodological assumptions of the clinical studies of Damasio and colleagues in patients with lesions in the ventromedial prefrontal cortex (Bechara et al., 1994). By way of illustration, one may note the terminology used to name the decks (e.g., advantageous and disadvantageous decks), so that deficit-based concepts remain embedded in the task's own methodological language. The research tradition in affective decision-making using the CGT still suffers from an underrepresentation of children from low-SES families in its samples. The present study adds to a still scarce literature that aims to reduce this historical gap and seeks to contribute to the characterization of decision-making processes in these populations.

Deficit-based models proposing that early-life stress inevitably impairs cognitive development have dominated the psychological literature. This approach has supported the notion that "future-oriented prudence" (i.e., the ability to delay gratification for a larger future reward) (Thompson et al., 1997) is a "good" skill, adaptive regardless of the ecological context in which development occurs. The results of the present study support conceptualizations drawn from the evolutionary LH framework that challenge the aforementioned prevailing view. On the one hand, the decision-making pattern observed in the middle/high-SES group is consistent with the notion that future-oriented behaviors are adaptive in safe and predictable environments. On the other hand, the present-oriented decision-making pattern observed in the low-SES group is in agreement with the evolutionary–developmental perspective: risky behaviors oriented toward immediate gain would be contextually functional responses to resource-poor conditions rather than deficits.

References

Andrews, G., & Moussaumai, J. (2015). Improving children's affective decision-making in the Children's Gambling Task. *Journal of Experimental Child Psychology*, *139*, 18–34. https://doi.org/10.1016/j.jecp.2015.05.005

ANEP. (2016). Relevamiento de características socioculturales de las escuelas públicas del Consejo de Educación Inicial y Primaria 2015. Uruguay.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological methods*, 24(1), 1.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, (1–3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295. https://doi.org/10.1126/science.275.5304.1293

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159–162. https://doi.org/10.1016/j.tics.2005.02.002

Brand, M., Labudda, K., & Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Networks, 19*(8), 1266-1276.

Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007). Decisions under ambiguity and decisions under risk: correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. *Journal of clinical and experimental neuropsychology, 29*(1), 86-99.

Brown, L. Shebenou, R. J. Johnsen, S. K. (1990). Test of Nonverbal Intelligence: A Language-free measure of cognitive ability, examiner's manual. Second Edition. Austin, Texas: Pro-ed.

Bunch, K. M., Andrews, G., & Halford, G. S. (2007). Complexity effects on the children's gambling task. *Cognitive Development*, *22*(3), 376–383. https://doi.org/10.1016/j.cogdev.2007.01.004

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., Jonides, J., Berman, M. G., Wilson, N. L., Teslovich, T., Glover, G., Zayas, V., Mischel, W., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences, 108*(36), 14998–15003. https://doi.org/10.1073/pnas.1108561108

Chen, E., Cohen, S., & Miller, G. E. (2010). How low socioeconomic status affects 2-year hormonal trajectories in children. *Psychological science, 21*(1), 31-37.

Crone, E. A., Bunge, S. A., Latenstein, H., & Van der Molen, M. W. (2005). Characterization of Children's Decision-making: Sensitivity to Punishment Frequency, Not Task Complexity. *Child Neuropsychology*, *11*(3), 245–263. https://doi.org/10.1080/092970490911261

Crone, E. A., Somsen, R. J., Beek, B. V., & Van der Molen, M. W. (2004). Heart rate and skin conductance analysis of antecedents and consequences of decision-making. *Psychophysiology*, *41*(4), 531–540. https://doi.org/10.1111/j.1469-8986.2004.00197.x

Crone, E. A., & Van der Molen, M. W. (2004). Developmental changes in real life decision-making: Performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Developmental Neuropsychology*, *25*(3), 251–279. https://doi.org/10.1207/s15326942dn2503_2

Crone, E. A., Vendel, I., & Van der Molen, M. W. (2003). Decision-making in disinhibited adolescents and adults: Insensitivity to future consequences or driven by immediate reward? *Personality and Individual Differences*, *35*(7), 1625–1641. https://doi.org/10.1016/S0191-8869(02)00386-0

Delgado, H., Aldecosea, C., Menéndez, Ñ., Rodríguez, R., Nin, V., Lipina, S., & Carboni, A. (2019-2020). [Unpublished data on Socioeconomic Status Differences in Children's Affective Decision-Making(Version 3). Universidad de la República. https://doi.org/10.6084/m9.figshare.15078204.v3

Demaree, H. A., Burns, K. J., & DeDonno, M. A. (2010). Intelligence, but not emotional intelligence, predicts Iowa Gambling Task performance. *Intelligence*, *38*(2), 249–254. https://doi.org/10.1016/j.intell.2009.12.004

Del Giudice, M. (2015). Self-regulation in an evolutionary perspective. In G. Gendolla, M. Tops, & S. Koole (Eds.), *Handbook of biobehavioral approaches to self-regulation* (pp. 25-41). Springer, New York.

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. Neuroscience & Biobehavioral Reviews, 35(7), 1562-1592.

Duran, C. A., & Grissmer, D. W. (2020). Choosing immediate over delayed gratification correlates with better school-related outcomes in a sample of children of color from low-income families. *Developmental Psychology*, *56*(6), 1107–1120. https://doi.org/10.1037/dev0000920

Ellis, B. J., Abrams, L. S., Masten, A. S., Sternberg, R. J., Tottenham, N., & Frankenhuis, W. E. (2020). Hidden talents in harsh environments. *Development and Psychopathology*, 1–19. https://doi.org/10.1017/S0954579420000887

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, *12*(4), 561–587. https://doi.org/10.1177/1745691617693054

Ellis, B. J., & Del Giudice, M. (2014). Beyond allostatic load: Rethinking the role of stress in regulating human development. *Development and Psychopathology*, *26*(1), 1–20. https://doi.org/10.1017/S0954579413000849

Ellis, B. J., & Del Giudice, M. (2019). Developmental adaptation to stress: An evolutionary perspective. *Annual Review of Psychology*, *70*, 111–139. https://doi.org/10.1146/annurev-psych-122216-011732

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk: The impact of harsh versus unpredictable environments on the evolution and development of life history strategies. *Human Nature, 20*(2), 204-268.

Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology*, *39*(5), 924–933. https://doi.org/10.1037/0012-1649.39.5.924

Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Development*, *73*(4), 1238–1248. https://doi.org/10.1111/1467-8624.00469

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain research*, *1110*(1), 166-174. https://doi.org/10.1016/j.brainres.2006.06.072

Fawcett, T. W., McNamara, J. M., & Houston, A. I. (2012). When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behavioural processes, 89*(2), 128-136.

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex*, *15*(1), 58–63. https://doi.org/10.1093/cercor/bhh108

Fenneman, J., & Frankenhuis, W. E. (2020). Is impulsive behavior adaptive in harsh and unpredictable environments? A formal model. *Evolution and Human Behavior, 41*(4), 261-273.

Fields, A., Bloom, P. A., VanTieghem, M., Harmon, C., Choy, T., Camacho, N. L., Gibson, L., Umbach, R., Heleniak, C., & Tottenham, N. (2021). Adaptation in the face of adversity: Decrements and enhancements in children's cognitive control behavior following early caregiving instability. *Developmental science, 24*(6), e13133. https://doi.org/10.1111/desc.13133

Frankenhuis, W. E., & de Weerth, C. (2013). Does early-life exposure to stress shape or impair cognition? *Current Directions in Psychological Science*, *22*(5), 407–412. https://doi.org/10.1177/0963721413484324

Frankenhuis, W. E., & Nettle, D. (2020). The strengths of people in poverty. *Current Directions in Psychological Science*, *29*(1), 16–21. https://doi.org/10.1177/0963721419881154

Frankenhuis, W. E., Panchanathan, K., & Nettle, D. (2016). Cognition in harsh and unpredictable environments. *Current Opinion in Psychology*, *7*, 76–80. https://doi.org/10.1016/j.copsyc.2015.08.011

Frankenhuis, W. E., Young, E. S., & Ellis, B. J. (2020). The hidden talents approach: Theoretical and methodological challenges. *Trends in Cognitive Sciences*, *24*(7), 569–581. https://doi.org/10.1016/j.tics.2020.03.007

Gao, S., Wei, Y., Bai, J., Lin, C., & Li, H. (2009). Young children's affective decision-making in a gambling task: Does difficulty in learning the gain/loss schedule matter? *Cognitive Development*, *24*(2), 183–191. https://doi.org/10.1016/j.cogdev.2008.07.006

Garon, N., Longard, J., Craig, B., & Kent, K. (2015). Loss frequency and awareness predict performance on a preschool variant of the Iowa Gambling Task. *Journal of Cognition and Development*, *16*(2), 286–301. https://doi.org/10.1080/15248372.2013.826662

Garon, N., & Moore, C. (2004). Complex decision-making in early childhood. *Brain and Cognition*, *55*(1), 158–170. https://doi.org/10.1016/S0278-2626(03)00272-0

Garon, N., & Moore, C. (2007a). Developmental and gender differences in future-oriented decision-making during the preschool period. *Child Neuropsychology*, *13*(1), 46–63. https://doi.org/10.1080/09297040600762701

Garon, N., & Moore, C. (2007b). Awareness and symbol use improves future-oriented decision-making in preschoolers. *Developmental Neuropsychology*, *31*(1), 39–59. https://doi.org/10.1080/87565640709336886

Garon, N., Moore, C., & Waschbusch, D. A. (2006). Decision-making in children with ADHD only, ADHD-anxious/depressed, and control children using a child version of the Iowa Gambling Task. *Journal of Attention Disorders*, *9*(4), 607–619. https://doi.org/10.1177/1087054705284501

Griskevicius, V., Ackerman, J. M., Cantú, S. M., Delton, A. W., Robertson, T. E., Simpson, J. A., Thompson, M. E., & Tybur, J. M. (2013). When the economy falters, do people spend or save? Responses to resource scarcity depend on childhood environments. *Psychological Science, 24*(2), 197–205. https://doi.org/10.1177/0956797612451471

Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: a life history theory approach. *Journal of personality and social psychology, 100*(6), 1015.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *Science, 344*(6186), 862-867.

Humphreys, K. L., Lee, S. S., Telzer, E. H., Gabard−Durnam, L. J., Goff, B., Flannery, J., & Tottenham, N. (2015). Exploration—Exploitation strategy is dependent on early experience. *Developmental Psychobiology*, *57*(3), 313–321. https://doi.org/10.1002/dev.21293

Jachimowicz, J. M., Chafik, S., Munrat, S., Prabhu, J. C., & Weber, E. U. (2017). Community trust reduces myopic decisions of low-income individuals. *Proceedings of the National Academy of Sciences, 114*(21), 5401-5406.

Kerr, A., & Zelazo, P. D. (2004). Development of "hot" executive function: The children's gambling task. *Brain and Cognition*, *55*(1), 148–157. https://doi.org/10.1016/S0278-2626(03)00275-6

Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., & Farah, M. J. (2018). Childhood socioeconomic status and executive function in childhood and beyond. *PloS one, 13*(8), e0202964.

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta−analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), e12529. https://doi.org/10.1111/desc.12529

Lerner, R. M. (2006). Developmental science, developmental systems, and contemporary theories of human development. In W. Damon & R. M. Lerner (Series Eds.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol 1. Theoretical models of human development* (6th ed., pp. 1–17). Hoboken, NJ: Wiley.

Lerner, R. M. (2018). *Concepts and theories of human development (4th ed.)*. New York: Routledge.

Li, D., Liu, T., Zhang, X., Wang, M., Wang, D., & Shi, J. (2017). Fluid intelligence, emotional intelligence, and the Iowa Gambling Task in children. *Intelligence*, *62*, 167–174. https://doi.org/10.1016/j.intell.2017.04.004

Lipina, S., Segretin, S., Hermida, J., Prats, L., Fracchia, C., Camelo, J. L., & Colombo, J. (2013). Linking childhood poverty and cognition: Environmental mediators of non−verbal executive control in an Argentine sample. *Developmental Science*, *16*(5), 697–707. https://doi.org/10.1111/desc.12080

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences*, *101*(45), 16075–16080. https://doi.org/10.1073/pnas.0406666101

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(6149), 976–980. https://doi.org/10.1126/science.1238041

Mata, F., Miranda, D. M., Sallum, I., Bechara, A., & Malloy-Diniz, L. F. (2013). Do general intellectual functioning and socioeconomic status account for performance on the Children's Gambling Task? *Frontiers in Neuroscience*, *7*(68), 1–9. https://doi.org/10.3389/fnins.2013.00068

Mittal, C., Griskevicius, V., Simpson, J. A., Sung, S., & Young, E. S. (2015). Cognitive adaptations to stressful environments: When childhood adversity enhances adult executive function. *Journal of Personality and Social Psychology*, *109*(4), 604–621. https://doi.org/10.1037/pspi0000028

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M., Thomson W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science*, *8*(1), 74-87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental science*, *10*(4), 464-480. https://doi.org/10.1111/j.1467-7687.2007.00600.x

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perera, M., & Cazulo, P. (2016). Índice de Nivel Socioeconómico Propuesta de actualización. *CINVE,* Montevideo.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2013). nlme: Linear and nonlinear mixed effects models. *R Package.*

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Richmond-Rakerd, L. S., Caspi, A., Ambler, A., d'Arbeloff, T., de Bruine, M., Elliott, M., Harrington, H., Hogan, S., Houts, R. M., Ireland, D., Keenan, R., Knodt, A. R., Melzer, T. R., Park, S., Poulton, R., Ramrakha, S., Rasmussen, L., Sack, E., Schmidt, A. T., Sison, M. L., … Moffitt, T. E. (2021). Childhood self-control forecasts the pace of midlife aging and preparedness for old age. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3), e2010211118. https://doi.org/10.1073/pnas.2010211118

Ritter, N., Kilinc, E., Navruz, B., & Bae, Y. (2011). Test Review: L. Brown, RJ Sherbenou, & SK Johnsen Test of Nonverbal Intelligence-4 (TONI-4). Austin, TX: PRO-ED, 2010. *Journal of Psychoeducational Assessment, 29*(5), 484-488.

Sheridan, M. A., & McLaughlin, K. A. (2014). Dimensions of early experience and neural development: deprivation and threat. Trends in cognitive sciences, 18(11), 580-585.

Stevens, C., Lauinger, B., & Neville, H. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An event related brain potential study. *Developmental Science*, *12*(4), 634–646. https://doi.org/10.1111/j.1467-7687.2009.00807.x

Suor, J. H., Sturge−Apple, M. L., Davies, P. T., & Cicchetti, D. (2017). A life history approach to delineating how harsh environments and hawk temperament traits differentially shape children's problem−solving skills. *Journal of Child Psychology and Psychiatry, 58*(8), 902-909.

Sturge-Apple, M. L., Suor, J. H., Davies, P. T., Cicchetti, D., Skibo, M. A., & Rogosch, F. A. (2016). Vagal tone and children’s delay of gratification: Differential sensitivity in resource-poor and resource-rich environments. *Psychological Science*, *27*(6), 885–893. https://doi.org/10.1177/0956797616640269

Suor, J. H., Sturge−Apple, M. L., Davies, P. T., & Cicchetti, D. (2017). A life history approach to delineating how harsh environments and hawk temperament traits differentially shape children’s problem−solving skills. *Journal of Child Psychology and Psychiatry*, *58*(8), 902–909. https://doi.org/10.1111/jcpp.12718

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development, 12*(2), 199-212.

Van Duijvenvoorde, A. C., Jansen, B. R., Bredman, J. C., & Huizenga, H. M. (2012). Age-related changes in decision making: Comparing informed and non informed situations. *Developmental Psychology, 48*(1), 192.

Vohs, K. D. (2013). The poor’s poor mental power. *Science*, *341*(6149), 969–970. https://doi.org/10.1126/science.1244172

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159–1177. https://doi.org/10.1177/0956797618761661

Webb, C. A., DelDonno, S., & Killgore, W. D. (2014). The role of cognitive versus emotional intelligence in Iowa Gambling Task performance: What’s emotion got to do with it? *Intelligence*, *44*, 112–119. https://doi.org/10.1016/j.intell.2014.03.008

Wilson, A. C., Lengua, L. J., Tininenko, J., Taylor, A., & Trancik, A. (2009). Physiological profiles during delay of gratification: Associations with emotionality, self-regulation, and adjustment problems. *Journal of Applied Developmental Psychology*, *30*(6), 780–790. https://doi.org/10.1016/j.appdev.2009.05.002

World Medical Association. (2014). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists*, *81*(3), 14–18.

Young, E. S., Griskevicius, V., Simpson, J. A., Waters, T. E., & Mittal, C. (2018). Can an unpredictable childhood environment enhance working memory? Testing the sensitized-specialization hypothesis. *Journal of Personality and Social Psychology*, *114*(6), 891–908. https://doi.org/10.1037/pspi0000124

Young, E. S., Frankenhuis, W. E., & Ellis, B. J. (2020). Theory and measurement of environmental unpredictability. *Evolution and Human Behavior, 41*(6), 550-556.

Zelazo, P. D., Qu, L., & Kesek, A. C. (2010). Hot executive function: Emotion and the development of cognitive control. In S. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 97–111). Washington, DC: American Psychological Association. https://doi.org/10.1037/12059-006

Colegas del CIBPsi instalando el recién adquirido equipo para EEG Biosemi de 64 canales
CIBPsi's colleagues with the recently acquired EEG Biosemi 64 channels equipment

En esta foto, compañeros del **CIBPsi** instalan electrodos del equipo EEG Biosemi de 64 canales para hacer, en 2012, los primeros estudios de electroencefalografía con un aparato de estas características y potencia. El equipo fue adquirido en 2010.

In this photo, colleagues from CIBPsi install electrodes of the Biosemi 64-channel EEG equipment to perform, in 2012, the first electroencephalography studies with a device of these characteristics and power. The equipment was acquired in 2010.

Social Avoidance in Depression: a Study Using a Social Decision-Making Task

Gabriela Fernández-Theoduloz[1], Valentina Paz[2], Eliana Nicolaisen-Sobesky[3], Alfonso Pérez[3], Abraham P. Buunk[4], Álvaro Cabana[2], & Victoria B. Gradin[2]

Abraham P. Buun, Álvaro Cabana https://orcid.org/0000-0002-8637-290X, Gabriela Fernández-Theoduloz https://orcid.org/0000-0002-9225-060X, Victoria B Gradin https://orcid.org/0000-0001-6599-3523, Eliana Nicolaisen-Sobesky https://orcid.org/0000-0002-4806-2863, Valentina Paz https://orcid.org/0000-0002-2401-5472

1 **Centro de Investigación Básica en Psicología (CIBPsi)** & Instituto de Psicología Clínica, Facultad de Psicología, Universidad de la República, Uruguay
2 **Centro de Investigación Básica en Psicología (CIBPsi**) & Instituto de Fundamentos y Métodos en Psicología. Facultad de Psicología. Universidad de la República, Uruguay
3 **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República, Uruguay
4 Department of Psychology. University of Groningen & Netherlands Interdisciplinary Demographic Institute. The Netherlands

Correspondence concerning this article should be addressed to Victoria B. Gradin, Center for Basic Research in Psychology (CIBPsi) & Instituto de Fundamentos y Métodos en Psicología, Facultad de Psicología, Tristan Narvaja 1674, Montevideo, Uruguay. Contact: **victoriagradin@gmail.com**

This is a free version, no exact replication -with the approval of the authors- of the pre-print of the manuscript Fernández-Theoduloz, G., Paz, V., Nicolaisen-Sobesky, E., Pérez, A., Buunk, A. P., Cabana, Á., & Gradin, V. B. (2019). Social avoidance in depression: A study using a social decision-making task. *Journal of Abnormal Psychology*, *128*(3), 234–244. It is available at **https://doi: 10.1037/abn0000415**

Abstract

Depression significantly affects interpersonal functioning. Social avoidance may play an important role in depression, limiting opportunities and social skills acquisition, contributing to the maintenance of social difficulties. In the last years, the need for studying social interactions using interactive tasks has been highlighted. This study investigated social avoidance in unmedicated depressed (n=26) and matched healthy control (n=26) participants, using a novel computerized social decision-making task (the TEAM task). In this task, participants choose between a social option (playing in a team with a co-player) and an individual option (playing alone). While the social option is more profitable from a material point of view, it can also be challenging due to social comparison and guilt feelings for failing the team. It was found that, the higher the rank of the co-player, the stronger the negative emotions (shame, guilt) reported by participants, and the more they opted for the individual option. Depressed participants reported significantly less positive (happiness) and more negative (shame, guilt, disappointment) feelings regarding the task. Importantly, depressed participants chose significantly more often the individual option compared to controls, which led to lower gains in this group. Furthermore, as the task progressed, controls selected the individual option less often, while depressed participants selected the individual option more often. Our findings illustrate the importance of social avoidance in depression and how this behavior can lead to negative consequences. They also highlight the role of social comparison and guilt related processes in underlying social avoidance in depression.

Key words: Depression; social comparison; social avoidance; social decision-making

Introduction

Depression is a prevalent disorder, ranked among the leading causes of disability worldwide (Kessler & Bromet, 2013). Crucially, depression has a profound impact on interpersonal functioning. People with depression report poor participation in social activities, having less support from social networks, poor social skills, difficulties in being assertive, poor intimate relationships, being unconfident and displaying excessive reassurance-seeking, and in general not enjoying but rather suffering through social interactions (Joiner & Timmons, 2009). Classic studies of social cognition in individuals with mental disorders have mostly used self- and observer-based questionnaires to study social interactions (Lis & Kirsch, 2016), as well as facial emotion perception or theory of mind tasks (Kohler, Hoffman, Eastman, Healey, & Moberg, 2011). While these approaches provide useful information, they do not asses active social interactions. In the last few years, there has been growing recognition of the importance of studying social interactions in relation to mental disorders (King-Casas & Chiu, 2012) and more specifically in depression (Gradi*n et* al., 2015, 2016; Kupferberg, Bicks, & Hasler, 2016; Pulcu & Elliott, 2015; Wang, Yang, Li, & Zhou, 2015), using paradigms such as behavioral economic tasks (Hasler, 2012), as well as other interactive tasks (Kupferberg, Hager, et al., 2016; Silk *et* al., 2014), that allow active social exchanges.

It has been hypothesized that avoidance, (Ferster, 1973; Lewinsohn, 1974) and in particular *social* avoidance (defined as the tendency to keep away from social situations (Goossens, 2014)), play a crucial role in depression (Trew, 2011). It has also been suggested that depression is associated with low assertiveness, social withdrawal, avoidance and shyness (Joiner, 2000). While social avoidance of interpersonal conflict may prevent the experience of negative outcomes, it may also result in the loss of social and material opportunities, lead to isolation and prevent the individual from improving social skills and from learning how to deal with interpersonal problems (Joiner, 2000; Joiner & Timmons, 2009; Trew, 2011). In the long run, this may contribute to maintaining social difficulties and depression (Trew, 2011). Recent work (Ottenbreit, Dobson, & Quigley, 2014; Trew, 2011) has noted that, while the construct of avoidance has received considerable attention in the study of anxiety disorders, it has been underemphasized in research on depression. Empirical studies are limited and have mostly used rating scales based on self-report questionnaires (Ottenbreit et al., 2014). This highlights the need for empirical research exploring avoidance in depression by using social decision-making tasks.

Here, we have developed a computerized social decision-making task (the TEAM task) to investigate social avoidance in depression. This task was inspired by real life situations in which we have to choose between doing something as part of a team (social option) or doing something individually (individual option). An example of this kind of situation would be the case of a student who has to decide between doing a course assignment with a partner or individually. While the social option may have advantages (i.e. learning from someone else, sharing the workload) it may also bring some drawbacks. For example, if the student perceives her partner as being brighter than her, this could lead to worries about not being good enough. In addition, she could be anticipating feelings of guilt for letting her partner down. If the negative cognitions and negative affect are relatively strong, the student might choose to do the assignment individually despite the loss of benefits. Examples of this kind of situations can be found in a wide variety of daily life activities, including work but also leisure activities such as sports or hobbies.

In the TEAM task, there are two elements that could contribute to social avoidance. First, teaming up with a partner that is perceived as superior may elicit thoughts such as “I'm not good enough”. This may trigger social comparison processes. Social comparison is a central feature of human social life, vital for human adaptation and survival, as it allows to assess abilities and aptitudes (Buunk & Gibbons, 2007; Festinger, 1954). Comparing oneself with others is a continuous process that can even happen automatically without awareness (Swallow & Kuiper 1988). In this framework, downward social comparison is as comparing oneself with those who are considered to be worse, while upward comparison refers to comparing oneself with those who are considered to be better. Downward comparison is usually associated to positive emotions such as relief. On the other hand, while upward social comparison can be helpful in providing information to assess and eventually improve our abilities, it can also be threatening, representing a chance for highlighting personal flaws and inadequacies (Swallow & Kuiper, 1988).

According to cognitive theories (Beck, 1979; Disner, Beevers, Haigh, & Beck, 2011), depression is characterized by a bias towards negativity in information processing and thinking. In particular, depression has been associated with a negative view of the self, with depressed individuals typically devaluing themselves, often being highly critical regarding their own abilities and in general seeing themselves as worthless and inadequate (Swallow & Kuiper, 1988). It has been proposed that social comparison could act as a trigger to the negative self-evaluations characteristic of depression, contributing to the etiology and maintenance of the disorder (Ahrens & Alloy, 1997; Buunk & Brenninkmeijer, 2000; Swallow & Kuiper, 1988). Furthermore, it is thought that social comparison could underlie the social withdrawal and self-imposed isolation typical of depression (Swallow & Kuiper, 1988). Following this line of thought, we hypothesized that in the TEAM task, depressed individuals would show an enhanced social avoidance response partly mediated by social comparison processes.

There is a second element that could contribute to social avoidance during the TEAM task. Apart from feeling that they are not as good as their partners, participants may also experience guilt due to viewing themselves as a liability for the team, and as a consequence may decide to go for the individual option. Excessive feelings of guilt are a key symptom of depression (American Psychiatric Association, 2013) and several studies using scales for measuring the construct of guilt have shown that individuals with depression experience elevated levels of guilt (Berrio*s et* al., 1992; Ghatavi, Nicolson, MacDonald, Osher, & Levitt, 2002; O'Connor, Berry, Weiss, & Gilbert, 2002; Pulcu, Zahn, & Elliott, 2013). Based on this, we expected that, in the TEAM task, enhanced feelings of guilt in depressed individuals would also contribute to an increased social avoidance response.

In summary, we developed a social decision-making task that allows measuring social avoidance. We predicted that during the TEAM task, outcomes in which the participant had a good performance would be associated with positive feelings, while outcomes in which the participant had a poor performance would be associated with negative feelings such as guilt and shame, particularly in cases where the co-player had done well. Regarding between group comparisons, it was hypothesized that depressed participants would show lower positive feelings due to anhedonia symptoms (Pizzagalli, 2014) and enhanced negative feelings, such as guilt and shame, in response to the task. Crucially, it was also expected that depressed participants would show an enhanced social avoidance response during the TEAM task compared to control volunteers. To our knowledge, this is the first study investigating social avoidance behavior in depression using a computerized social decision-making task.

Methods

Participants The study was approved by the Research Ethics Committee from the School of Psychology, Universidad de a República (Comité de Ética en Investigación de la Facutad de Psicología, Universidad de la República) (protocol number: 191175-001397-14). Written informed consent was obtained from all participants. Data were acquired from participants with depressive symptoms and healthy controls. The study was advertised through the university social networks. Those who were interested in taking part completed the Spanish version of the Beck Depression Inventory-II (BDI-II) (Beck, Ward, & Mendelson, 1961; Sanz, Perdigón, & Vázquez, 2003) on a website and were invited to self-nominate either for the depression or control group. Applicants who self-nominated for the depression group and had a score ≥16 on the BDI-II, as well as applicants from the control group who scored below 16 on the BDI-II, were invited to a recruitment session. In this session, volunteers were screened for depression and other psychiatric symptoms using the Spanish version of the MINI International Neuropsychiatric Interview (reliability measured as Cohen's kappa coefficient > 0.80 for MDD (Mordal, Gundersen, & Bramness, 2010); in the current study reliability of the MINI Plus was not assessed) (Ferrando, Bobes, & Gibert, 2004; Sheehan et al., 1998). In addition, during the interview participants were screened using the BDI-II. Inclusion criteria for the depression group were: satisfying the DSM-IV criteria for an episode of depression, a score ≥16 on the BDI-II (this requirement had to be met on three occasions: when they completed the BDI on the website, on the recruitment session and on the experimental session) and at least 3 weeks of not taking psychiatric medication. Participants in the control group had no current or past history of depression or any other psychiatric disorder. One depression and one control participant were excluded from the analysis for not believing in the task 'cover-story' (see paradigm description). The final sample consisted of 52 participants, 26 in each group. All participants were female, and the groups did not differ in age, study area or years of education (see Table 1 for details).

Table 1. Participant characteristics

	Control	Depression	Significance
n	26	26	
Age	24.03±4.54	25.51±4.77	*p* = .25 *NS*
Years of education	14.85±1.85	15.1±2.43	*p* = .56 *NS*
Area of education: Technology and Science / Health Science / Social Science and Arts	2 / 14 / 10	2 / 14 / 10	
BDI-II	1.23±1.68	29.54±8.59	*p* < .001
BDI-II – guilt sub-score	0.08±0.272	1.58±0.758	*p* < .001
PSI-Sociotropy	68.12±15.25	93.19±13.30	*p* < .001
PSI-Autonomy	68.12±14.69	91.77±11.05	*p* < .001
IIP-64-Total	54.27±23.02	104.73±32.83	*p* < .001

Note: Values are given as mean ± standard deviation; BDI, Beck Depression Inventory (scores from the experimental session); PSI, Personal Style Inventory; IIP, Inventory of Interpersonal Problems; STAI, State-Trait Anxiety Inventory for Adults; LSAS, Liebowitz Social Anxiety Scale; RSES, Rosenberg Self-Esteem Scale; SSES, State Self-esteem Scale; PANAS, Positive Affect Negative Affect Scale; INCOM, Iowa-Netherlands Comparison Orientation Measure; *p*-values of the independent-samples *t* test; NS, no significant difference between groups

Table 1 (Cont.) Participant characteristics

	Control	Depression	Significance
n	26	26	
IIP(domineering/controlling)	7.35±5.28	9.50±5.62	p = .16 *NS*
IIP (vindictive/self-centred)	4.12±2.76	8.96±6.85	p < .005
IIP (cold/distant)	3.46±3.43	10.19±6.56	p < .001
IIP (socially inhibited)	4.54±4.04	16.19±8.26	p < .001
IIP (non assertive)	7.50±4.33	18.42±6.81	p < .001
IIP (overly accommodating)	9.38±4.21	15.62±6.03	p < .001
IIP (self-sacrificing)	11.50±5.74	16.58±5.76	p = .003
IIP (intrusive/needy)	6.42±4.01	9.27±6.41	p = .06 *NS*
STAI-Trait	12.08±5.5	42.08±6.95	p < .001
STAI-State	8.73±7.7	32.27±11.12	p < .001
LSAS-Total	21.88±14.13	64.56±23.53	p < .001
LSAS-fear/anxiety scale	11.5±7.97	32.92±11.51	p < .001
LSAS-avoidance scale	10.38±7.16	31.64±13.1	p < .001
RSES	35.27±3.37	21.62±4.23	p < .001
SSES-Total	86.35±6.54	53.46±15.40	p < .001
SSES-Performance	31.04±2.39	18.85±5.73	p < .001
SSES-Social	33.12±1.93	20.81±6.72	p < .001
SSES-Appearance	22.19±3.5	13.81±5.44	p < .001
PANAS-Positive affect	34.65±6.55	18.88±3.82	p < .001
PANAS-Negative affect	17.00±2.32	28.42±7.46	p < .001
INCOM	29.73±4.69	37.19±4.00	p < .001

Note: Values are given as mean ± standard deviation; BDI, Beck Depression Inventory (scores from the experimental session); PSI, Personal Style Inventory; IIP, Inventory of Interpersonal Problems; STAI, State-Trait Anxiety Inventory for Adults; LSAS, Liebowitz Social Anxiety Scale; RSES, Rosenberg Self-Esteem Scale; SSES, State Self-esteem Scale; PANAS, Positive Affect Negative Affect Scale; INCOM, Iowa-Netherlands Comparison Orientation Measure; *p*-values of the independent-samples *t* test; NS, no significant difference between groups

Participants were assessed with several psychological rating scales (see Supplementary Materials and Table 1 for details).

The TEAM task A social decision-making behavioral task (Figure 1) was developed to assess social decision making. Participants were told that they would be playing a game with real people through a computer network, and that two of those co-players were in nearby labs doing the same procedure with other researchers. In reality, the task was pre-programmed and there were no real co-players. Participants were instructed on how to perform the TEAM task. First, they were taught how to perform a time-estimation test (Boksem, Kostermans, & De Cremer, 2011) (Figure 1a).

In this test, a red circle is shown; then, a sky-blue circle appears replacing the red one, and participants have to press the spacebar between 500ms and 1000ms after the color changed. Participants were told that depending on how well they did on the test, they could be ranked as a three, two or one-star player, with the three/one-star player being the most/least accurate. Participants were told that the two co-players that were in nearby labs were going to perform the test at the same time, and that the three of them were going to see each other rankings at the same time. The outcome of the test was pre-programmed and participants always received one star, while the other two co-players were classified as a two and as a three-star player (Figure 1b). Thus, this first part of the task allowed to create a social hierarchy (Zink et al., 2008) with participants always being ranked at the bottom. This procedure was implemented to facilitate upward social comparisons during the task (i.e. more occasions in which the participant performed the test incorrectly while the co-player did it well).

After establishing the social hierarchy, the core part of the TEAM task would begin (Figure 1c). In each trial, the participant had to choose between playing with a co-player (social option) or playing individually (individual option). Depending on the trial, the co-player could be a one, two or three-star player. In case of choosing the social option, the participant would be paired with a co-player of the category allocated to the trial. Furthermore, participants were told that they would be paired with a co-player who had also accepted to play with a co-player of the participant's rank. Thus, all players were supposed to make the same kind of decisions and play the game in the same role. In the case of the social option, the participant had to perform the time-estimation test simultaneously with the co-player, and the possible outcomes were: both did it correctly, one did it correctly and the other one did it incorrectly, and both did it incorrectly. In the first case, both earned 22 points, whereas in the remaining cases both players earned 20 points (Figure 1d). In case of choosing the individual option, the participant performed the time estimation test alone; if

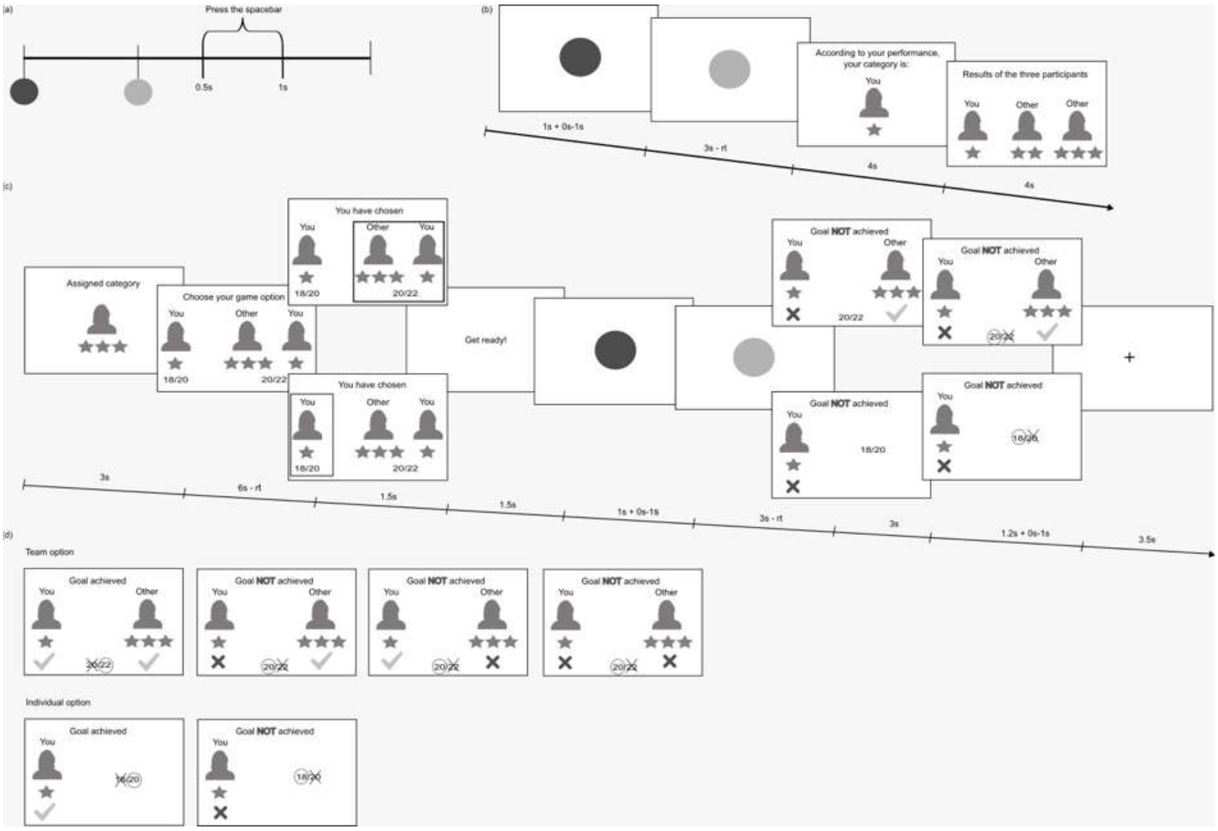


Figure 1. The TEAM task. (a) Time-line of the time-estimation test. (b) Establishment of a social hierarchy. (c) Example of a trial. (d) Payoff matrix. rt: reaction time. The original version in Spanish is provided in the Supplementary Materials (Figure S1).

the test was performed correctly, she would earn 20 points, otherwise she would get 18 points. Note that from a purely material point of view, it was always better to choose the social option. However, as mentioned above, the social option implies social comparison and guilt related processes which may induce participants to choose the individual option. It is worth mentioning that social comparisons processes are present both at the decision time when the participant has to decide whether to make a team or not (depending on the hierarchy of the co-player) (Boksem, Kostermans, Milivojevic, & De Cremer, 2012; Yin Wu, Zhou, van Dijk, Leliveld, & Zhou, 2011; Zink et al., 2008) as well as in the outcome time when the performances of the two players are shown (Boksem et al., 2011; Kedia, Mussweiler, & Linden, 2014; Ma et al., 2011; Qi, Raiha, Wu, & Liu, 2018; Qiu et al., 2010; Yan Wu, Zhang, Elieson, & Zhou, 2012).

The participant was told that if during the game there had been trials where she ended up playing in a team with one of the participants that were in nearby labs, at the end of the session she would be introduced to this co-player. This procedure was implemented to reinforce the social aspect of the task. Furthermore, studies have shown that social comparison processes depend on whether the participant anticipates actual contact with the comparison other (Bunk & Gibbons, 2007).

The task was programmed using Psychopy (Peirce, 2007, 2008) and had three types of trials, corresponding to whether the participant could play with a co-player who had one, two or three stars. Subjects performed 45 trials of the task, 15 in each condition, for about 18 minutes. Outcomes were manipulated so that three-star co-players had a 100% rate of correct responses in the time-estimation test, two-star co-players had 60% and one-star co-players had a 30% rate. As the participant was a one-star player, she had a 30% rate of correct responses. Thus, participants performance in the time estimation test was fixed.

After finishing the task, participants rated their subjective emotional reaction to the different situations they faced during the task: the possibility of having co-players of each category, the outcomes of the trials and the possibility of meeting other participants. Participants rated the emotions of happiness, anger, sadness, guilt, shame and disappointment on a nine-point Likert scale.

At the end of the session, participants were debriefed. None of the participants manifested discomfort regarding the cover-story. All participants received the same reward (a cinema ticket).

Analysis of behavioral data and emotional responses to the task Binary logistic generalized linear models were used to analyze participant's decisions during the game. Generalized estimated equations were used to estimate model parameters, thus adjusting for correlations due to repeated observations within each participant over the trials. We used a logit link function and assumed an exchangeable working correlation structure. The statistical significance level was set at .05. The participant´s decision (playing individually or with a co-player) was entered as the binary dependent variable. The variable subject was entered as a repeated effect variable and group as a fixed effect variable. Trial number and trial type (having the possibility of playing with a one, two or three-star co-player) were set as covariates. Group, trial number and trial type main effects were explored. Also, the following interactions were set in the model: group*trial number, group*trial type, trial number*trial type and group*trial number*trial type. Effect size is presented as odds ratio for the group variable.

A mixed ANOVA was used to examine the effect of group and game type (playing with a one, two or three-star co-player or individually) on emotional responses about having to play in these four situations. A mixed ANOVA was used to examine the effect of group and trial outcome (for the social option) on emotional responses when facing those outcomes. Effect size was calculated using partial eta-squared (η^2_p).

Results

Clinical ratings Between-group *t* tests were used to study the differences between the depression and control groups in the clinical ratings. Depressed participants scored higher than controls on measures of depression, anxiety, social anxiety and negative affect while scoring lower on measures of self-esteem and positive affect. The depression group also scored significantly

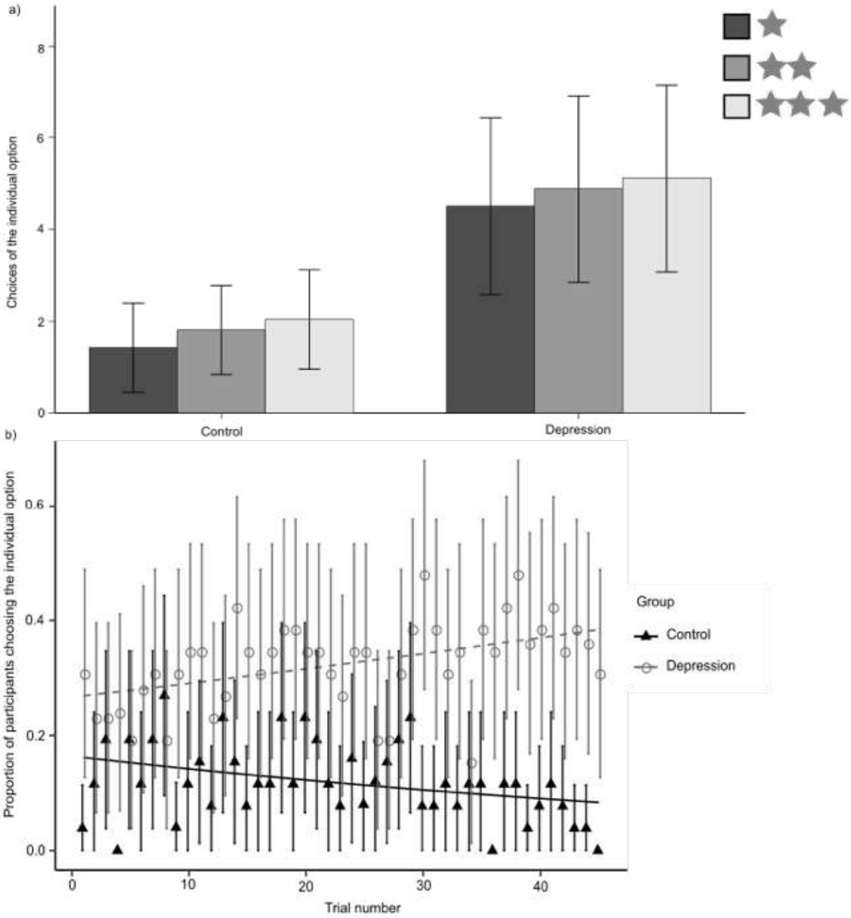


Figure 2**.** Behavioral results. (a) Participants chose more often the individual option the higher the stars of the co-player. Depressed participants chose the individual option significantly more often than controls. Error bars denote 95% confidence intervals. (b) Control participants selected less the individual option as the task progressed while depressed participants did not show this behavior. Lines show mean values predicted by the model. Error bars denote 95% confidence intervals.

higher in sociotropy and autonomy and in social comparison orientation. In addition, as in previous studies (Gradin *et* al., 2015) the depression group scored significantly higher in interpersonal problems in the following domains: vindictive/self-centered, cold/distant, socially inhibited, nonassertive, overly accommodating, and self-sacrificing (see Table 1).

Behavioral results The generalized linear model identified a significant main effect of group on decision making (X^2(1, N = 52) = 10.755, p = .001, OR = 3.61, 95% IC [1.68, 7.75]), with depressed participants choosing the individual option significantly more often than controls. As a consequence, depressed participants earned a lower number of points during the task compared to controls (p = .007) (see Figure 2a and Supplementary Table S2). In addition, a significant interaction between group and trial number was found (Wald X^2(2, N = 52) = 9.627, p = .002). This was related to the fact that control participants selected the individual option less often as the task went on, while depressed participants exhibited the opposite behavioral pattern (see Figure 2b). Also, a marginally significant main effect of trial-type (X^2(1, N = 52) = 3.557, p = .059) was found, with participants choosing more often the individual option the higher the number of stars of the co-player. All other effects and interactions were not significant (see Supplementary Material for analyses on reaction times).

Additionally, we tested a regression model without interaction terms to simplify the interpretation of the main effects of group, trial number and trial type. This analysis yielded significant effects for group (X^2 (1, N = 52) = 10.428, p = .001) and trial type (X^2 (2, N = 52) = 3.935, p = .047) and no significant effect for trial number.

As some of the depressed participants also experienced symptoms of social anxiety, which strongly relate to social avoidance, we performed additional analyses controlling for social anxiety levels measured with the LSAS. These analyses also yielded significant results for the effects of group and the group*trial number interaction (see Supplementary Material for the details on these analyses).

Emotional response

Emotions related to each of the task game types After performing the TEAM task, participants rated their emotions (happiness, shame, guilt, anger and sadness) about having to play in each of the game types (i.e. playing with a one, two or three-star co-player or individually) (Figure 3a and Supplementary Table S3).

For the emotion of happiness, a significant main effect (F(1.381,69.035) = 11.185, p < .001, η^2_p = .183, 95% CI [.04, .33]) and a significant linear trend (F(1,50) = 10.896, p = .002, η^2_p = .179, 95% CI [.03, .35]) were found for game type. Post-hoc pairwise comparisons after Bonferroni correction identified significant differences in happiness between the social and the individual games with the social games eliciting higher happiness than the individual game (p ≤ .006). A significant main effect of group was also found (F(1,50) = 9.334, p = .004, η^2_p = .157, 95% CI [.02, .33]), with the control group reporting higher happiness than the depression group.

For the emotion of shame, a significant main effect (F(1.457,72.864) = 12.671, p < .001, η^2_p = .202, 95% CI [.06, .35]) and a linear trend (F(1,50) = 16.136, p < .001, η^2_p = .244, 95% CI [.06, .42]) for game type were found, with the three-star co-player eliciting more shame than the other game types and in turn the two-star co-player eliciting more shame than the one-star co-player and the individual game (p ≤ .023). A significant main effect of group was found (F(1,50) = 15.458, p < .001, η^2_p = .236, 95% CI [.06, .41]), with the depression group reporting more shame than controls.

For the emotion of guilt, a significant main effect (F(1.614,80.711) = 6.165, p = .006, η^2_p = .110, 95% CI [.01, .24]) and linear trend (F(1,50) = 8.170, p = .006, η^2_p = .140, 95% CI [.01, .31]) were found for the game type, with the three-star co-player eliciting more guilt than playing individually (p = .033). The remaining effects and interactions were not found significant.

Emotions related to each of the task outcomes Participants reported their emotions regarding the task's outcomes for the social option (Figure 3b and Supplementary Table S3).

For the emotion of happiness, a significant main effect of outcome was found (F(2.236,111.814) = 70.276, p < .001, η^2_p = .584, 95% CI [.46, .66]), with the outcome "*You correct, Other correct*" eliciting the highest happiness, followed by "*You correct, Other wrong*", then by "*You wrong, Other correct*" and finally by "*You wrong, Other wrong*" (p ≤ .012).

A significant main effect of outcome was found for the emotions of shame (F(1.615,80.763) = 31.724, p < .001, η^2_p = .388, 95% CI [.22, .51]) and guilt (F(1.819,90.958) = 25.732, p < .001, η^2_p = .340, 95% CI [.18, .46]) with the outcome "*You wrong, Other correct*" eliciting higher guilt and shame than the outcome "*You wrong, Other wrong*" (p = .001 for shame and p = .008 for guilt) and this outcome in turn eliciting more shame and guilt than the remaining outcomes (p ≤ .002).

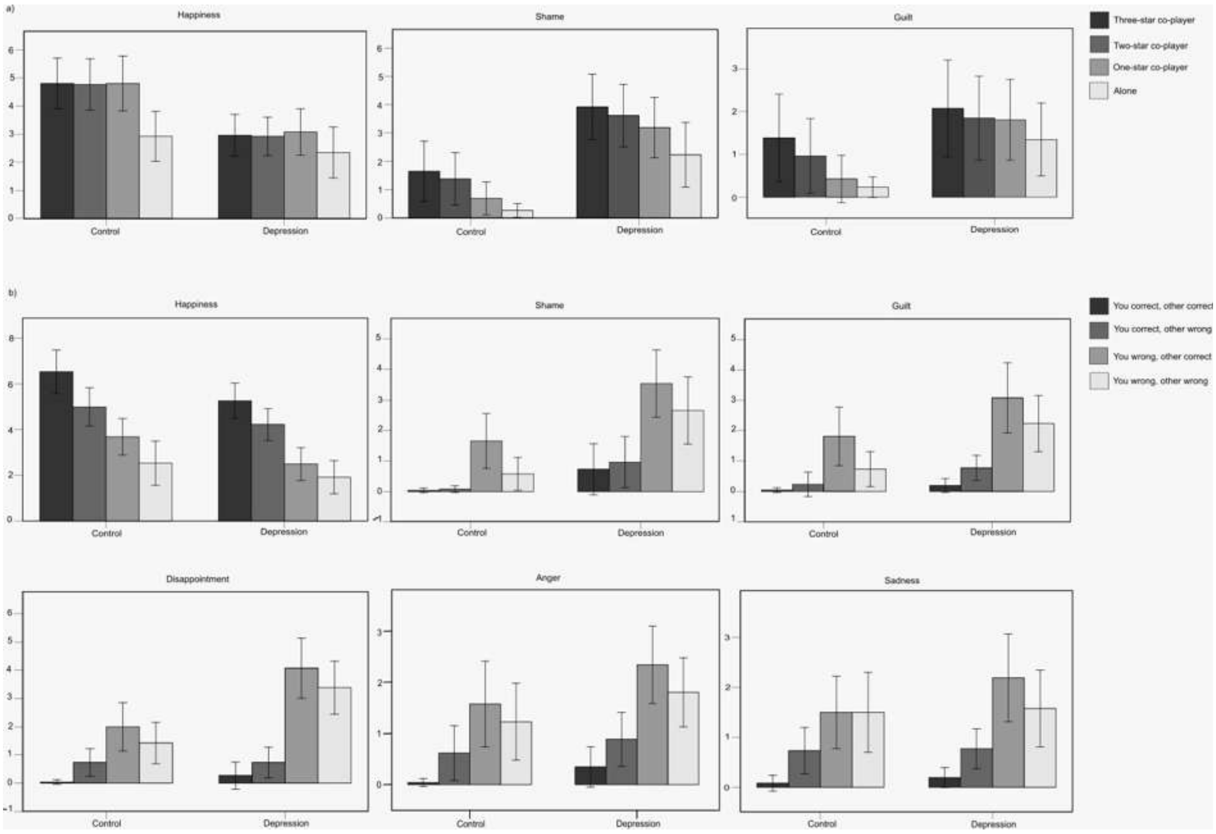


Figure 3. Emotional responses to the TEAM Task. (a) Emotional responses for each game type. Error bars denote 95% confidence intervals. (b) Emotional responses for the social outcomes. Error bars denote 95% confidence intervals.

There was also a significant main effect of outcome for the emotions of disappointment (F(1.852,92.614) = 43.546, p < .001, η^2_p = .466, 95% CI [.31, .57]), anger (F(2.096,104.788) = 25.327, p < .001, η^2_p = .336, 95% CI [.18, .45]) and sadness (F(1.776,88.791) = 24.464, p < .001, η^2_p = .329, 95% CI [.17, .45]), with the outcomes where the participant performed poorly eliciting more disappointment, anger and sadness than the outcomes where the participant performed well (p < .005).

A significant main effect of group was found for happiness (F(1,50) = 4.617, p = .037, η^2_p = .085, 95% CI [.0, .25]), shame (F(1,50) = 10.149, p = .002, η^2_p = .169, 95% CI [.02, .34]), guilt (F(1,50) = 7.452, p = .009, η^2_p = .130, 95% CI [.01, .30]) and disappointment (F(1,50) = 10.191, p = .002, η^2_p = .169, 95% CI [.02, .35]), with the depression group reporting less happiness and more shame, guilt and disappointment than controls. For shame a significant interaction was found (F(1.615,80.763) = 3.587, p = .041, η^2_p = .067, 95% CI [.00, .18]), with depressed participants reporting significantly more shame for the outcomes "*You wrong, Other correct*"(p = .009), "*You wrong, Other wrong*"(p = .001) and "*You correct, Other wrong*"(p = .035), but not for the outcome "*You correct, Other correct*". For disappointment, a significant interaction was also found (F(1.852,92.614) = 7.161, p = .002, η^2_p = .125, 95% CI [.02, .25]), with depressed participants reporting more disappointment than controls for the outcomes where the participant performed poorly (p = .003) but not for the outcomes where the participant performs correctly.

The remaining effects and interactions were not significant.

Emotions related to a possible encounter with co-players

Depressed participants reported less happiness (t(50) = -4.780, p < .001, d = -1.35, 95% CI [-4.315, -1.762]) and more sadness (t(50) = 2.439, p = .018, d = 0.69, 95% CI [.2, .2.03]), shame (t(50) = 4.327, p < .001, d = 1.22, 95% CI [1.57, 4.28]) and disappointment (t(50) = 3.062, p = .004, d = 0.87, 95% CI [.06, 2.87]) about the possibility of meeting their co-players in comparison to controls (Supplementary Table S3).

Within the depression group, we searched for correlations between decision making and emotional reports, but no significant correlations were found.

Discussion

Effect of the co-player rank on emotional and behavioral responses

It was found that the higher the rank of the co-player, the higher the negative emotions (shame, guilt) reported by participants about having to play with that kind of co-player. Behaviorally, a marginally significant effect was found with participants choosing more frequently the individual option the higher the rank of the co-player. From a purely material point of view this wouldn't be reasonable, as the higher the rank of the co-player, the higher the chances for the team earning the maximum number of points. However, the higher the rank of the co-player, the higher also the chances of the participant ending up in the situation in which she performs poorly while the co-player performs well. This situation implies an upward social comparison plus dealing with the fact that the team has not reached its maximum because of the participant's fault. In agreement with this, participants reported the highest levels of shame and guilt for this outcome of the task. Thus, the negative feelings related to the anticipation of this outcome seem to be driving the increased social avoidance response the higher the rank of the co-player.

These findings are in line with research that shows that upward social comparison can be experienced in a negative way (Buunk & Gibbons, 2007) that increases negative affect (Fuhr, Hautzinger, & Meyer, 2014). Our findings are also in line with research indicating that anticipation of guilt feelings for letting other people down may lead individuals to avoid forming interdependent partnerships with people they see as more competent than themselves (Wiltermuth & Cohen, 2014).

Effect of depression on emotions and decision making

Depressed and control volunteers differed in the way they reacted to the TEAM task. In particular, depressed participants reported higher feelings of shame for all the social outcomes except for when both participant and co-player performed well; higher levels of disappointment when having a bad performance; and less happiness and more sadness, shame and disappointment about the possibility of meeting their co-players. Crucially, regarding decision making, depressed volunteers showed an increased social avoidance response, choosing the individual option significantly more often than controls. Furthermore, it was observed that while control participants selected the individual option less frequently as the task went on, depressed participants showed the opposite behavior selecting the individual option more frequently as the task progressed. This suggest that while the balance between the attractiveness and the challenge of the social option was more positive for controls as the task went on, depressed participants did not show the same kind of learning.

These findings are in agreement with the proposal that avoidance (Blalock & Joiner, 2000; Ferster, 1973; Lewinsohn, 1974; Spurrell & McFarlane, 1995), and especially social avoidance (Joiner, 2000), is a key aspect of depression. Our findings are also in line with studies reporting an association between avoidance and depression by using self-reported questionnaires (Ottenbreit et al., 2014; Quigley, Wen, & Dobson, 2017).

While social avoidance, and especially avoidance of interpersonal conflict, may preclude individuals from experiencing negative interpersonal outcomes, it may contribute to depression in several ways (Joiner, 2000). First, interpersonal avoidance may result in the loss of social and material opportunities, and loss is an important trigger for depression (Heikkinen et al., 1997). Second, interpersonal avoidance may imply a diminution in social reinforcement and social support. Finally, social avoidance may lead to isolation and decreased social skills acquisition and learning, further contributing to interpersonal problems and the maintenance of depression (Lewinsohn, 1974; Trew, 2011).

One factor that makes the social option stressful is that it implies social comparison processes. Thus, an increased sensitivity to social comparison in depression could be contributing to the increased negative emotions and enhanced social avoidance observed in this group. This is in agreement with a theory (Swallow & Kuiper, 1988) that postulates social comparison as a trigger for negative self-evaluations and social withdrawal in depression. Our findings are also in agreement with several studies indicating that dysphoric subjects interpret social comparison information in a less self-serving way and that they focus on the fact that others are better off than they are, confirming their low rank and status (Buunk & Brenninkmeijer, 2000), and with a study showing that when facing an imaginary upward comparison, depressed individuals show a stronger decrease in positive affect compared to controls (Bäzner, Brömer, Hammelstein, & Meyer, 2006).

Other approaches have also linked social comparison to depression (Buunk & Brenninkmeijer, 2000). In the context of the involuntary subordinate strategies theory (Price, Sloman, Gardner, Gilbert, & Rohde, 1994), depression is understood as a state of involuntary subordination, characterized by a sense of defeat, frustration and hopelessness, a lack of challenging and exploratory behavior, and the use of strategies signaling acknowledgement of one's low status. From this perspective, the enhanced social avoidance response that we observed in depressed individuals could be interpreted as a deficit in challenging and exploratory behavior as well as a signal of no threat. In the same line, a recent study showed avoidance of competitive situations in depression (Kupferberg, Hager, et al., 2016).

Using the self-reported rating scale Iowa-Netherlands Comparison Orientation Measure (INCOM) (Gibbons & Buunk, 1999) , we found that depressed participants obtained a higher score in social comparison orientation compared to controls. Social comparison orientation refers to the disposition that a person has to compare herself with others. It has been suggested that especially those individuals with a high degree of uncertainty about themselves and a tendency to focus attention on the self, such as people with low self-esteem, depression or high neuroticism, would be particularly likely to engage in social comparison processes (Gibbons & Buunk, 1999). Our finding of a higher social comparison score in depressed participants is consistent with this hypothesis, as well as with a previous study also reporting an increased social comparison orientation in depression using the INCOM scale (Bäzne*r et* al., 2006). This suggests that depressed participants are more concerned and spent more time and energy ruminating about their standing relative to others. This supports the notion that the increased social avoidance response observed during the TEAM task in depression relates to avoiding self-evaluative information to prevent discomfort and is not due to a lack of interest in social comparison information.

Apart from social comparison processes, social avoidance during the TEAM task may also be driven by anticipation of guilt feelings for an eventual poor performance leading to the team not earning the maximum number of points. In our study, depressed participants reported higher feelings of guilt in relation to the task compared to controls, and enhanced feelings of guilt are a typical symptom of depression (American Psychiatric Association, 2013). Thus, although no significant correlation was found between guilt feelings and decision making within the depression group, it is likely that anticipation of guilt feelings for failing the team may be contributing to an enhanced social avoidance response in this group. This is in agreement with evidence showing that highly guilt-prone individuals are more likely to anticipate negative feelings about letting team partners down by underperforming, and therefore avoid situations in which outcomes are interdependent with others whom they see as more competent (Wiltermuth & Cohen, 2014). Future studies should further explore the relationship between guilt feelings and social avoidance in depression.

Possible limitations of the study should be noted. First, a university sample was used, which could limit generalizability of results. This recruitment method was used to facilitate recruitment of unmedicated depressed participants. Second, the sample consisted only of women and so results may not be generalize to men. Third, emotional responses were measured retrospectively. Thus, it is possible that differences in the recall of emotions may have contributed to the observed between group differences. One possibility for alleviating this would have been the implementation of on-line measures of emotions (i.e. measuring emotional reactions during the task, at the precise moments that events occur). However, interrupting the task to ask participants to reflect and answer on their emotions brings the drawback that it may affect the experience of the task (i.e. reflecting on how you feel and reporting how you feel may affect how you further perceive the task). Therefore, to preserve ecological validity retrospective measures were applied instead of on-line measures of emotions. Fourth, future work might include a pre-post measure of emotions using a general scale, such as the PANAS (Positive Affect Negative Affect Scale) (Dufey & Fernández, 2012; Watson, Clark, & Tellegen, 1988) and a pre-post measure of guilt and shame such as the Guilt and Shame Proneness scale (Cohen, Wolf, Panter, & Insko, 2011). Fifth, apart from showing symptoms of depression, our clinical sample also exhibited social anxiety symptoms. While our findings remained significant after controlling for social anxiety levels, it is difficult to completely disentangle the effects of depression and social anxiety on social avoidance given the high levels of comorbidity. Alternatively, it would be interesting to study social avoidance within the Research Domain Criteria (RDOC) framework, which understands psychopathology in terms of fundamental components (e.g., executive functioning, affect regulation, person perception) that span the full range of human behavior from normal to abnormal, instead of using the existing psychiatric categories (Sanislow *et al*., 2010; Woody & Gibb, 2015). Sixth, since the TEAM task is a novel instrument further work on task validation would be of relevance. Finally, it would be interesting to test variations of the task such as ranking the participant as a two-star player and observe the preferences for upward or downward comparisons.

In summary, our study provides evidence for altered social decision making in depression using a social decision-making behavioral task. We observed that when having the possibility of playing in teams, unmedicated depressed participants opted more often than healthy controls to play alone, with this behavior leading to lower earnings. Our findings highlight the role of social avoidance in depression and how this behavior may lead to negative consequences and loss of opportunities in people's daily lives. Our study also highlights the role of social comparison and guilt-related processes in underlying social avoidance in depression.

Acknowledgements

Special thanks to all the volunteers who participated in the research and to the Center for Basic Research in Psychology (CIBPsi). This study was funded by Comisión Sectorial de

Investigación Científica (CSIC), Universidad de la República (Uruguay) and a Master Scholarship from the Comisión Académica de Posgrados, Universidad de la República (Uruguay) to G.F.T. A.C. and V.B.G. were funded by Universidad de la República, Agencia Nacional de Investigación e Innovación (ANII) and the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Ethic statements

The study was approved by the Research Ethics Committee from the School of Psychology, Universidad de la República (Comité de Ética en Investigación de la Facutad de Psicología, Universidad de la República) (protocol number: 191175-001397-14).

References

Ahrens, A. H., & Alloy, L. B. (1997). Social comparison processes in depression. *Health, Coping, and Well-Being: Perspectives from Social Comparison Theory*, 389–410.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Bäzner, E., Brömer, P., Hammelstein, P., & Meyer, T. D. (2006). Current and former depression and their relationship to the effects of social comparison processes. Results of an internet based study. *Journal of Affective Disorders*, *93*(1–3), 97–103. https://doi.org/10.1016/j.jad.2006.02.017

Beck, A. T. (1979). *Cognitive Therapy and the Emotional Disorders*. New York: Plume.

Beck, A. T., Ward, C., & Mendelson, M. (1961). Beck depression inventory (BDI). *Archives of General Psychiatry*, *4*(6), 561–571.

Berrios, G. E., Bulbena, A., Bakshi, N., Dening, T. R., Jenaway, A., Markar, H., … Mitchell, S. L. (1992). Feelings of Guilt in Major Depression. *British Journal of Psychiatry*, *160*(06), 781–787. https://doi.org/10.1192/bjp.160.6.781

Blalock, J. A., & Joiner, T. E. (2000). Interaction of Cognitive Avoidance Coping and Stress in Predicting Depression/Anxiety. *Cognitive Therapy and Research*, *24*(1), 47–65. https://doi.org/10.1023/A:1005450908245

Bobes, J., Badia, X., Luque, A., Garcia, M., González, M. P., & Dal-Ré, R. (1999). Validación de las versiones en español de los cuestionarios Liebowitz Social Anxiety Scale, Social Anxiety and Distress Scale y Sheehan Disability Inventory para la evaluación de la fobia social. *Med Clin (Barc), 112*, 530–538.

Boksem, M. A. S., Kostermans, E., & De Cremer, D. (2011). Failing where others have succeeded: Medial Frontal Negativity tracks failure in a social context. *Psychophysiology*, *48*(7), 973–979. https://doi.org/10.1111/j.1469-8986.2010.01163.x

Boksem, M. A. S., Kostermans, E., Milivojevic, B., & De Cremer, D. (2012). Social status determines how we monitor and evaluate our performance. *Social Cognitive and Affective Neuroscience*, *7*(3), 304–313. https://doi.org/10.1093/scan/nsr010

Buunk, A., Belmont, J., Peiró, J., Zurriaga, R., & Gibbons, F. (2005). Diferencias individuales en la comparación social: propriedades de la escala española de orientación hacia la comparación social. *Revista Latinoamericana de Psicología, 37*(3), 561–579. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0120-05342005000300008&lng=pt&nrm=iso&tlng=es

Buunk, A. P., & Gibbons, F. X. (2006). Social comparison orientation: a new perspective on those who do and those who don't compare with others. In S. Guimond (Ed.), *Social Comparison and Social Psychology: Understanding cognition, intergroup relations and culture* (pp. 15–33). Cambridge: Cambridge University Press.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3–21. https://doi.org/10.1016/j.obhdp.2006.09.007

Buunk, B., & Brenninkmeijer, V. (2000). Social comparison processes among depressed individuals: Evidence for the evolutionary perspective on involuntary subordinate strategies. *Subordination and Defeat: An Evolutionary Approach to Mood Disorders and Their Therapy*, 147–164.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947–966. https://doi.org/10.1037/a0022641

Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews. Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Dufey, M., & Fernández, A. M. (2012). Validez y confiabilidad del Positive Affect and Negative Affect Schedule (PANAS) en estudiantes universitarios chilenos. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, *1*(34), 157–173.

Ferrando, L., Bobes, J., & Gibert, J. (2004). MINI Internation Neuropsychiatric Interviews. Versión en Español. Sheehan DV & Lecrubier Y.

Ferster, C. B. (1973). A functional analysis of depression. *American Psychologist*, *28*(10), 857–870. https://doi.org/10.1037/h0035605

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fuhr, K., Hautzinger, M., & Meyer, T. D. (2014). Are Social Comparisons Detrimental for the Mood and Self-Esteem of Individuals with an Affective Disorder? *Cognitive Therapy and Research*, *39*(3), 279–291. https://doi.org/10.1007/s10608-014-9656-2

Ghatavi, K., Nicolson, R., MacDonald, C., Osher, S., & Levitt, A. (2002). Defining guilt in depression: a comparison of subjects with major depression, chronic medical illness and healthy controls. *Journal of Affective Disorders*, *68*(2–3), 307–315. https://doi.org/10.1016/S0165-0327(01)00335-4

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, *76*(1), 129–142. https://doi.org/10.1037//0022-3514.76.1.129

Goossens, L. (2014). Affinity for Aloneness in Adolescence and Preference for Solitude in Childhood. In R. J. Coplan & J. C. Bowker (Eds.), *The handbook of solitude : psychological perspectives on social isolation, social withdrawal, and being alone* (pp. 150–166). West Sussex: John Wiley & Sons.

Gradin, V. B., Pérez, A., MacFarlane, J. A., Cavin, I., Waiter, G., Engelmann, J., … Steele, J. D. (2015). Abnormal brain responses to social fairness in depression: an fMRI study using the Ultimatum Game. *Psychological Medicine*, *45*(06), 1241–1251. https://doi.org/10.1017/S0033291714002347

Gradin, V. B., Pérez, A., Macfarlane, J. A., Cavin, I., Waiter, G., Tone, E. B., … Steele, J. D. (2016). Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. *Psychological Medicine*, *46*(06), 1289–1300. https://doi.org/10.1017/S0033291715002834

Hasler, G. (2012). Can the neuroeconomics revolution revolutionize psychiatry? *Neuroscience and Biobehavioral Reviews*, *36*(1), 64–78. https://doi.org/10.1016/j.neubiorev.2011.04.011

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*(6), 895–910. https://doi.org/10.1037//0022-3514.60.6.895

Heikkinen, M. E., Henriksson, M. M., Isometsä, E. T., Marttunen, M. J., Aro, H. M., & Lönnqvist, J. K. (1997). Recent Life Events and Suicide in Personality Disorders. *Journal of Nervous and Mental Disease*, *185*(6), 373–381. https://doi.org/10.1097/00005053-199706000-00003

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, *56*(6), 885–892. https://doi.org/10.1037/0022-006X.56.6.885

Joiner, T. E. (2000). Depression's Vicious Scree: Self-Propagating and Erosive Processes in Depression Chronicity. *Clinical*

Psychology: Science and Practice, *7*(2), 203–218. https://doi.org/10.1093/clipsy.7.2.203

Joiner, T. E., & Timmons, K. A. (2009). Depression in Its Interpersonal Context. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 322–339). New York: The Guilford Press.

Kedia, G., Mussweiler, T., & Linden, D. E. J. (2014). Brain mechanisms of social comparison and their influence on the reward system. *NeuroReport*, *25*(16), 1255–1265. https://doi.org/10.1097/WNR.0000000000000255

Kessler, R. C., & Bromet, E. J. (2013). The Epidemiology of Depression Across Cultures. *Annual Review of Public Health*, *34*(1), 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

King-Casas, B., & Chiu, P. H. (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biological Psychiatry*, *72*(2), 119–125. https://doi.org/10.1016/j.biopsych.2012.03.033

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Research*, *188*(3), 303–309. https://doi.org/10.1016/j.psychres.2011.04.019

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2016.07.002

Kupferberg, A., Hager, O., Fischbacher, U., Brändle, L., Haynes, M., & Hasler, G. (2016). Testing the social competition hypothesis of depression using a simple economic game. *British Journal of Psychiatry Open*, *2*(2), 163–169. https://doi.org/10.1192/bjpo.bp.115.001362

Lewinsohn, P. (1974). A behavioral approach to depression. In R. Friedman & M. Katz (Eds.), *The psychology of depression: Contemporary theory and research.* (p. 318). Oxford: John Wiley & Sons.

Liebowitz, M. R. (1987). Social Phobia. In D. F. Klein (Ed.) (Vol. 22, pp. 141–173). S. Karger AG. https://doi.org/10.1159/000414022

Lis, S., & Kirsch, P. (2016). Neuroeconomic Approaches in Mental Disorders (pp. 311–330). https://doi.org/10.1007/978-3-642-35923-1_16

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., & Weber, B. (2011). Empathic responses to others' gains and losses: An electrophysiological investigation. *NeuroImage*, *54*(3), 2472–2480. https://doi.org/10.1016/j.neuroimage.2010.10.045

Mordal, J., Gundersen, & Bramness, J. G. (2010). Norwegian version of the Mini-International Neuropsychiatric Interview: Feasibility, acceptability and test-retest reliability in an acute psychiatric ward. *European Psychiatry*, *25*(3), 172–177. https://doi.org/10.1016/j.eurpsy.2009.02.004

O'Connor, L. E., Berry, J. W., Weiss, J., & Gilbert, P. (2002). Guilt, fear, submission, and empathy in depression. *Journal of Affective Disorders*, *71*(1–3), 19–27. https://doi.org/10.1016/S0165-0327(01)00408-6

Ottenbreit, N. D., Dobson, K. S., & Quigley, L. (2014). An examination of avoidance in major depression in comparison to social anxiety disorder. *Behaviour Research and Therapy*, *56*, 82–90. https://doi.org/10.1016/j.brat.2014.03.005

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. https://doi.org/10.3389/neuro.11.010.200

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Price, J., Sloman, L., Gardner, R., Gilbert, P., & Rohde, P. (1994). The social competition hypothesis of depression. *The British Journal of Psychiatry*, *164*(3), 309–315. https://doi.org/10.1192/bjp.164.3.309

Pulcu, E., & Elliott, R. (2015). Neural origins of psychosocial functioning impairments in major depression. *The Lancet Psychiatry*, *2*(9), 835–843. https://doi.org/10.1016/S2215-0366(15)00237-0

Pulcu, E., Zahn, R., & Elliott, R. (2013). The Role of Self-Blaming Moral Emotions in Major Depression and Their Impact on Social-Economical Decision Making. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00310

Qi, Y., Raiha, S., Wu, H., & Liu, X. (2018). Corrigendum to "Social value orientation modulates context-based social comparison preference in the outcome evaluation: An ERP study" [Neuropsychologia 112 (2018) 135–144] (S002839321830085X)(10.1016/j.neuropsychologia.2018.02.028). *Neuropsychologia*, *113*, 151. https://doi.org/10.1016/j.neuropsychologia.2018.04.005

Qiu, J., Yu, C., Li, H., Jou, J., Tu, S., Wang, T., … Zhang, Q. (2010). The impact of social comparison on the neural substrates of reward processing: An event-related potential study. *NeuroImage*, *49*(1), 956–962. https://doi.org/10.1016/j.neuroimage.2009.08.025

Quigley, L., Wen, A., & Dobson, K. S. (2017). Avoidance and depression vulnerability: An examination of avoidance in remitted and currently depressed individuals. *Behaviour Research and Therapy*, *97*, 183–188. https://doi.org/10.1016/j.brat.2017.07.015

Robins, C. J., Ladd, J., Welkowitz, J., Blaney, P. H., Diaz, R., & Kutcher, G. (1994). The personal style inventory: Preliminary validation studies of new measures of sociotropy and autonomy. *Journal of Psychopathology and Behavioral Assessment*, *16*(4), 277–300. https://doi.org/10.1007/BF02239408

Rojas-Barahona, C. A., Zegers, B., & Förster, C. E. (2009). La escala de autoestima de Rosenberg: Validación para Chile en una muestra de jóvenes adultos, adultos y adultos mayores. *Revista Médica de Chile, 137*(6), 791–800. https://doi.org/10.4067/S0034-98872009000600009

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton NJ: Princeton University Press.

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., … Cuthbert, B. N. (2010). Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology*, *119*(4), 631–639. https://doi.org/10.1037/a0020909

Sanz, J., Perdigón, A. L., & Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2 . Propiedades psicométricas en población general. *Clínica y Salud*, *14*(3), 249–280.

Sheehan, D. V, Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59 Suppl 20*, 22-33;quiz 34-57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9881538

Silk, J. S., Siegle, G. J., Lee, K. H., Nelson, E. E., Stroud, L. R., & Dahl, R. E. (2014). Increased neural response to peer rejection associated with adolescent depression and pubertal development. *Social Cognitive and Affective Neuroscience*, *9*(11), 1798–1807. https://doi.org/10.1093/scan/nst175

Spielberger, C. D. (1968). *The State-trait Anxiety Inventory (STAI): Test Manual for Form X*. Consulting Psychologists Press.

Spielberger, C., Gorsuch, R., & Lushene, R. (1982). *Manual del cuestionario de ansiedad estado/rasgo (STAI)*. Madrid: TEA Ediciones.

Spurrell, M. T., & McFarlane, A. C. (1995). Life-events and psychiatric symptoms in a general psychiatry clinic: The role of intrusion and avoidance. *British Journal of Medical Psychology*, *68*(4), 333–340. https://doi.org/10.1111/j.2044-8341.1995.tb01840.x

Swallow, S. R., & Kuiper, N. A. (1988). Social comparison and negative self-evaluations: An application to depression. *Clinical Psychology Review*, *8*(1), 55–76. https://doi.org/10.1016/0272-7358(88)90049-9

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, *31*(7), 1156–1168. https://doi.org/10.1016/j.cpr.2011.07.007

Wang, Y., Yang, L.-Q., Li, S., & Zhou, Y. (2015). Game Theory Paradigm: A New Tool for Investigating Social Dysfunction in Major Depressive Disorders. *Frontiers in Psychiatry*, *6*, 128. https://doi.org/10.3389/fpsyt.2015.00128

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wiltermuth, S. S., & Cohen, T. R. (2014). "I'd only let you down": Guilt proneness and the avoidance of harmful interdependence. *Journal of Personality and Social Psychology*, *107*(5), 925–942. https://doi.org/10.1037/a0037523

Woody, M. L., & Gibb, B. E. (2015). Integrating NIMH Research Domain Criteria (RDoC) into depression research. *Current Opinion in Psychology*, *4*, 6–12. https://doi.org/10.1016/j.copsyc.2015.01.004

Wu, Y., Zhang, D., Elieson, B., & Zhou, X. (2012). Brain potentials in outcome evaluation: When social comparison takes effect. *International Journal of Psychophysiology*, *85*(2), 145–152. https://doi.org/10.1016/j.ijpsycho.2012.06.004

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., & Zhou, X. (2011). Social Comparison Affects Brain Responses to Fairness in Asset Division: An ERP Study with the Ultimatum Game. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00131

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron*, *58*(2), 273–283. https://doi.org/10.1016/j.neuron.2008.01.025

El CIBPsi en la seire LASchool
The CIBPsi at the LASchool series

La **4th Latin American School for Education, Cognitive and Neural Sciences** se celebró del 10 al 22 de marzo de 2014 en Punta del Este, Uruguay. El objetivo principal de la Escuela Latinoamericana de Educación, Ciencias Cognitivas y Neurales -LASchool- fue construir puentes entre estas disciplinas, con el propósito de fomentar innovaciones basadas en la ciencia en todas las áreas de la Educación. Fue un evento especial en muchos niveles, donde estudiantes y profesores trabajan juntos para explorar nuevas soluciones y generar propuestas de proyectos potencialmente relevantes para el desarrollo, diseño e implementación de prácticas educativas efectivas. El CIBPsi fue uno de los organizadores uruguayos junto con el Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea (Ceibal) y el Núcleo Interdisciplinario de Ciencias Cognitivas (NICC). La relación de las Ciencias Cognitivas uruguayas con los principales estudiosos regionales e internacionales como Sebastián Lipina, Marisa Carrasco, Mariano Sigman, Marcela Peña, Sidarta Ribeiro, Jacques Mehler†, Stanislas Dehaene, Kathryn Hirsch-Pasek, Elizabeth Spelke, Sidney Strauss, Manuel Carreiras, entre otros, fueron impulsadas por este evento.

The **4th Latin American School for Education, Cognitive and Neural Sciences** was held on March 10-22, 2014 in Punta del Este, Uruguay. The main objective of the Latin American School for Education, Cognitive and Neural Sciences -LASchool- was to build bridges between these disciplines, with the purpose of fostering science-based innovations in all areas of Education. It was a special event at many levels, where students and faculty work together to explore new solutions and generate project proposals potentially relevant for the development, design and implementation of effective educational practices. The CIBPsi was one of the Uruguayan organizers with the Plan Ceibal and the Núcleo Interdisciplinario de Ciencias Cognitivas (NICC). The relation of the Uruguayan Cognitive Sciences with main regional and international scholars such as Sebastian Lipina, Marisa Carrasco, Mariano Sigman, Marcela Peña, Sidarta Ribeiro, Jacques Mehler†, Stanislas Dehaene, Kathryn Hirsch-Pasek, Elizabeth Spelke, Sidney Strauss, Manuel Carreiras, between others, were impulsed by this event.

EEG signatures of elementary composition: disentangling genuine composition and expectancy processes

Emilia Fló[1,2], Álvaro Cabana[1,3], Juan C. Valle-Lisboa[1,3]

1**Centro de Investigación Básica en Psicología (CIBPsi)**, Facultad de Psicología, Universidad de la República. Montevideo, Uruguay
2 Laboratorio de Neurociencias, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay
3 Instituto de Fundamentos y Métodos, Facultad de Psicología, Universidad de la República. Montevideo, Uruguay

Álvaro Cabana https://orcid.org/0000-0002-8637-290X, Emilia Fló https://orcid.org/0000-0002-1986-2802, Juan Valle-Lisboa https://orcid.org/0000-0002-5662-5937

Author note. Juan Valle Lisboa is now at Sección Biofísica y Biología de Sistemas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.

Correspondence concerning this manuscript should be addressed to **eflo@fcien.edu.uy** (Emilia Fló), **acabana@psico.edu.uy** (Álvaro Cabana), or (Juan C. Valle-Lisboa). Juan Valle-Lisboa, Facultad de Ciencias, Iguá 4225, Montevideo 11600, Uruguay.

This is a free version -with the approval of the authors- of the pre-print of the manuscript Fló, E., Cabana, A. & Valle-Lisboa, J. C. (2020). EEG signatures of elementary composition: Disentangling genuine composition and expectancy processes. *Brain and Language, 209*, pp. 104837. Available at **https://doi.org/10.1016/j.bandl.2020.104837**

Abstract

We present an adaptation of Bemis & Pylkkänen's (2011) two-word, composition-list paradigm to study elementary composition in Spanish using electroencephalography. Our main objective is to determine if EEG is sensitive enough to detect a composition-related activity and to evaluate whether the expectancy of participants to compose contributes to this signal. Although we found relevant activity at the expected channels and times, we also found putative composition-related activity before the second word onset. Using a threshold-free cluster permutation analysis combined with linear models we show a task progression effect for the composition task that is not present for the list task. In a second experiment we evaluate two-word composition while avoiding differential expectancy effects across stimuli by incorporating all conditions to a single task. In this case, we failed to find any significant composition-related activity. We suggest that the composition-related activity measured with EEG may be at least in part carried by expectancy processes arising from the block design of the experiment. We conclude that for this paradigm it is important to complement the block based experiments with those controlling for expectancy, and to use powerful data-driven techniques to analyze the data

Keywords: Linguistic composition; EEGExpectancy effects; Trial effects; Cluster based permutation

Introduction

The ability to combine words in order to represent and convey new meanings is a funda mental operation in the comprehension and production of language (Martin & Baggio, 2020). One of the challenges in language research is to account for the functional neuroanatomical basis that underlie these processes (Friederici, 2017; Hagoort et al., 2009). More specifically, the challenge is to understand how the meaning of individual words is combined into complex meaning representations.

Linguists recognize that different levels of knowledge are involved in the production or understanding of an utterance. In particular, syntax, semantics and world knowledge interact strongly when building the meaning of an expression (Hagoort, 2019). Research on the brain basis of language has attempted to disentangle these processes by designing tasks in which only one of these levels is varied (Snijders et al., 2009; Friederici et al., 2000; Humphries et al., 2005; Mazoyer et al., 1993; Kutas & Hillyard, 1980). Nevertheless, on many occasions these approaches involve elaborate stimuli which elicit neural responses related to general performance. Additionally, it is questionable whether they manage to successfully isolate a specific linguistic processes. Furthermore, these paradigms do not usually shed light on the computations that occur at every step as words are combined to form a unified concept. According to Pylkkänen and collaborators, in order to determine the computational contributions of each type of processing, it is crucial to empirically characterize composition at its most basic level as a starting point from where to refine the study of syntactic and semantic computations (Pylkkänen et al., 2011; Poeppel et al., 2012).

Elementary composition

Inspired by the classical sentences versus list of words tasks, Pylkkänen's research group (Bemis & Pylkkänen, 2011, 2013c,a,b; Westerlund & Pylkkänen, 2014; Pylkkänen et al., 2014) designed a paradigm based on two word modifier-noun phrases. The initial study of this series (Bemis & Pylkkänen, 2011) introduced a simple paradigm to evaluate composition at its minimum by restricting stimuli to pairs of adjectives and nouns, and comparing subject's brain response to a non-word-noun control condition. Furthermore, a list task was implemented in which noun-noun conditions and non-word-noun conditions were presented to subjects. The rationale behind the experimental design and data analysis was to find brain regions for which there was a difference in activity between the two-word condition and the one-word condition in the composition task, and no (or a smaller) difference between conditions in the list task. By combining task and number of words in their design, they argue that any difference between conditions in the composition task that is not present in the list task cannot be due to an effect of word number. Applying an hypothesis driven cluster permutation analysis on MEG recordings, they report that the left anterior temporal lobe (LATL) and the ventromedial prefrontal cortex (vmPFC) show an increased response only for nouns preceded by adjectives. Interestingly, these responses develop early: activity in LATL occurs 184 - 255 ms after noun-onset, and the vmPFC response was reported 331 - 480 ms post-stimulus. The authors interpret LATL and vmPFC activities as a reflection of the combinatorial processes elicited by the binding of adjective-noun stimuli. This paradigm has been adapted to study semantic and syntactic operations in MEG and fMRI studies (see Westerlund & Pylkkänen (2014); Zhang & Pylkkänen (2015); Ziegler & Pylkkänen (2016); Zaccarella & Friederici (2015); Zaccarella et al. (2017)), and although other regions have shown activation for nouns in elementary combinatorial contexts (Bemis & Pylkkänen, 2011, 2013a; Pylkkänen et al., 2014), the most consistent brain area eliciting activation across studies is the LATL. This region's involvement in composition is supported by a series of fMRI studies on conceptual combination (Baron et al., 2010; Baron & Osher son, 2011; Coutanche & Thompson-Schill, 2015). It has been postulated that the LATL is involved not in syntactic or semantic computations per se, but in the addition of conceptual features in order to construct complex conceptual representations (Poortman & Pylkkänen, 2016; Patterson et al., 2007; Ralph et al., 2016). In this line, the early LATL response obtained under the two-word paradigm in MEG experiments has shown to be sensitive to the characteristics of the concepts combined such that the specificity of both head and modifier modulate this signal (Westerlund et al., 2015; Zhang & Pylkkänen, 2015; Pylkkänen, 2019, 2020). In addition, this activity is elicited by the composition of complex numbers but not for two-word numerical phrases, (Prato & Pylkkänen, 2014; Blanco-Elorrieta & Pylkkänen, 2016), and by the addition of semantic features to an individual representation and not to multiple entities (Poortman & Pylkkänen, 2016). Therefore this basic composition-related activity seems to reflect LATL's involvement in conceptual composition.

Elementary composition measured by EEG

When neural activity occurs in a synchronous manner across a great number of neurons, the coherent field potentials and local magnetic fields produced become big enough to be picked up by EEG and MEG, respectively. Even though there is great overlap between the information provided by these techniques they have some important differences. Firstly, because electrical conductivity varies across the layers of tissue that separate the cortical sources and the scalp, electrical signals are reduced and distorted. This has lower impact on MEG signal as magnetic permeability is more consistent (Okada et al., 1999). Furthermore, whereas EEG can detect both tangential and radial dipoles, MEG is more sensitive to source orientation. MEG is not able to capture neuronal currents oriented radially as they do not generate a magnetic field outside the head (Williamson & Kaufman, 1981; Malmivuo & Plonsey, 1995), however sources meeting this criteria have been shown to correspond with relatively small regions of the cortex located at the crests of gyri (Hillebrand & Barnes, 2002). EEG source reconstruction requires numerous electrodes, knowing their positions with precision, having head shapes measurements and accurate estimates of tissue conductivities (Michel et al., 2004). As magnetic fields are not distorted by the different brain structures, it is typically considered that MEG is better suited to resolve source localization. Importantly, both techniques have millisecond resolution and therefore are useful when the temporal course of a neural process is of interest (Okada et al., 1999; Luck, 2005).

In spite of MEG and EEG similarities, few studies have attempted to obtain an EEG marker of elementary composition following the two-word, composition-list paradigm established in the MEG literature. To our knowledge, an adaptation to study syntactic composition was implemented in Segaert et al. (2018) by measuring oscillatory changes in brain activity elicited by the syntactic binding of pairs of pseudowords, and only one study carried out a replication of Bemis & Pylkkänen initial study. Neufeld et al. (2016) performed a classical event related potential (ERP) analysis on EEG data obtained under this task. These authors selected a time window where the effect is expected according to the original study, and averaged the results over anterior and posterior electrodes. Their electrode grouping was motivated by trying to relate the combinatorial activity with classical N400 context effects. Although they found a broadly distributed negative composition-related activity in EEG recordings, they did not find a significant interaction between task (composition, list) and number of words (two-word, one-word), rendering their results inconclusive. Interestingly, their preprocessing approach enabled the authors to test for differences in the time window preceding the critical noun. A classical analysis on a visually selected time interval showed a difference between the composition task conditions before composition could have been achieved. They interpret this precombinatorial activity as reflecting the building of a syntactic structure to allocate the incoming noun.

The presence of a process taking place before the onset of the second word only in combinatorial contexts, points to an alternative interpretation of the experimental results. Given that the experiments in Neufeld et al. (2016) and Bemis & Pylkkänen (2011) were both based on a block design, it is possible that the structural properties of the stimuli in the different blocks influence to some extent their result. In a two-word trial during the composition task, the first word indicates with certainty to subjects that they would have to combine both items. Therefore, the processing of the first word was contingent on second word processing, and this was not the case during the two-word condition in the list task. Accordingly, it is possible that the activity identified as composition-related is conflated with anticipatory processes (c.f. Molinaro et al. (2012)).

Expectancy and task progression effects

A well-studied electrophysiological signature of expectancy-based processes is the Contingent Negative Variation (CNV), a slow negative event-related potential (ERP) associated to expectancy and prediction, which develops in a gradual manner during the realization of a task (Walter et al., 1964). This complex expectancy wave is modulated by task demands (Jacobson & Gans, 1981; Tecce et al., 1976; Rebert et al., 1967), sensory processing (Loveless, 1975; Gaillard, 1976) and motor preparation (Irwin et al., 1966; Brunia & Vingerhoets, 1981; Rohrbaugh & Gaillard, 1983). CNV is reported in the literature as a negative potential that increases as the contingency between two stimuli is learned (Walter et al., 1964; Poon et al., 1974; Proulx & Picton, 1980; Cohen, 1969; Hillyard, 1969), with a post-learning behavior that is task-dependent (Donald, 1980). Besides CNV, task progression effects have also been reported for classical ERPs such as the P300 elicited during oddball paradigms, which shows a behavior consistent with habituation effects (Ravden & Polich, 1998; Polich, 1989; Barry et al., 2019). Likewise, N400 amplitude decreases during word repetition tasks (Bentin & McCarthy, 1994; Ströberg et al., 2017).

This phenomenon challenges the ordinary analysis followed for ERPs, as averaging across trials neglects the possibility that the cognitive process of interest may vary as task unfolds. Furthermore, the prevalent averaging procedure reduces data to one data point per subject per condition, restraining the analysis to rigid ANOVAs. In contrast, modelling between trial variation and using mixed regression or non-parametric analyses allows the possibility to evaluate task progression effects (Vossen et al., 2011; Brush et al., 2018; Volpert-Esmond et al., 2018).

The present study

The purpose of the present study is to investigate whether EEG is sensitive to detect a composition-related activity elicited by a Spanish adaptation of Bemis & Pylkkänen's paradigm, and to separate a genuine composition from other non-specific expectancy-related responses. To this end, we carried out an adaptation of Bemis and Pylkkänen original experiment (Bemis & Pylkkänen, 2011) to Spanish while recording EEG activity (Experiment 1). Importantly, we used three methods to analyze the EEG data. In the first place, to be able to compare our results with the existent EEG replication (Neufeld et al., 2016), we followed their traditional ERP analysis. Following the original Bemis & Pylkkänen study, we performed an adapted version of their cluster permutation method on both ERP and time-frequency power representations. Finally, as expectancy processes change across trials and in order to avoid bias introduced by parameter selection, we implemented a threshold free cluster permutation analysis that included trial number as a predictor variable to test for task progression effects on EEG activity. If the reported activity in Neufeld et al. (2016) is in fact related to composition, we would expect it to be also present during a task in which participants cannot anticipate if a given word will have to be used to perform composition. Conversely, if this activity is not elicited it would suggest the alternative hypothesis, that this brain activity is at least in part due to anticipatory processes generated by task demands. Therefore, to examine this possibility we introduce a novel task to evaluate two word composition that avoids differential expectancy effects across conditions (Experiment 2).

Experiment 1

Materials and methods

Participants Twenty-nine non-colorblind Uruguayan undergraduate students participated in this experiment (22 female, average age, 25.31 ± 0.56). All subjects were native Rioplatense Spanish speakers, right handed, and had normal or corrected vision. The experiments were approved by the ethics committee of the Facultad de Psicología, Universidad de la República. All participants gave informed consent and were not awarded any economic or academic retribution, according to the national established guidelines (Decree Nº379/008).

Experimental design The original paradigm to study elementary composition included two tasks: a composition and a list task. The composition task consisted of a one-word condition in which subjects were presented with a non-word followed by a word denoting a noun (*xkq boat*), and a two-word condition in which subjects saw two words, an adjective followed by a noun (*red boat*). After each condition subjects had to match the presented verbal stimulus to an image. In the list task, subjects were presented with a one-word condition exactly the same as in the composition task, and a two-word condition in which two consecutive nouns were displayed (*cup boat*). Subjects had to indicate whether a subsequent picture matched any preceding word. In contrast to English, in Spanish adjectives such as color, shape or evaluation tend to be used in a post-nominal way, and their pre-nominal uses are mainly restricted to literary resources as is the use of epithets (Bosque & Demonte, 1999). Taking this into account we adapted the composition task so that subjects were presented with *noun adjective* (NA) trials in the two-word condition and *non-word adjective* (XA) trials in the one-word condition. We also devised two versions of the list task. In the list of adjectives task, subjects saw either *adjective adjective* (AA) trials (two-word condition) or *non-word adjective* (XA) trials (one-word condition). In this task the critical second word was also an adjective (*noun adjective VS. adjective adjective*), matching the composition task. However, using a list of adjectives could be problematic as color adjectives are typically more abstract than regular nouns. Hence, subjects were presented with a list of nouns task in which they saw *noun noun* (NN) trials and *non-word noun* (XN) trials in the two-word and one-word conditions, respectively. During the composition task participants were instructed to answer whether the image matched the preceding words. In contrast, in the list tasks subjects had to answer whether the image matched any of the preceding words. Half of the participants started the experiment with the composition task, and half started with the list tasks. The list task order and yes/no response hand was counterbalanced across subjects. Subjects were encouraged to answer as quickly and accurately as possible. Each task consisted of 200 trials that were preceded by a 40 trial practice to ensure that participants understood the task and learned the response keys. During each trial participants saw a fixation cross for 1 s and all stimuli except for the images were presented for 300 ms and were followed by a 300 ms blank screen. The images presented at the end of each trial remained on screen until subjects pressed a key or after 3 s (see Figure 1A). Following the end of each trial a sound was elicited to encourage participants to blink at that time. The inter-trial interval was randomly varied between 0.8 - 1.5 s. Stimuli presentation was coded in Psychopy (Peirce, 2007) and displayed on a CRT monitor with a 60 Hz refresh rate.

Stimuli construction In order to generate the stimuli, 11 nouns - *bote, cable, cepillo, globo, gorro, lápiz, reloj, teléfono, tenedor, tren, zapato* (boat, cable, thoothbrush, balloon, hat, pencil, clock, telephone, fork, train, shoe) and 11 color adjectives - *amarillo, azul, blanco, celeste, marrón, naranja, negro, rojo, rosado, verde, violeta* - (yellow, blue, white, sky-blue, brown, orange, black, red, pink, green, purple) were selected. Nouns and adjectives were matched for frequency (p = 0.059), number of letters (p = 0.51), number of substitution neighbors (p = 0.92), number of phonemes (p = 0.50), number of syllables (p = 0.58), number of homophones (p = 0.22), and number of phonological neighbors (p = 0.95). It was not possible to compare words on familiarity, imageability and concreteness as there was no available information for 7 out of the 11 color adjectives. Non-words (*brnlqs, slgrl, grsd, vrpng, jlcrfsmt, cxgnff, drbcw, tphn, dpjzb, pkrdt, vqdfnsm*) were constructed using Wuggy software (Keuleers & Brysbaert, 2010), and did not differ in number of letters from the adjectives (p = 0.40) nor from the nouns (p = 0.87). p-values correspond to a two tails t-test and properties were taken from Duchon et al. (2013).

For the composition task a python script was created to automatically select 10 nouns and 10 adjectives, and to combine them in order to generate 100 trials. For the one-word condition the 11 non-words and the 11 adjectives were combined, 21 items were randomly selected and discarded. This was done independently for every subject. Half of the trials for each condition were incongruent. In the two-word condition a trial was incongruent if the noun (25 trials) or the adjective (25 trials) did not match the image. In the one-word condition a trial was incongruent if the adjective did not match the image (50 trials). A second python script was created to generate the stimuli for the list tasks. For each subject for each list task, permutations of the 11 words were obtained and 10 items were randomly selected and discarded. For the one-word condition the 11 non-words and the 11 words (nouns or adjectives) were combined, 21 items were randomly selected and discarded. For the one-word condition of both tasks, a trial was incongruent if the image did not match the word (50 trials). For the two-word condition a trial was incongruent if the image did not match any of the preceding words (50 trials).

We evaluated the corpus bigram frequency of all the noun-adjective pairs using the Spanish corpus of Google Ngram (Michel et al., 2011) using the Phrasefinder API (https://phrasefinder.io/api). The Spanish corpus registers a total of 40.053.844 bigrams. Of the 121 possible noun-adjective pairs, 60 were not found in the corpus. The highest frequency was for the pair *lápiz rojo* (red pencil) with a frequency of 10579 instances (0.026%) and the lowest non-zero frequency was for the pair *gorro rosado* (pink cap) with a frequency of 49 (1.2×10^{-4}%). The overall mean relative bigram frequency was 8.9×10^{-4}%, 95% CI [3.9×10^{-4}%, 1.4×10^{-3}%]. None of the possible noun-noun pairs appear in the Google Ngrams Spanish corpus.

Data analysis

Behavioral data analysis Response times were measured from image onset until subjects pressed the no or yes keys. For each subject and task, reaction times were analyzed by removing incorrect and missing responses as well as trials in which response times were over or under two standard deviations. A linear mixed model with task as a three level factor (*composition - list of nouns - list of adjectives*) and number of words as a two-level factor (*two-word - one-word*) as fixed effects, and subject, first and second word as random intercepts was fitted to log-transformed reaction times, using the *lme4* R package (Bates et al., 2015; R Core Team, 2017). Accuracy data was analyzed using generalized linear mixed models using a logit link function with task as a three level factor (*composition - list of nouns - list of adjectives*), number of words as a two-level factor (*two-word - one-word*) and image congruency as a two-level factor (*congruent - incongruent*) as fixed effects, and random intercepts for each subject, first and second word. We used Wald χ^2tests to evaluate factor main effects and their interactions, while Wald Z tests were employed in pairwise comparisons. Bonferroni adjustments were used to correct for multiple comparisons.

EEG recording and preprocessing EEG signal was recorded using a Biosemi Active-Two system (Biosemi, B.V., Amsterdam, Netherlands). Sixty-four Ag-AgCl scalp electrodes where placed on a head cap following the location and label of the 10-20 system (Jasper, 1958). Ocular movements were monitored by 4 electrooculographic (EOG) electrodes (above, below the left eye, and on the outer canthi). The activity recorded was referenced online to the common mode sense (CMS; active electrode) and grounded to a passive electrode (Driven Right Leg, DRL), creating a feedback loop that drives the average potential of the participant to the AD-box reference potential. Data was digitized with a sample rate of 512 Hz with a fifth-order low-pass sinc filter with a -3 dB cutoff at 410Hz.

Data was preprocessed in MATLAB using fieldtrip toolbox (Oostenveld et al., 2011). Continuous data was two-pass filtered with a second-order high-pass Butterworth filter at 0.1 Hz and a fourth-order low-pass Butterworth filter at 30 Hz. Data was epoched 0.2 s prior to first word onset until 1.2 s (at image presentation). Epochs were baselined to activity 200 ms preceding first word onset. Noisy trials and channels were rejected following Junghöfer et al. (2000). Trials in which participants responded incorrectly were discarded. The remaining trials for each condition were averaged in order to obtain one ERP per subject per condition.

For the composition task the number of rejected electrodes and trials was 1.52 ± 1.16 and 23.19 ± 6.97, respectively. A t-test showed no difference in the number of trials rejected between conditions (p = 0.16). For the list of adjectives task the average of discarded electrodes was 2.17 ± 1.97 and the number of trials rejected was 21, 76 ± 6.80, there was no difference in the number of trials rejected between conditions (p = 0.12). For the list of nouns task the average of electrodes rejected corresponded to 1.55 ± 1.22 and the average for trials was 23.30 ± 7.22. No difference between conditions for the number of rejected trials was found (p = 0.96). A one-way ANOVA was conducted to compare the effect of task on number of channels rejected, no difference was found ($F(2, 78) = 1.86$, p = 0.16). Accordingly, a second one-way ANOVA showed no difference between task and number of discarded trials ($F(2, 78) = 0.008$, p = 0.99).

Subjects' ERP signal-to-noise ratio was evaluated following Parks et al. (2016). A lower bound of signal to noise ratio confidence interval (SNRLB) of 3.0 dB was used as threshold to ensure signal quality. For the composition task the SNRLB ranged from 3.54 to 21.87 dB (mean = 11.25, median = 10.26 dB, SD = 5.23 dB). The SNRLB for the list of adjectives task ranged from 0.15 to 16.10 dB (mean = 8.31 dB, median = 8.34 dB, SD = 3.41 dB). Subjects 13 and 19 failed to meet the SNRLB criterion. Finally for the list of nouns task the SNRLB was between 4.37 and 22.14 dB (mean = 11.76 dB, median = 10.46 dB, SD = 3.66 dB). In order to compare the data between tasks we excluded from the analysis the two participants that did not met the critical value in the list of adjectives task.

Cluster permutation analysis on ERPs We used a cluster permutation analysis to evaluate all electrodes and epoch data points from -0.2 s to 1.2 s (until image presentation). We used two-tailed t-tests to contrast the difference between the composition conditions (NA - XA) with the difference between the list of nouns conditions (NN - XN). Independently, we used the same approach to contrast differences between the composition (NA - XA) conditions with the difference between the list of adjectives conditions (AA - XA). This analysis aims at finding differences between the composition task conditions that are not present between the list task conditions. Additionally, as we decided to use two controls, conditions in the list of nouns and in the list of adjectives were also contrasted to test for possible differences.

In each analysis, a t-test was performed on every sample and t values were clustered depending on if they exceeded a dependent samples t-test threshold of $p < 0.05$ (two tailed). t values for each data point within each cluster were summed in order to obtain a value per cluster. The maximum negative and positive cluster statistic values were kept. This was done for 5000 permutations of the data resulting in a null hypothesis distribution of the statistic, against which we tested the real data. We considered the critical α level here to be 0.025. For all the analyses using this method on ERPs we set to three the minimum number of electrode neighbors that had to be significant for a given time point in order to be part of a cluster. All cluster permutation analyses were conducted on MATLAB using Fieldtrip toolbox (Oostenveld et al., 2011) and neighbors were defined following Biosemi 64 neighbors template.

Threshold free cluster permutation analysis In order to explore task progression effects on ERPs we performed threshold-free cluster enhancement (TFCE) (Mensen & Khatami, 2013; Smith & Nichols, 2009) analyses on t-values derived from linear models fitted to epoched EEG data. TFCE is a non-parametric cluster randomization method inspired by random field theory that uses an enhanced statistic designed to boost weaker but broadly distributed signals. In contrast to the cluster permutation method implemented in Fieldtrip (Oostenveld et al., 2011), this method does not have threshold as a free parameter, reducing the researcher degrees of freedom (Wicherts et al., 2016).

We fitted two linear models to each time point and electrode using data from both tasks combined (as in the previous analyses). A simple model contained parameters for trial number, task (composition vs. list), number of words (two-word vs. one-word), their interaction (task × number of words), and the interaction between task and trial number. This last parameter was introduced to test for different effects of task progression across tasks. A second model (the "complete" model) also included the interaction between number of words and trial number, and the triple interaction (task × number of words × trial number). Also, to further dissect the relation between task progression and experimental condition, we fitted follow-up models with trial number, number of words and their interaction as parameters for each task separately. For each time and electrode, t-values were obtained for each parameter estimate. Following Mensen & Khatami (2013), positive and negative t-values were separated and t_{tfce} statistics were computed. The maximum (in absolute value) t_{tfce} statistic was used in cluster randomization tests with 5000 permutations to obtain p-values for each parameter of each model.

Voltage correlation to reaction times In order to quantify the association between voltage and participants' reaction times we used the rmcorr R package to obtain repeated-measures correlation coefficients (Bakdash & Marusich, 2018). This analysis takes into account the fact that observations within participants are not independent and has greater statistical power as no averaging across individuals is performed, similar to linear mixed model approaches. We selected the electrode-time point pair for which the t_{tfce} statistic for the trial parameter was maximal and averaged voltage across a 100 ms time window centered on that point. Using this data, we obtained correlation coefficients between trial-to-trial voltage values and reaction times for the composition and lists tasks.

Experiment 2

Material and Methods

Participants Thirty-nine non-colorblind Uruguayan undergraduate students participated in this experiment (23 female, average age, 23.13 ± 3.65). All subjects were native Rioplatense Spanish speakers, right handed, and had normal or corrected vision.

Experimental design In this experiment we tried to maintain the design similar to Experiment 1 while trying to reduce possible confounding of expectancy in composition effects. During this task participants were presented with four different conditions: *noun adjective: NA*, *non-word adjective: XA*, *noun noun: NN* and *non-word noun: XN*. In all conditions, each word pair was followed by an image. Subjects were instructed to answer if the image matched the preceding verbal material and were encouraged to answer as quickly and accurately as possible. Hence, there were two-word conditions (NA and NN) and one-word conditions (XA and XN), and composition was required only for the NA condition. A representation of the trials is shown in Figure 5B. The main differences with the previous experiment were that subjects saw all types of trials in the same block, and the image presented in NN trials had two elements. The task consisted of 400 trials that were preceded by a 40 trial practice. In each trial, participants saw a fixation cross for 1 s and all stimuli except for the images were presented for 300 ms followed by a 300 ms blank screen. Every subject was presented with 100 trials of each condition, and half of the trials of every condition were incongruent to the image. For the XN and XA conditions, a trial was incongruent if the image did not match the noun or adjective respectively (50 trials for each condition). For the NA condition a trial was incongruent if the image did not match the noun (25 trials) or did not match the adjective (25 trials). Finally, for the NN condition a trial was incongruent if the image did not match the first noun (25 trials) or the second noun (25 trials).

In order to obtain results comparable to Experiment 1, we conceptually organized the aforementioned four conditions as two tasks with two conditions each. The composition task comprised of a *two-word* (NA) and a *one-word* (XA), condition, and a *list of nouns* task also comprised of a *two-word* (NN) and a *one-word* (XN) condition.

Stimuli construction The same pool of words and non-words used in the Experiment 1 was employed to create the stimuli. A python script was coded to combine words and non-words for each subject in order to create 400 trials. For the XN, XA and the NA conditions, a combination of the 11 elements of each pool was made, resulting in 121 pairings with the desired structure.

Subsequently, 21 items were randomly discarded. For the NN condition, permutations of the nouns were generated and 10 items were discarded, resulting in 100 pairings. Both the NN and the NA pairs have the same bigram frequency properties as the ones used in Experiment 1.

Data analysis

Behavioral data analysis Response times were measured from image onset until subjects pressed the no or yes keys. For each subject, reaction times were analyzed by removing incorrect and missing responses as well as trials in which response times were over or under two standard deviations. A linear mixed model with task as a two level factor (*composition* - *list of nouns*) and number of words as a two-level factor (*two-word* - *one-word*) as fixed effects, and subject, first and second word as random intercepts was fitted to log-transformed reaction times, using the lme4 R package (Bates et al., 2015; R Core Team, 2017). Accuracy data was analyzed using generalized linear mixed models using a logit link function with task as a two-level factor (*composition* - *list of nouns*), number of words as a two-level factor (*two words* - *one-word*) and image congruency as a two-level factor (*congruent* - *incongruent*) as fixed effects, and random intercepts for each subject and second word. We used Wald χ^2 tests to evaluate main effects and their interactions, while Wald Z tests were employed in pairwise comparisons. Bonferroni adjustments were used to correct for multiple comparisons.

EEG recording and preprocessing We followed the same recording and preprocessing steps implemented in the first ex

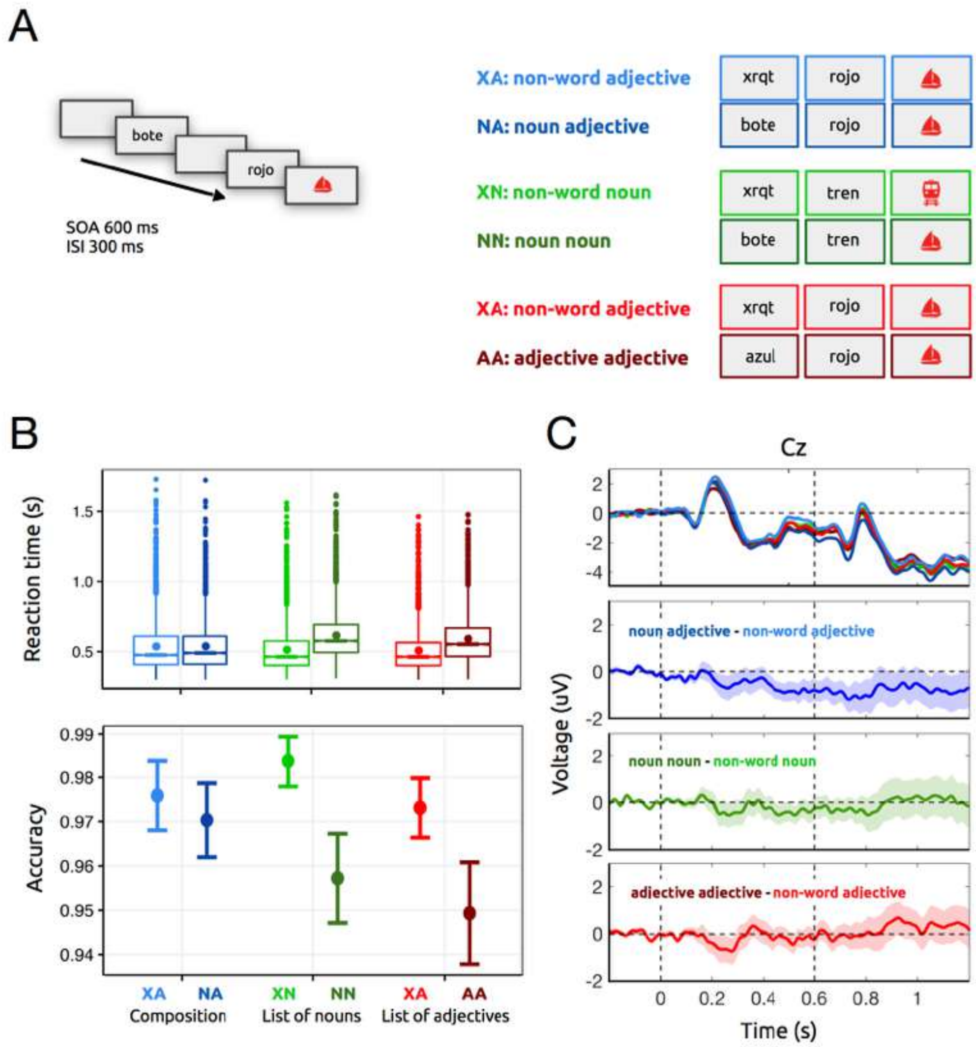


Figure 1: Experiment 1. Spanish adaptation of Bemis & Pylkkänen's (2011) task. A. Subjects were presented with three tasks: a composition task and two list tasks with two-word and one-word conditions. B. Reaction times and accuracy results. Bars indicate 95% confidence interval. C. Top graph shows ERP grand average for each condition of each task superimposed for Cz. Bottom graphs show mean differences between conditions for each task with 95% confidence interval for Cz. XA: non-word adjective, NA: noun adjective, XN: non-word noun, NN: noun noun, and AA: adjective adjective. Vertical dashed lines indicate first and second stimulus onset.

periment. The number of rejected electrodes and trials was 4.44 ± 2.21 and 55.75 ± 15.28, respectively. A one-way repeated-measures ANOVA was conducted to test for differences in the number of trials across conditions, no difference was found $F(3, 124) = 0.74$, $p = 0.53$. Subjects' ERP SNRLB ranged from 1.74 to 20.31 dB (mean = 10.99, median = 11.52 dB, SD = 4.78 dB). Subjects 3, 12, 14, 15, 26 and 29 failed to meet the SNRLB criterion and were excluded from the analysis.

EEG data analysis A cluster permutation analysis and a repeated-measures correlation analysis between trial-to-trial voltage values and reaction times were carried out following the procedures described for Experiment 1.

Further analyses In order to test for specific hypotheses and compare to published results we performed a classical ERP analysis of Experiment 1. Furthermore both experiments were submitted to time-frequency analysis coupled to a cluster permutation test. The details are described in the Supplementary Materials section.

Results

Experiment 1: Adaptation to Spanish of the original study

Behavioral results We found a significant main effect of task ($\chi^2(2) = 54.67$, $p < 0.001$) as well as an effect of number of words ($\chi^2(1) = 729.02$, $p < 0.001$). More interestingly, an interaction between task and number of words was also significant ($\chi^2(2) = 471.35$, $p < 0.001$). Pairwise comparisons showed no difference in reaction times between the two-word condition and the one-word condition for the composition task ($Z = -0.81$, $p = 0.42$). Contrarily, reaction times were smaller for the one-word condition for the list of nouns ($Z = -27.37$, $p < 0.001$) and for the list of adjectives ($Z = -22.89$, $p < 0.001$), compared to the two-word conditions. Participant's accuracy was in line with the reaction time results. There was a main effect of task ($\chi^2(2) = 11.03$, $p = 0.004$) and number of words ($\chi^2(1) = 33.66$, $p < 0.001$), as well as a congruency effect ($\chi^2(1) = 7.63$, $p = 0.006$) which was not further explored. The analysis shows a significant interaction between task and number of words ($\chi^2(2) = 11.44$, $p = 0.003$). Pairwise comparisons revealed no difference between conditions in accuracy for the composition task ($Z = 1.23$, $p = 0.22$) and a significant difference between conditions in the list of nouns ($Z = 5.48$, $p < 0.001$), as well as for the list of adjectives task ($Z = 4.26$, $p < 0.001$) (Figure 1B).

Cluster permutation analysis results The permutation cluster analysis to test the interaction between the list of adjectives and composition tasks and the number of words yielded a significant negative cluster comprised of 16 electrodes with a central-frontal distribution ($p = 0.007$, t = 1.01 - 1.20s) (Figure 2A; see Figure 1 in Supplementary Material for th topographical distribution of the cluster). A post-hoc analysis was carried out for each task taking only the electrodes and data points that participate in the interaction cluster. No cluster was obtained for the list of adjectives. We found a significant negative voltage cluster for the composition task ($p = 5.0 \times 10^{-4}$, t = 1.01- 1.20 s).

Supplementary Material for the topographical dis tribution of the cluster). The post-hoc analysis showed no difference between conditions for the list of nouns task, and two significant negative voltage clusters for the composition task ($p = 8.0 \times 10^{-4}$, t = 0.95 - 1.15 s and $p = 6.6 \times 10^{-3}$, t = 0.86 - 094 s). We also carried out the same analysis contrasting both list tasks. No significant clusters were obtained.

TFCE cluster permutation analyses and linear models For the sake of brevity we present results for the composition task and the list of nouns task, since results for both list tasks follow the same pattern.

Threshold-free cluster enhancements statistics were obtained

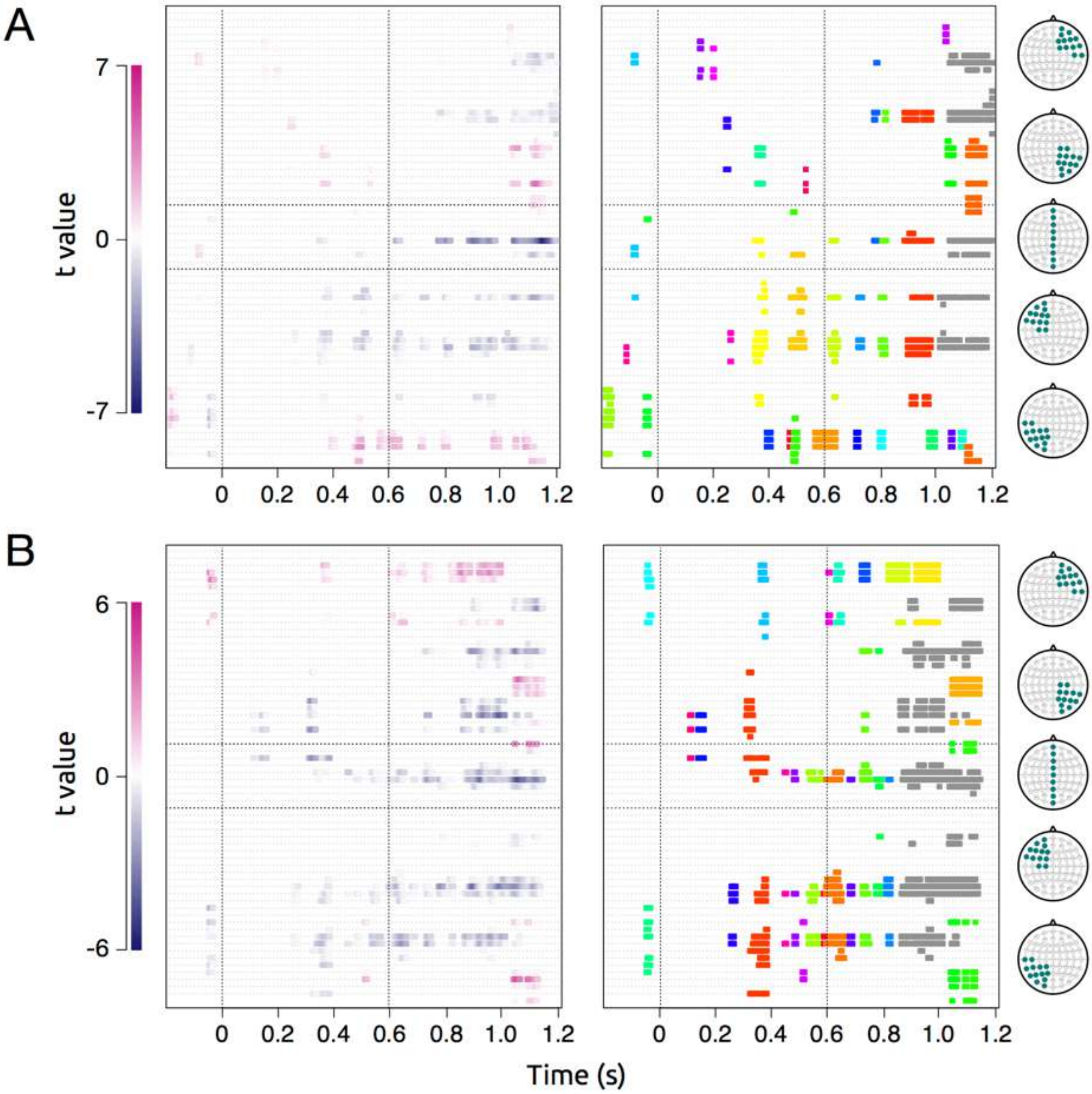


Figure 2: Cluster permutation analysis results for the interaction between task and number of words. A. Composition vs. list of adjectives. Left: t-values for the clusters obtained. Right: Electrode–time points clusters. Color code represents data points that participate in a given cluster. Only the gray cluster is statistically significant (p = 0.007, t = 1.01 - 1.20 s). B. Composition vs. list of nouns. Left: t-values for the clusters obtained. Right: Electrode–time clusters. Only the gray cluster is statistically significant (p = 0.006, t = 0.86 - 1.15 s).

The same analysis comparing the list of nouns and composition tasks showed similar results. A significant negative cluster was obtained for the interaction between task and number of words ($p = 0.006$, t = 0.86 - 1.15 s) composed of 26 electrodes with a central distribution (Figure 2B; see Figure 2 in

for two linear models fit ted to the EEG data, the simple and the complete model (See Table 1 in supplementary material).

For the simple model, we found significant effects of task (FCz, t = 0.423 s, $t_{tfce} = -2.68 \times 10^3$, $p = 0.027$), number of words (PO7, t = 0.503 s, $t_{tfce} = 8.55 \times 10^3$, $p < 0.001$), trial

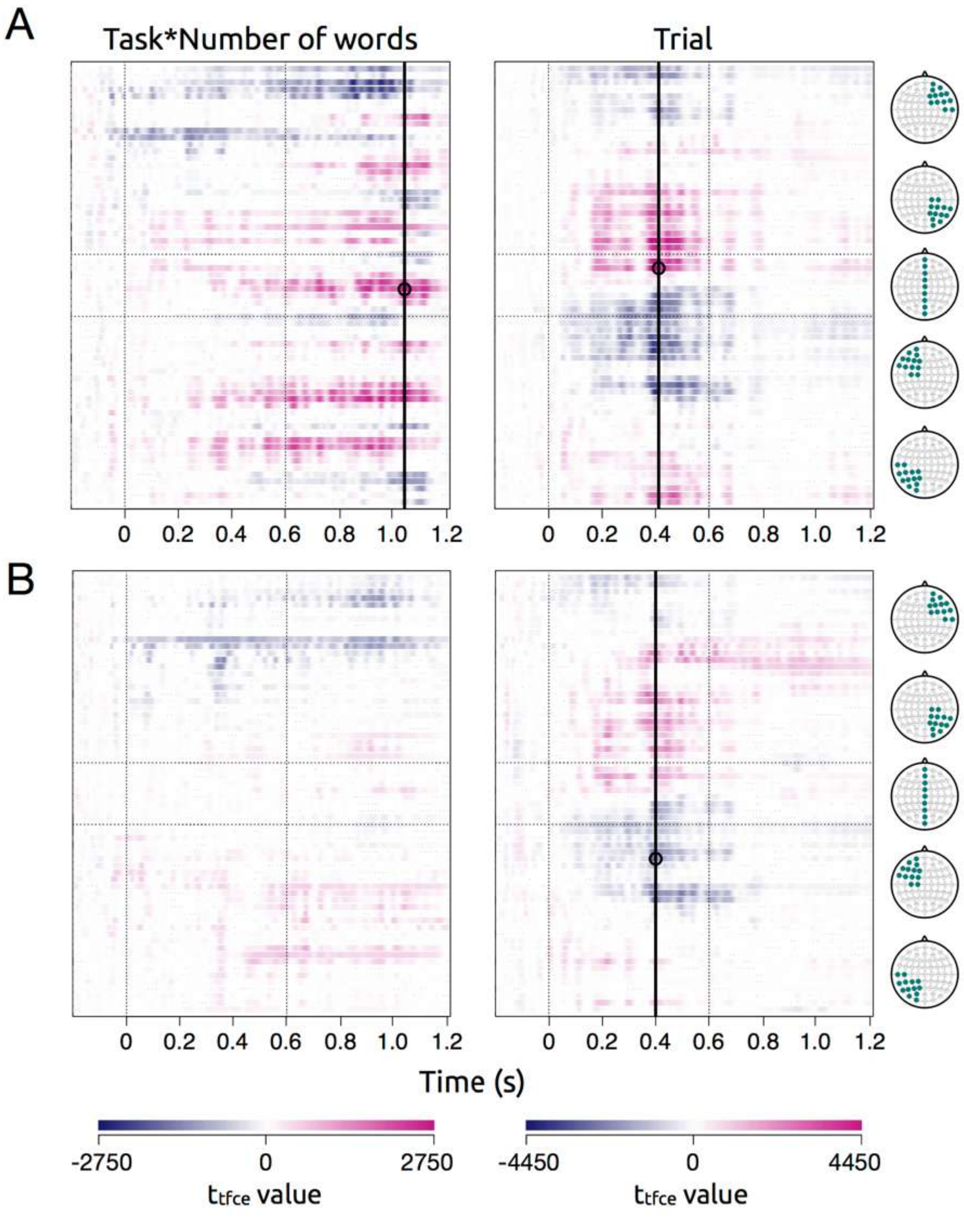


Figure 3: Threshold-free cluster enhancement results on t-values derived from linear models fitted to EEG data from experiment 1. A. Simple model t_{tfce} statistics for the interaction between task and number of words (left) and for trial (right). B. Complete model t_{tfce} statistics for the interaction between task and number of words (left) and for trial (right). Line and circle indicate time-electrode maximum statistic values that reached significance in permutation tests

number (Oz, t = 0.415 s, $t_{tfce} = 4.45 \times 10^3$, $p < 0.001$) (Figure 3A) and a significant interaction between task and number of words (Cz, t = 1.05 s, $t_{tfce} = 2.75 \times 10^3$, $p = 0.027$) (Figure 3A). Importantly, we found a significant effect for the interaction between task and trial number (F2, t = 0.460 s, $t_{tfce} = -2.46 \times 10^3$, $p = 0.043$).

The analysis of the complete model with all the interactions showed the following results. A significant task effect (F1, t = 0.417 s, $t_{tfce} = -1.82 \times 10^3$, $p = 0.032$), a significant effect of number of words (FC3, t = 0.398 s, $t_{tfce} = -2.84 \times 10^3$, $p = 0.024$), a significant trial number effect (F3, t = 0.404 s, $t_{tfce} = -2.82 \times 10^3$, $p = 0.026$) (Figure 3B, Figure 4A), no significant interaction between task and number of words (C6, t = 0.374 s, $t_{tfce} = -1.44 \times 10^3$, $p = 0.49$) (Figure 3B), no significant interaction between task, number of words and trial number (Iz, t = 0.915, $t_{tfce} = -1.26 \times 10^3$, $p = 0.67$), no significant interaction between task and trial number (F1, t = 0.398s, $t_{tfice} = 1.41 \times 10^3$, $p = 0.52$) and no significant interaction between number of words and trial number (Iz, t = 0.899 s, $t_{tfce} = 1.72 \times 10^3$, $p = 0.30$).

For each task a model was fitted to compare the one-word

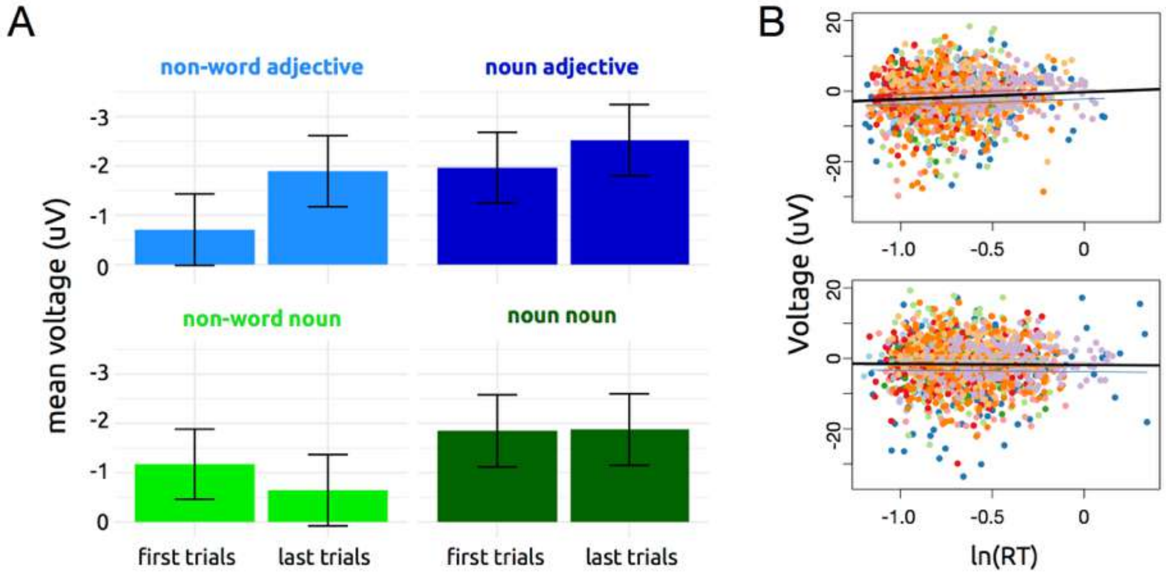


Figure 4: Effect of task progression on the evoked response and reaction times for the composition and list of nouns tasks. Voltage values correspond to the average in a window centered at the maximum value for the trial statistic (F3, 352–452 ms). A. Voltage values for the first and last 33 trials of each condition for each task were averaged. There is an increase in negative voltage as the composition task unfolds. Composition task conditions are shown in shades of blue, list of nouns conditions in shades of green. Error bars represent 95% confidence intervals. B. Correlations between voltage and reaction times for the composition task (top) and the list of nouns task (bottom).

and two-word conditions. For the composition task an effect of number of words (FC1, t = 0.439 s, $t_{tfce} = -2.52 \times 10^3$, $p = 0.020$), an effect of trial (F3, t = 0.402 s, $t_{tfce} = -2.68 \times 10^3$, $p = 0.013$), and no effect for an interaction between number of words and trial (Iz, t = 0.888 s, $t_{tfce} = 1.50 \times 10^3$, $p = 0.31$) was found.

For the list of nouns task an effect of number of words (POz, t = 0.499 s, $t_{tfce} = 2.23 \times 10^3$, $p = 0.020$), no effect of trial (C6, t = 0.723 s, $t_{tfce} = 1.58 \times 10^3$, $p = 0.20$), and no effect for an interaction between number of words and trial (F5, t = 0.061 s, $t_{tfce} = -1.26 \times 10^3$, $p = 0.49$) was obtained.

Correlation between voltage and reaction times For the composition task, lower reaction times correlated with more negative potentials (F3, rrm (1436) = 0.055, 95% CI [0.004, 0.107], $p = 0.036$). This correlation was not found for the list of nouns (F3, rrm (1428) = -0.015, 95% CI [-0.067, 0.037], $p = 0.56$) (Figure 4 B).

Classical ERP and Time-Frequency analysis In the Supplementary Materials we present the results for the classical ERP and the time-frequency analyses. As in Neufeld et al. (2016), we found no interaction between task and number of words, although post-hoc analyses yielded a difference between conditions for the composition task and no difference between conditions for the list tasks. No significant differences were found for the composition task in the time points evaluated before second word onset. Finally, we found no significant clusters for the interaction between task and number of words in any of the studied power bands (gamma, alpha and beta bands).

Experiment 2. Behavioral Results For reaction times, a main effect of task was found ($\chi^2(1) = 458.36$, $p < 0.001$) as well as an effect of number of words ($\chi^2(1) = 546.51$, $p < 0.001$). Moreover, the interaction between task and number of words

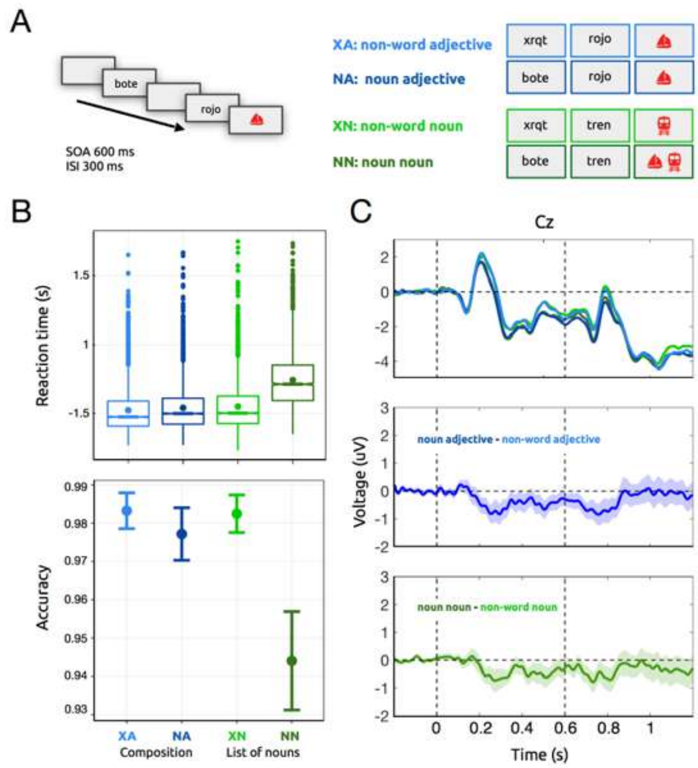


Figure 5: Experiment 2. Composition without expectancy. A. Participants were presented with two two word conditions: noun adjective (NA) and noun noun (NN) conditions, and two one-word conditions: non-word adjective (XA) and non-word noun (XN) conditions. B. Reaction times and accuracy results. Bars indicate 95% confidence interval. C. Top graph shows ERP grand averages for each condition at Cz. Bottom graphs show mean differences between conditions with 95% confidence intervals. Vertical dashed lines indicate first and second stimulus onset.

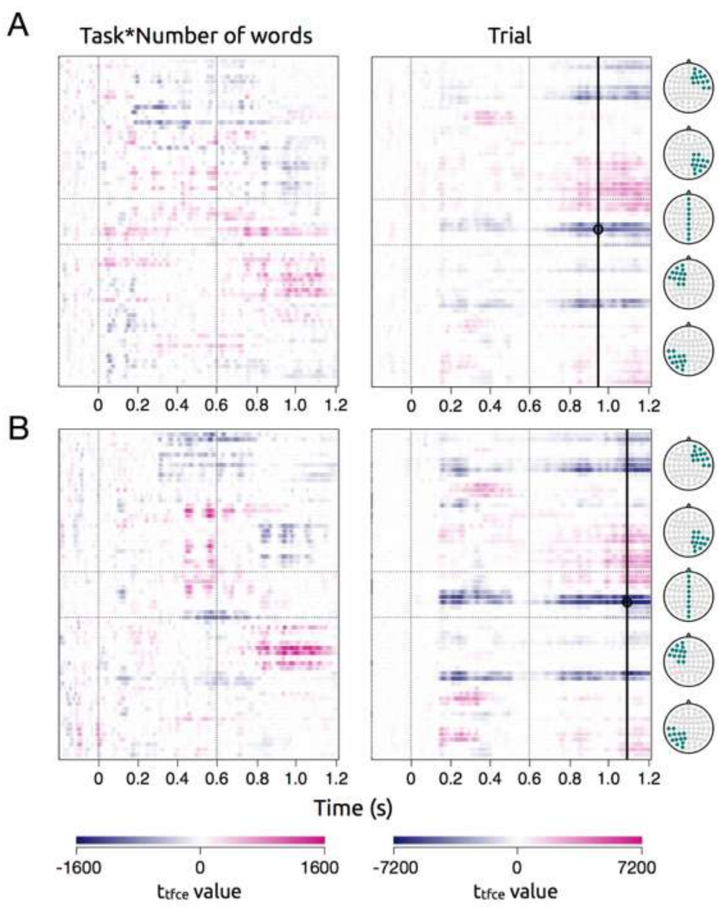


Figure 6: Threshold-free cluster enhancement results on t-values derived from linear models fitted to EEG data from experiment 2. A. Simple model ttf ce statistics for the interaction between task and number of words (left) and for trial (right). B. Complete model ttf ce statistics for the interaction between task and number of words (left) and for trial (right). Line and circle indicate time-electrode maximum statistic values that reached significance in permutation tests.

was significant ($\chi^2(1)$ = 1284.94, p < 0.001). Pairwise comparisons were tested between number of words for each group. The two-word and one-word conditions in the list of nouns task were different ($Z = -37.04$, $p < 0.001$). Furthermore, there was a significant difference between conditions for the composition group as well ($Z = -4.10$, $p < 0.001$) (Figure 5B).

Cluster permutation analysis results To evaluate an interaction between number of words (two-word – one-word) and task (composition – list of nouns) we carried out a cluster permutation analysis on the EEG data averaged for each subject. The cluster permutation analysis yielded 9 positive and 8 negative clusters; however, none of them reached significance ($p > 0.64$). To further explore this result, we proceeded to do a post-hoc analysis for the difference between each pair of conditions. The comparison between the composition task conditions yielded: three significant positive voltage clusters (1) from 0.617 to 0.836 s, $p < 0.001$, (2) from 0.336 to 0.397 s, $p = 0.006$ and (3) from 0.217 to 0.279 s, $p = 0.007$; and two significant negative voltage clusters: (1) from 0.195 to 0.434 s, $p < 0.001$, (2) from 0.621 to 0.859 s, $p < 0.001$. The analysis for the list of nouns conditions gave a comparable result: two significant positive voltage clusters (1) from 0.314 to 0.533 s, $p < 0.001$ and (2) from 0.555 to 0.756, $p < 0.001$. Furthermore, three significant negative voltage clusters were obtained: (1) from 0.199 to 0.543 s, $p < 0.001$, (2) from 0.625 to 0.758 s, $p = 0.002$ and (3) from 0.762 to 0.843 s, $p = 0.008$ (see Figure 3 in Supplementary Material).

TFCE cluster permutation analyses and linear models In order to keep analyses similar to Experiment 1, we fitted two linear models to each time-electrode point in epoched EEG data, and subjected each parameter to the TFCE procedure (see Table 1 in supplementary material). The simple model yielded no significant effect of task (FCz, t = 0.872 s, $t_{tfce} = -1.37 \times 10^3$, $p = 0.48$), a significant effect of number of words (PO8, t = 0.267 s, $t_{tfce} = 9.07 \times 10^3$, $p < 0.001$), a significant trial number effect (FCz, t = 0.964 s, $t_{tfce} = -4.89 \times 10^3$, $p < 0.001$) (Figure 6A), no significant interaction between task and number of words (FC6, t = 0.527 s, $t_{tfce} = -955.2$, $p = 0.87$) (Figure 6A) and no significant effect for the interaction between task and trial (PO4, t = 0.950 s, $t_{tfce} = -1.42 \times 10^3$, $p = 0.49$).

The complete model with all the trial interactions showed the following results. We found no significant task effect (P7, t = 0.911 s, $t_{tfce} = 2.21 \times 10^3$, $p = 0.33$), a significant effect of number of words (P8, t = 0.269 s, $t_{tfce} = 1.70 \times 10^4$, $p = 0.001$), a significant trial number effect (FCz, t = 1.08 s, $t_{tfce} = -7.19 \times 10^3$, $p < 0.001$) (Figure 6B), no significant interaction between task and number of words (F7, t = 0.950 s, $t_{tfce} = 1.52 \times 10^3$, $p = 0.30$) (Figure 6B), no significant interaction between task, number of words and trial number (CP4, t = 0.572, $t_{tfce} = -1.24 \times 10^3$, $p = 0.59$), no interaction between task and trial number (P7, t = 0.007s, $t_{tf\,ice} = 1.28 \times 10^3$, $p = 0.62$) and no effect between number of words and trial number (PO8, t = 0.972 s, $t_{tfce} = -1.59 \times 10^3$, $p = 0.25$)

Correlation between voltage and reaction times We found a significant positive correlation between the average voltage for FCz between 950 - 1050 ms (t_{tfce} maximum statistic was obtained at around 1000 ms on that electrode) and reaction times on a trial-to-trial basis. Lower reaction times correlated with more negative potentials (FCz, rrm (10376) = 0.069, 95% CI [0.050, 0.088], $p < 0.001$) across conditions.

Time-frequency analysis In the Supplementary Materials we present the results of the time-frequency analysis. We found no significant clusters for the interaction between task and number of words for the gamma, alpha and beta bands.

In regard to accuracy, there was a main effect of task ($\chi^2(1)$ = 11.78, p < 0.001) and number of words ($\chi^2(1)$ = 47.32, p < 0.001), as well as a congruency effect ($\chi^2(1)$ = 42.88, p < 0.001). The analysis showed a significant interaction between task and number of words ($\chi^2(1)$ = 15.86, p < 0.001). Pairwise comparisons revealed no difference between conditions for the composition task (Z = 1.91, p = 0.057) and a significant difference between conditions in the list of nouns (Z = 8.36, p < 0.001) (Figure 5B).

Discussion

In this work we have adapted Bemis & Pylkkänen's (2011) experimental paradigm in order to study elementary language composition in Spanish using EEG. As the original experiment was done in MEG and with English stimuli, our adaptation required many changes. In contrast to English, canonical adjectivation of a noun in Spanish is done post nominally, this imposed the need to add a second control task to keep the critical word type consistent in both the composition and the list task. Furthermore, we adapted the analysis to EEG data, and used two different cluster finding techniques.

Despite these important modifications, we found comparable results to those of the original study. The binding of two items into a single concept yielded a processing advantage, as shown by faster reaction times and a lower error rate for the two-word condition of the composition task compared to conditions in which composition was not possible. Our cluster permutation analysis comparing brain responses to the composition task and the list of nouns task, showed a significant interaction between task and number of words 260 - 550 ms after second word onset. This effect was driven by a difference between two-word and one-word conditions in the composition task. A similar result in a later time window (410 - 600 ms) was found when comparing the composition and the list of adjectives tasks. Although consisting of different word classes (nouns and adjectives), the clusters obtained in both comparisons showed a similar temporal and topographical distribution. Furthermore, no difference was found between the list tasks, suggesting that both tasks were equally appropriate as controls. This results were not accompanied by differences in the non-phase locked activity as no modulation of power was found for gamma, alpha and beta bands (see Supplementary Material for analyses and results).

To the best of our knowledge this is the first study to show an EEG signal temporally consistent with the MEG composition-related activity using an unbiased rigorous cluster permutation analysis and Spanish stimuli. Although a previous study using EEG was pub lished (Neufeld et al., 2016), the reported composition-related activity is not supported by the interaction analysis between task and number of words, leaving open the possibility that their result is indexing the presence of two words over one word, and not a genuine com position activity. This inconclusive result may be a consequence of the electrode partition choice and the time window targeted.

Interestingly, Neufeld et al. (2016) reported a difference between the composition task conditions before second word onset. The authors interpret this pre-combinatorial activity as a syntactic building process, arguing that the adjective-noun syntactic structure is initialized before second word onset to allocate the expected noun. Even though carrying the same classical ERP analysis on our data did not yield the exact same results (see Supplementary Material), our cluster permutation analysis output (Figure 2A and Figure 2B) shows that the same electrodes that comprise the significant clusters seem to also take part in smaller clusters before second word onset (this is clear for anterior and posterior left hemisphere electrodes). Although Neufeld and collaborators' interpretation is reasonable, we propose an alternative explanation. For all tasks and conditions, after seeing the first word participants knew with certainty what type of word would follow. Nevertheless, it was only the case for the composition task that subjects had to manipulate the first word in relation to the second word. In this way, participants' processing of the first word was conditional to the second word, and a contingency between the first and second stimuli had to be established only during the composition task. This pre-composition activity could then be more related to general expectancy rather to a specific syntactic mechanism.

Given that the experimental design imposes the need to establish a relation between the stimuli, CNV is a good candidate explanation for our results. In particular, this component's amplitude is sensitive to attentional demands (Low et al., 1967; Rebert et al., 1967; Jacobson & Gans, 1981; Simons et al., 1979; Tecce & Scheff, 1969) and task progression (Walter et al., 1964), and it has been consistently shown that CNV increases as the contingency between two stimuli is learned (Proulx & Picton, 1980; Cohen, 1969; Hillyard, 1969). Consistent with CNV behavior, a TFCE permutation analysis allowed us to evidence the presence of a task progression effect specific to the composition task, i.e. amplitude increased on a trial-to-trial basis (Figure 4A). The maximum effect of trial number occurred before second word onset and was independent of condition. Moreover, a more exhaustive model including all interac tions between predictor variables rendered the crucial interaction between number of words and task non-significant. This results suggest that the composition-related activity obtained in our EEG experiment is contaminated by an anticipatory process. In agreeement with the electrophysiological response described above, we found a correlation between the voltage amplitude for each trial at the location of maximum effect of trial number (as indicated by the TFCE analysis) and the participants's response times (Figure 4B), such that response times are lower as voltage values get more negative, suggesting its involvement in response preparation. Importantly, this correlation was significant only for the composition task. These results suggest that the composition task shows electrophysiological and behavioral patterns compatible with non-stationary and learning effects which are not present for the list task, indicating that the tasks differ not only in the composition requirement. Hence, the crucial manipulation designed to identify a neural signature of composition may not allow to separate an elementary composition activity from an expectancy-based response in EEG.

In order to test this hypothesis we designed an experiment to engage participants in combining two words, but ensuing that expectancy processes affect both tasks similarly. For this, conditions were grouped in a single block such that each two-word trial started with a noun, and an adjective or another non-composable noun could follow. In this manner, participant's expectancy would be equally affecting both tasks, as on any given trial a noun gave no indication of whether the subsequent word would be a composable item or not. Behavioral data indicates that subjects were unifying both elements when a noun was followed by an adjective, as reaction times were similar to the one-word condition and lower than the two nouns condition. However, a cluster permutation analysis on ERP data yielded no interaction between task and number of words. Similarly, no differences in the frequency domain were found (see Supplementary Material). The comparison of each two-word condition to its one-word control shows very similar temporal and topographical distributions (see Fig. 3 in Supplementary Material). After the initial noun is presented, a negative activity starting before second word onset develops and extends to the time window where a composition effect would be expected, whether the second word allows composition or not. This suggests that the activity we detected in the first experiment may not reflect basic composition, as no interaction effect is obtained when controlling for expectancy. Moreover, a TFCE permutation analysis employing the same models used to evaluate the first experiment showed no significant interaction between task and number of words. Interestingly, a trial number effect was identified after second word onset, which could reflect an anticipatory response to the image to which subjects were required to answer. In this line, negative voltage was correlated with lower reaction times. Although the maximum trial number effect takes place after the second word, following the same behavior than for the composition task a comparable trial number effect can be noticed 200 ms after first word onset (see Fig. 6B).

It is important to point out that contrary to experiments in which a typical CNV is observed, Bemis & Pylkkänen's paradigm consists of two linguistic stimuli followed by an image, and therefore two expectancy processes would be elicited: one between the first and second word, and another between second word onset and image presentation. For all tasks and conditions, after second word presentation an anticipatory activity was probably elicited as subjects had to maintain the verbal material available in memory and prepare to give a motor response. We argue that this process would be equal for all conditions. However, a critical difference across tasks is the contingency between stimuli in the composition task, manifested as a task progression effect. In order to determine if our results are compatible with a CNV interpretation it would be relevant to implement the same task with a larger number of trials. If a learning effect accounts for our results, the increase in amplitude of the CNV should reach a plateau once the contingency is fully established. Another interesting alternative would be to increase the inter stimulus intervals (ISIs) between first word, second word and image presentation. This would allow the negative potentials to develop in time, enabling a comparison with the two-component response described for the CNV (Weerts & Lang, 1973; Loveless & Sanford, 1974). In our experiment, the short time that separates the three stimuli probably produces a superposition of anticipatory waves, preventing a proper characterization of these components.

Is this anticipatory response in part specific to language processing? According to pre dictive coding models (Farmer et al., 2013; Clark, 2013; Kutas & Federmeier, 2011) the expectancy-related activity could be part of language processing; in particular, it could be elicited by a composition-related process. It is clear though that the activity we observed in our first experiment (only in the composition task) does not correlate with linguistic com position since it appears to be present for both tasks in experiment 2. In this case both conditions with nouns as first word create a similar expectancy as the continuation is not known. Crucially, after the second word the situation is disambiguated, but we do not mea sure any difference in activity between conditions NA and NN that could be attributable to composition.

It has been shown that a negative potential is elicited when a delay is introduced before sentence-final words or when a specific linguistic stimulus can be predicted (Besson et al., 1997; León-Cabrera et al., 2017; Kaan & Carlisle, 2014). Evidence against a simple syntactic interpretation of this anticipatory response (Neufeld et al., 2016) comes from Bentin (1987). In this work, lexical stimuli separated by long ISIs were used to study neural responses to semantic expectancy. On some trials the first stimulus was a word, and subjects were asked to respond whether the next word was an antonym of the first. In contrast, on trials starting with a non-word, subjects had to perform a lexical decision task on the second stimulus. The elicited complex response (characterized by a sustained negative activity) had a larger amplitude in the antonym task that required the semantic content of the first word to be held in memory. This supports the possibility that the pre-combinatorial activity addressed in Neufeld et al.'s and in our experiments may reflect an anticipatory mechanism related to linguistic processing that is not specific to syntax or semantic composition.

A point that could be raised is whether the similar electrophysiological response for the *noun adjective* and *noun noun* conditions in our second experiment reflects subjects' at tempt to compose both nouns. Although noun-noun compounds are widely used in Spanish, compounding is not overly productive. There are hierarchical compounds, where the head noun can be in first or last position (*hombre araña* "spiderman", *hidroterapia* "hydrotherapy") and concatenative compounds which according to the final representation can be considered coextensional (*cantante-bailarín* "singer-dancer"), additive (*espacio-tiempo* "space-time") or intersective (*centro-derecha* "center-right") (Moyna, 2011). Nevertheless, none of the possi ble noun pairs used as stimuli correspond to established lexical compounds and we verified that they are not used in natural Spanish constructions. Although we cannot be certain that composition was not attempted by subjects, it would be a bad strategy to perform the task. Participants would benefit from maintaining the verbal material as independent units in order to check each word against the images presented. Moreover, open noun-noun composition would lead to ambiguous results (is a *shoe fork* a type of fork or a type of shoe?). Even though reaction times are higher and accuracy is lower for the noun-noun condition, participants still show a very good performance. This is not surprising, as during reading or hearing a sentence there is no certainty about which word will follow a specific stimuli and people can still correctly compose meanings. Therefore, the uncertainty introduced by our task is in this sense ecological.

It could be argued that our results are related to the linguistic differences between Spanish and English, specifically in relation to the order of adjectival modification, such that presentation of the head-noun in the first position elicits a linguistic process that would be absent when presenting an adjective. Nevertheless, there is evidence from Arabic that the LATL shows an increased response to noun phrases resulting from postnominal modification (Westerlund et al., 2015) which is no different from that observed using English stimuli. Moreover, this activity is elicited for English color-object noun phrases presented in reverse order (object-color) when the task demanded subjects to match both words to a single image representation (Bemis & Pylkkänen, 2013c). Notice that our first experiment reproduces all these features. Besides, the initial suggestion of a pre-composition activity for nouns preceded by adjectives was described for EEG using English noun-phrases in their canonical order, so there is at least one published report in English showing this pre-composition expectancy. It might also be questioned whether the difference we observe might be due to the lexical properties of the stimuli used. The original experiment consisted of monosyllabic stimuli whereas we use multisyllabic words. Nevertheless they are moderately frequent and easy to read (well within the window where reading time is insensitive to word length (Rayner, 1998)). In addition, following the English stimuli used in the previous studies the vast majority of words in our study are mono-morphemic, with the word "telephone" as the only exception, and hence we do not think this difference could explain the results.

Irrespective of what turns out to be the explanation of our results, it is doubtful that the activity we found in our first experiment is entirely a composition-based activity. Certainly, our work shows that some results previously published could be confounded with expectancy based processes. This has been addressed clearly for Neufeld et al. (2016); however this issue is also present in other reports using visual (Bemis & Pylkkänen, 2011, 2013c,a) and auditory (Bemis & Pylkkänen, 2013a) stimuli. In these cases, the authors do not analyze or plot the activity elicited before second word onset, and they do not discuss the potential issues that their experimental design entails. Furthermore, in studies in which expectancy would equally affect noun-adjective constructions (Westerlund & Pylkkänen, 2014; Westerlund et al., 2015; Ziegler & Pylkkänen, 2016; Zhang & Pylkkänen, 2015), conditions are only compared to each other or contrasted with a non word-word condition, disregarding the crucial list control condition. Importantly, these possible objections cannot be applied to all the studies produced on this subject. In particular, the compositional interpretation is supported by production studies. These authors show an increase in LATL's activity when subjects had to describe pictures with adjective-noun constructions in comparison to pictures described by enlisting two concepts (Pylkkänen et al., 2014; Prato & Pylkkänen, 2014; Blanco-Elorrieta et al., 2018). Although experiments were designed such that conditions were arranged in separate blocks there were no expectancy differences across stimuli. Thus our results do not question the involvement of LATL in linguistic composition, but show that EEG measures of brain responses elicited by this paradigm do not provide reliable evidence for an elementary composition marker, and suggest that some experimental designs used in MEG may be subject to the same limitation. On this line, we cannot rule out that the expectancy-related activity showed in this work is more apparent in EEG. This technique could be more sensitive to capture this anticipatory neural activity than MEG thus hindering the detection of a signal specific to composition. Additionally, it is possible that the hypothesis driven analyses carried out on specific regions of interest in the MEG studies manage to isolate composition from other processes.

In summary, we successfully adapted Bemis & Pylkkänen (2011) minimal composition paradigm to EEG and a language like Spanish that uses post-nominal adjectivation. We found an increased negativity for nouns followed by adjectives in a time window consistent with the composition-related activity described in the MEG literature. We have also shown the relevance of applying data-driven analyses that take into consideration task development effects, and adapted appropriate methods to do so. Finally, we introduced a non-blocked variant of the experiment to separate the contributions of composition and general expectancy effects to the measured signals. We suggest that the composition-related activity measured with EEG may be at least in part carried by expectancy-related processes arising from the block design of the experiment. Whether it is possible to find an unequivocal electrophysiological marker of elementary composition is a question that remains open and should be addressed in future work.

References

Bakdash, J. Z., & Marusich, L. R. (2018). *rmcorr: Repeated Measures Correlation*. URL: https://cran.r project.org/package=rmcorr.

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage*, *55*, 1847–1852. https://doi.org/10.1016/j.neuroimage.2011.01.066.

Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience*, *1*, 44–51. https://doi.org/10.1080/17588920903548751.

Barry, R. J., Steiner, G. Z., De Blasio, F. M., Fogarty, J. S., Karamacoska, D., & MacDonald, B. (2019). Components in the P300: Don't forget the Novelty P3! *Psychophysiology*, (pp. 1–15). https://doi.org/10.1111/psyp.13371.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 251–264. http://www.jstatsoft.org/v67/i01/. https://doi.org/10.18637/jss.v067.i01.

Bemis, D. K., & Pylkkänen, L. (2011). Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases. *Journal of Neuroscience*, *31*, 2801–2814. https://doi.org/10.1523/JNEUROSCI.5003-10.2011

Bemis, D. K., & Pylkkänen, L. (2013a). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, *23*, 1859–1873. https://doi:10.1093/ cercor/bhs170.

Bemis, D. K., & Pylkkänen, L. (2013b). Combination across domains: An MEG investigation into the relationship between mathematical, pictorial, and linguistic processing. *Frontiers in Psychology*, *3*, 1–20. https://doi.org/10.3389/

fpsyg.2012.00583

Bemis, D. K., & Pylkkänen, L. (2013c). Flexible Composition: MEG Evidence for the Deployment of Basic Combinatorial Linguistic Mechanisms in Response to Task Demands. *PLoS ONE*, *8*. https://doi.org/10.1371/journal.pone.0073949.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language*, *31*, 308–327. https://doi.org/10.1016/0093-934X(87)90077-0.

Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 130–149. https://doi.org/10.1037/0278-7393.20.1.130.

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*, 3–23. https://doi.org/10.1016/S0301-0511(96)05215-5.

Blanco-Elorrieta, E., Kastner, I., Emmorey, K., & Pylkkänen, L. (2018). Shared neural correlates for building phrases in signed and spoken language. *Scientific Reports*, *8*, 1–10. https://doi.org/10.1038/s41598-018-23915-0.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Composition of complex numbers: Delineating the computational role of the left anterior temporal lobe. *NeuroImage*, *124*, 194–203. https://doi.org/10.1016/j.neuroimage.2015.08.049.

Bosque, I., & Demonte, V. (1999). *Gramática Descriptiva de la Lengua Española*. Number v. 1 in Nebrija y Bello. Libreria Tirant lo Blanch. URL: https://books.google.fr/books?id=atcp{_}KuI4ysC.

Brunia, C., & Vingerhoets, A. (1981). Opposite hemisphere differences in movement related potentials preceding foot and finger flexions. *Biological Psychology*, *13*, 261–269. URL: http://linkinghub.elsevier.com/retrieve/pii/0301051181900417. https://doi.org/10.1016/0301-0511(81)90041-7.

Brush, C. J., Ehmann, P. J., Hajcak, G., Selby, E. A., & Alderman, B. L. (2018). Using Multilevel Modeling to Examine Blunted Neural Responses to Reward in Major Depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*, 1032–1039. URL: https://doi.org/10.1016/j.bpsc.2018.04.003. https://doi.org/10.1016/j.bpsc.2018.04.003.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. https://doi.org/10.1017/S0140525X12000477.

Cohen, J. (1969). Very slow brain potentials relating to expectancy: The CNV. In *Average evoked potentials: Methods, results, and evaluations*. (pp. 143–198). Washington: US National Aeronautics and Space Administration. URL: https://ntrs.nasa.gov/search.jsp?R=19700007575 http://content.apa.org/books/13016-004. https://doi.org/10.1037/13016-004.

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, *25*, 2584–2593. https://doi.org/10.1093/cercor/bhu057.

Donald, M. W. (1980). Memory, Learning and Event-Related Potentials. In H. H. Kornhubek, & L. Deecke *January 4, 2023* (Eds.), *Progress in Brain Research* (pp. 615–627). Elsevier volume 54 of *Progress in Brain Research*. URL: http://linkinghub.elsevier.com/retrieve/pii/S0079612308616817. https://linkinghub.elsevier.com/retrieve/pii/S0079612308616817. http://www.sciencedirect.com/science/article/pii/S0079612308616817. https://doi.org/10.1016/S0079-6123(08)61681-7.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*, 1246–1258. https://do.org/10.3758/s13428-013- 0326-1.

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, *36*, 211–212. URL: https:// www.cambridge.org/core/product/identifier/S0140525X12002312/type/journal{_}article. https://doi.org/10.1017/ S0140525X12002312.

Friederici, A. D. (2017). *Language in our brain: the origins of a uniquely human capacity*. URL: https: //mitpress.mit.edu/books/language-our-brain.

Friederici, A. D., Meyer, M., & Cramon, D. Y. V. (2000). Auditory Language Comprehension: An Event Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language*, *300*, 289–300. https://doi.org/10.1006/brln.2000.2313.

Gaillard, A. (1976). Effects of warning-signal modality on the contingent negative variation (CNV). *Biological Psychology*, *4*, 139–153. URL: http://linkinghub.elsevier.com/retrieve/pii/0301051176900132. https://doi.org/10.1016/0301-0511(76)90013-2.

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, *366*, 55–58. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aax0289. https://doi.org/10.1126/science.aax0289.

Hagoort, P., Baggio, G., & WIllems, R. M. (2009). Semantic unification. In *The cognitive neurosciences, 4th ed.* (pp. 819–835). Cambridge, MA, US: Massachusetts Institute of Technology.

Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, *16*, 638–650. https://doi.org/10.1006/nimg.2002.1102.

Hillyard, S. A. (1969). Relationships between the contingent negative variation (CNV) and reaction time. *Physiology and Behavior*, *4*, 351–357. https://doi.org/10.1016/0031-9384(69)90188-7.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of Anterior Temporal Cortex to Syntactic and Prosodic Manipulations During Sentence Processing. *Human Brain Mapping*, *138*, 128–138. https://doi.org/10.1002/hbm.20148.

Irwin, D. A., Knott, J. R., McAdam, D. W., & Rebert, C. S. (1966). Motivational determinants of the "contingent negative variation". *Electroencephalography and Clinical Neurophysiology*, *21*, 538–543. https://doi.org/10.1016/0013-4694(66)90172-6.

Jacobson, G. P., & Gans, D. P. (1981). The Contingent Negative Variation as an Indicator of Speech Discrimination Difficulty. *Journal of Speech Language and Hearing Research*, *24*, 345. URL: http://jslhr. pubs.asha.org/article.aspx?doi=10.1044/jshr.2403.345. https://doi.org/10.1044/jshr.2403.345.

Jasper, H. H. (1958). *The Tenn Twenty Electrode System of The International Federation*. https://doi.org/10.1016/0013-4694(58)90053-1.

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*, 523–532. https://doi.org/10.1017/S0048577200980624.

Kaan, E., & Carlisle, E. (2014). ERP indices of stimulus prediction in letter sequences. *Brain Sciences*, *4*, 509–531. https://doi.org/10.3390/brainsci4040509.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633. https://doi.org/10.3758/BRM.42.3.627.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*, 621–647. URL: http://www.annualreviews.org/doi/10.1146/annurev.psych.093008.131123. https://doi.org.10.1146/annurevpsych.093008.131123.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.7350657. https://doi.org/10.1126/science.7350657.

León-Cabrera, P., Rodríguez-Fornells, A., & Morís, J. (2017). Electrophysiological correlates of semantic anticipation during speech comprehension. *Neuropsychologia*, *99*, 326–334. URL: http://dx.doi.org/10.1016/j.neuropsychologia.2017.02.026. https://doi.org/10.1016/j.neuropsychologia.2017.02.026.

Loveless, N. E. (1975). The effect of warning interval on signal detection and event-related slow potentials of the brain. *Perception & Psychophysics*, *17*, 565–570. https://doi.org/

10.3758/BF03203970.

Loveless, N. E., & Sanford, A. J. (1974). Slow potential correlates of preparatory set. *Biological Psychology*, *1*, 303–314. https://doi.org/10.1016/0301-0511(74)90005-2.

Low, M. D., Coats, A. C., Rettig, G. M., & McSherry, J. W. (1967). Anxiety, attentiveness-alertness: A phenomenological study of the CNV. *Neuropsychologia*, *5*, 379– 384. URL: http://www.rocksbackpages.com/Library/Article/eels-the-freak-shall-inherit-the-earth http: //linkinghub.elsevier.com/retrieve/pii/0028393267900097. https://doi.org/10.1016/0028-3932(67)90009-7.

Luck, S. J. S. J. (2005). *An introduction to the event-related potential technique*. MITPress

Malmivuo, J., & Plonsey, R. (1995). *BioelectromagnetismPrinciples and Applications of Bioelectric and Biomagnetic Fields* volume 15. Oxford University Press. URL: http://www. oxfordscholarship.com/view/10.1093/acprof:oso/9780195058239.001.0001/acprof-9780195058239. https://doi.org/10.1093/acprof:oso/9780195058239.001.0001.

Martin, A. E., & Baggio, G. (2020). Modelling meaning composition from formalism to mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190298. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0298. https://doi.org/10.1098/rstb.2019.0298.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., & Salamon, G. (1993). The Cortical Representation of Speech. *Journal of Cognitive Neuroscience*, *5*, 467–479. https://doi.org/10.1162/jocn.1993.5.4.467.

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *NeuroImage*, *67*, 111–118. URL: http://dx.doi.org/10.1016/j.neuroimage.2012.10.027. https://doi.org/10.1016/j.neuroimage.2012.10.027.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*, 2195–2222. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245704002135. https://doi.org/10.1016/j.clinph.2004.06.001.

Michel, J. B., Kui Shen, Y., Presser Aiden, A., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182. https://doi.org/10.1126/science.1199644.

Molinaro, N., Carreiras, M., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, *59*, 3488–3501. URL: http://dx.doi.org/10.1016/j.neuroimage.2011.11.009. doi:10.1016/j.neuroimage.2011.11.009.

Moyna, M. I. (2011). *Compound Words in Spanish: Theory and History*.

Neufeld, C., Kramer, S. E., Lapinskaya, N., Heffner, C. C., Malko, A., & Lau, E. F. (2016). The Electrophysiology of Basic Phrase Building. *PLOS ONE*, *11*, e0158446. URL: https://dx.plos.org/10.1371/journal.pone.0158446. https://doi.org/10.1371/journal.pone.0158446.

Okada, Y., Lähteenmäki, A., & Xu, C. (1999). Comparison of MEG and EEG on the basis of somatic evoked responses elicited by stimulation of the snout in the juvenile swine. *Clinical Neurophysiology*, *110*, 214–229. https://doi.org/10.1016/S0013-4694(98)00111-4.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869.arXiv:156869.

Parks, N. A., Gannon, M. A., Long, S. M., & Young, M. E. (2016). Bootstrap Signal-to-Noise Confidence Intervals: An Objective Method for Subject Exclusion and Quality Control in ERP Studies. *Frontiers in Human Neuroscience*, *10*, 1–15. URL: http://journal.frontiersin.org/Article/10.3389/fnhum.2016.00050/abstract. https://doi.org/10.3389/fnhum.2016.00050.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987. https://doi.org/10.1038/nrn2277.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. URL: http://dx.doi.org/10.1016/j.jneumeth.2006.11.017. doi: 10.1016/j.jneumeth.2006.11.017. Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a New Neurobiology of Language. *Journal of Neuroscience*, *32*, 14125–14131. URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3244-12.2012. https://doi.org/10.1523/JNEUROSCI.3244-12.2012.

Polich, J. (1989). Habituation of P300 from auditory stimuli. *Psychobiology*, *17*, 19–28. https://doi.org/10.3758/BF03337813.

Poon, L. W., Thompson, L. W., Williams, R. B., & Marsh, G. R. (1974). Changes of Antero-Posterior Distribution of CNV and Late Positive Component as a Function of Information Processing Demands. *Psychophysiology*, *11*, 660–673. URL: http://doi.wiley.com/10.1111/j.1469-8986.1974.tb01135.x. https://doi.org/10.1111/j.1469-8986.1974.tb01135.x.

Poortman, E. B., & Pylkkänen, L. (2016). Adjective conjunction as a window into the LATL's contribution to conceptual combination. *Brain and Language*, https://doi.org/10.1016/j.bandl.2016.07.006.

Prato, P. D., & Pylkkänen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00524.

Proulx, G. B., & Picton, T. W. (1980). The CNV during Cognitive Learning and Extinction. *Progress in Brain Research*, *54*, 309–313. https://doi.org/10.1016/S0079-6123(08)61640-4.

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*, 62–66. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aax0050. https://doi.org/10.1126/science.aax0050.

Pylkkänen, L. (2020). Neural basis of basic composition: What we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*. https://doi.org/10.1098/rstb.2019.0299.

Pylkkänen, L., Bemis, D. K., & Blanco, E. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition*, *133*, 371–384. https://doi.org/10.1016/j.cognition.2014.07.001.

Pylkkänen, L., Brennan, J. R., & Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, *26*, 1317–1337. URL: http://www.tandfonline.com/doi/abs/10.1080/01690965.2010.527490. https://doi.org/10.1080/01690965.2010.527490.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.r-project.org/.

Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42–55. URL: http://dx.doi.org/10.1038/nrn.2016.150. https://doi.org/10.1038/nrn.2016.150.

Ravden, D., & Polich, J. (1998). Habituation of P300 from visual stimuli. *International Journal of Psychophysiology*, *30*, 359–365. https://doi.org/10.1016/S0167-8760(98)00039-7.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psycho logical Bulletin*, *124*, 372–422. https://doi.org/10.1037/0033-2909.124.3.372.

Rebert, C. S., McAdam, D. W., & Knott, J. R. (1967). Slow potential change in human brain related to level of motivation. https://doi.org/10.1037/h0024146.

Rohrbaugh, J. W., & Gaillard, A. W. (1983). 13 Sensory and Motor Aspects of the Contin gent Negative Variation. In *Tutorials in ERP Research: Endogenous Components* (pp. 269– 310). URL: http://linkinghub.elsevier.com/retrieve/pii/S0166411508620440. https://linkinghub.elsevier.com/retrieve/pii/S0166411508620440. https://doi.org/10.1016/S0166-4115(08)62044-0.

Segaert, K., Mazaheri, A., & Hagoort, P. (2018). Binding language: Structuring sentences through precisely timed oscillatory mechanisms. *European Journal of Neuroscience*, *1*, 1–12. https://doi.org/10.1111/ejn.13816.

Simons, R. F., ÖHman, A., & Lang, P. J. (1979). Anticipation and Response Set: Cortical, Cardiac, and Electrodermal Correlates. *Psychophysiology*, *16*, 222–233. https://doi.org/10.1111/j.1469-8986.1979.tb02982.x.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*, 83–98. URL: http://dx.doi.org/10.1016/j.neuroimage.2008.03.061. https://doi.org/10.1016/j.neuroimage.2008.03.061.

Snijders, T. M., Vosse, T., Kempen, G., Berkum, J. A. V., Petersson, K. M., & Hagoort, P. (2009). Retrieval and Unification of Syntactic Structure in Sentence Comprehension: an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex*, *19*, 1493–1503. https://doi.org/10.1093/cercor/bhn187.

Ströberg, K., Andersen, L. M., & Wiens, S. (2017). Electrocortical N400 effects of semantic satiation. *Frontiers in Psychology*, *8*, 1–14. https://doi.org/10.3389/fpsyg.2017.02117.

Tecce, J. J., Savignano-Bowman, J., & Meinbresse, D. (1976). Contingent negative variation and the distraction-arousal hypothesis. *Electroencephalography and Clinical Neurophysiology*, *41*, 277–286. https://doi.org/10.1016/0013-4694(76)90120-6.

Tecce, J. J., & Scheff, N. M. (1969). Attention Reduction and Suppressed Direct-Current Potentials in the Human Brain. *Science*, *164*, 331–333. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.164.3877.331. https://doi.org/10.1126/science.164.3877.331.

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, *55*, 1–12. https://doi.org/10.1111/psyp.13044.

Vossen, H., Van Breukelen, G., Hermens, H., Van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: a mixed regression approach. *International Journal of Methods in Psychiatric Research*, *16*, n/a–n/a. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-79951999327{&}partnerID=tZOtx3y1. http://doi.wiley.com/10.1002/mpr.348. https://doi.org/10.1002/mpr.348. arXiv:1106.4512.

Walter, W. G., Cooper, R., Aldrige, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent Negative Variation: An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain. *Nature*, *203*, 380. URL: https://doi.org/10.1038/203380a0http://10.0.4.14/203380a0.

Weerts, T. C., & Lang, P. J. (1973). The effects of eye fixation and stimulus and response location on the Contingent Negative Variation (CNV). *Biological Psychology*, *1*, 1–19. https://doi.org/10.1016/0301-0511(73) 90010-0.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language*, *141*, 124–134. URL: http://dx.doi.org/10.1016/j.bandl.2014.12.003. https://linkinghub.elsevier.com/retrieve/pii/S0093934X1400190. https://doi.org/10.1016/ j.bandl.2014.12.003.

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59–70. URL: http://dx.doi.org/10.1016/j.neuropsychologia.2014.03.001. https://doi.org/10.1016/j.neuropsychologia.2014.03.001.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*, 1–12. https://doi.org/10.3389/fpsyg.2016.01832.

Williamson, S., & Kaufman, L. (1981). Biomagnetism. *Journal of Magnetism and Magnetic Materials*, *22*, 129–201. URL: https://linkinghub.elsevier.com/retrieve/pii/0304885381900780. https://doi.org/10.1016/0304-8853(81)90078-0.

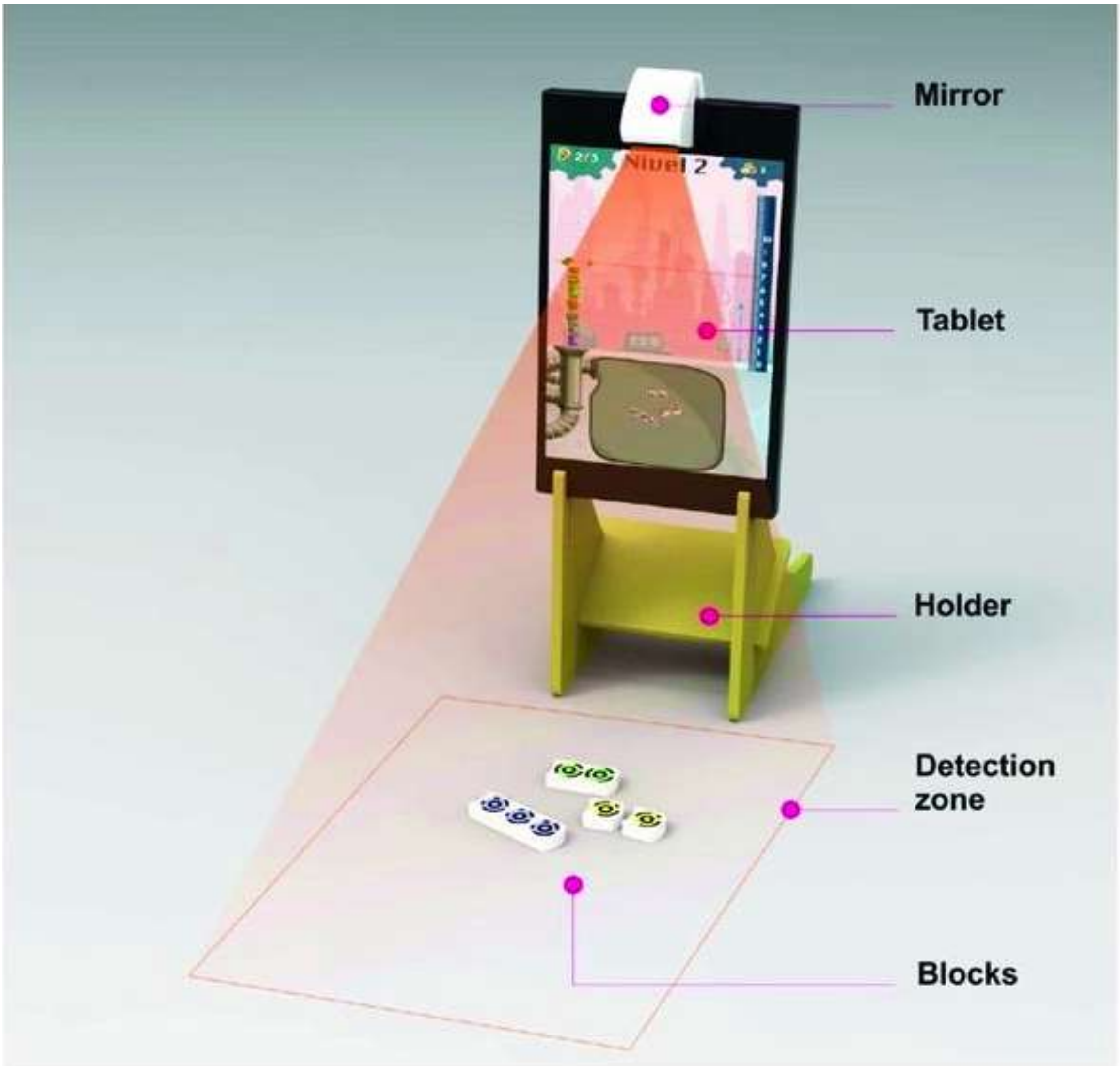
Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology*, *6*, 1–9. https://doi.org/10.3389/fpsyg.2015.01818.

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2017). Building by Syntax: The Neural Basis of Minimal Linguistic Structures. *Cerebral cortex (New York, N.Y.: 1991)*, *27*, 411–421. https://doi.org/10.1093/ cercor/bhv234.

Zhang, L., & Pylkkänen, L. (2015). NeuroImage The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, *111* , 228–240. URL: http://dx.doi.org/10.1016/j.neuroimage.2015.02.028. https://doi.org/10.1016/j.neuroimage.2015.02.028.

Ziegler, J., & Pylkkänen, L. (2016). Neuropsychologia Scalar adjectives and the temporal unfolding of semantic composition: An MEG investigation. *Neuropsychologia*, *89*, 161–171. URL: http://dx.doi.org/ 10.1016/j.neuropsychologia.2016.06.010. https://doi.org/10.1016/j.neuropsychologia.2016.06.010.

Ceibal Tangible (CETA)



CETA (Ceibal Tangible) es un dispositivo que adapta tablets distribuidas por Plan Ceibal a la modalidad de Interacción tangible. Esto es, permite que puedan llevarse adelante actividades realizadas con objetos que son detectados por la computadora a través de un sistema de visión por computadora. CETA ha sido desarrollado a través de financiaciones de ANII y Fundación Ceibal y se han desarrollado videojuegos para la estimulación de las matemáticas, así como, audiojuegos para niños ciegos y con baja visión. El inicio de CETA nace en el **CIBPsi** a partir del interés de algunas investigadoras que habían trabajado en dispositivos de interacción tangible para niños con parálisis cerebral en colaboración con docentes y estudiantes de Facultad de Ingeniería (Proyecto Periscopio). A partir de la llegada de aplicaciones comerciales que explotaban esta modalidad de interacción con fines educativos un grupo de investigadores del centro se puso a pensar en cómo generar una solución de bajo coste que pudiera ser creada para las tablets que Ceibal comenzaba a distribuir. La propuesta fue presentada a varias convocatorias de I + D y finalmente fue financiada en 2017. A partir de allí comenzó un largo viaje de investigación y desarrollo que prosigue hasta nuestros días.

CETA (Ceibal Tangible) is a device that adapts tablets distributed by Plan Ceibal to the tangible interaction modality. That is, it allows activities to be carried out with objects that are detected by the computer through a computer vision system. CETA has been developed through funding from ANII and Ceibal Foundation and has developed video games for mathematics stimulation and also audio games for blind and low vision children. The beginning of CETA was born at **CIBPsi** from the interest of some researchers who had worked on tangible interaction devices for children with cerebral palsy in collaboration with teachers and students from the School of Engineering (Periscopio Project). After the arrival of commercial applications that exploited this interaction modality for educational purposes, a group of researchers from the center began to think about how to generate a low-cost solution that could be created for the tablets that Ceibal was beginning to distribute. The proposal was submitted to several I + D calls and was finally funded in 2017. From there began a long journey of research and development that continues to this day.

Blues in Two Different Spanish-speaking Populations

Fernando González-Perilli[1,2]*, Ignacio Rebollo[1], Alejandro Maiche[1] and Analía Arévalo[3]

Fernando González Perilli https://orcid.org/0000-0001-5832-716X, Ignacio Rebollo https://orcid.org/0000-0002-4119-995, Alejandro Maiche https://orcid.org/0000-0002-5006-1544, Analía Arévalo https://orcid.org/0000-0002-3793-8081

1 **Centro de Investigación Básica en Psicología (CIBPsi)**. Universidad de la República. Montevideo, Uruguay
2 Facultad de Información y Comunicación. Universidad de la República. Montevideo, Uruguay
3 Departamento de Neurologia, Faculdade de Medicina. Universidade de São Paulo. São Paulo, Brazil

Correspondence concerning this article should be addressed to Fernando González Perillil. Facultad de Información y Comunicación. Universidad de la República. San Salvador 1944, 11200. Montevideo, Uruguay. Contact: **fgonzalezperilli@gmail.com**

This is a free version -with the the consent of the authors- of the open access manuscript González-Perilli, F., Rebollo, I., Maiche, A., & Arévalo, A. (2017). Blues in two different Spanish-speaking populations. *Frontiers in Communication*, *2*, 18. Available at **https://doi.org/10.3389/fcomm.2017.00018**

Abstract

Several studies investigating color discrimination across languages have shown a facilitation effect in groups that employ more than one term to refer to a given color. While Uruguayans use "*azul*" to refer to dark blue and "*celeste*" for light blue, Spaniards use "*azul*" for dark blue and the compound terms "*azul celeste*" or "*azul claro*" for light blue. In this study, Uruguayan and Spanish participants discriminated between pairs of color stimuli that lie at different distances from each other on the blue color spectrum in three different sessions: a session with no interference (basic task), one with verbal and one with visual interference. Only the Uruguayans were more accurate at distinguishing between stimuli associated with different color terms. Furthermore, while both Uruguayans and Spaniards showed a category effect in response times, the effect was strongest for Uruguayans when items were closer to each other on the color spectrum (i.e., more difficult). This study is unique in that we observed different Whorfian effects in two groups that speak the same language but differ in their use of color-specific terms. Our results contribute to the discussion of whether and to what extent language or other cultural variables affect the perception of different color categories.

Keywords: color perception, categorical perception, linguistic relativity, Sapir–Whorf hypothesis, cross-cultural cognition

Introduction

To what extent do language and/or culture affect the way we process and organize the information and experiences that make up our world? The work of Sapir, Whorf, and others sparked this famous debate at least a century ago, and these questions continue to interest academics across fields to this day (Whorf, 1956; Lucy and Shweder, 1979; Kay and Kempton, 1984; Vygotsky, 1987; Lupyan, 2012; Levelt, 2014).

Most investigations addressing this topic have been characterized as either descriptive, simply reporting interesting differences between two or more languages, or aiming to explain how observed disparities are associated with different cognitive processes (Zlatev and Blomberg, 2015). These two perspectives are also associated with weak and strong versions of the Sapir–Whorf hypothesis (lan guage and thought are interrelated vs. language determines thought, Brown, 1976). Both hypotheses have been criticized for being trivial and non-informative (weak version) or theoretically and/or methodologically wrong (strong version) (Bloom and Keil, 2001).

Zlatev and Blomberg (2015) propose approaching each investigation according to whether the focus is on the structure of language or on its implementation (discourse). Traditional cognitive approaches focus on abstract structural aspects of language and search for innate universal fea tures. On the other hand, linguistic relativism concentrates on how the phenomenon of categorical perception (CP, Harnad, 2005) is affected by different contextual factors, such as language and culture.

According to Lucy (1997), there are three "logical components" that are typically taken into account when studying linguistic relativity: (1) the distinction between language and thought, (2) the mechanisms explaining the instantiation of a possible influence, and (3) the identification of other factors involved in the phenomenon.

Regarding the first point, relativists often agree with a broad definition of *thought*, not just as a conscious reflective process (as understood in folk psychology) but also involving less aware, automatic processes, such as perception and categorization. Moreover, language and perception are not understood as isolated modules—as in classic cognitivism (Pylyshyn, 1999)—but are thought to interact with a myriad of processes. Thus, the role of verbal labels affecting perception and categorization is a key issue in contemporary approaches (Thierry, 2016). How basic cognitive processes are influenced by implicit recovery of linguistic (but also contextual and sociocultural information) is another key question, which involves points 2 and 3.

Therefore, the key notion leading the research on linguistic relativity is not whether minds are dependent on a given language but how verbal labels and categories interact with cognition across different contexts (Thierry, 2016; Zhong et al., 2017). Topics currently being studied include: cross-cultural comparisons (i.e., Boroditsky, 2001; Casasanto, 2008), the exploration of categorical effects under different interference conditions (i.e., Roberson and Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007), and the time course of the effect, which informs whether perception or higher cognitive processes are involved (Mo et al., 2011; Clifford et al., 2012; He et al., 2014; Forder et al., 2017).

One line of research within this debate concerns the way in which different languages divide color space. The key question within this work is whether these varying linguistic representations affect performance on tasks that are seemingly non-linguistic. In other words, does the way in which a particular language categorizes colors affect the way its speakers think about and organize color in their minds, even in the absence of an explicitly linguistic task? One special case—that of the color *blue*—has been studied by researchers across a number of languages, including Greek (Androulaki et al., 2006; Athanasopoulos, 2009; Thierry et al., 2009), Italian (Bimler and Uusküla, 2014), Japanese (Athanasopoulos et al., 2010), Korean (Roberson et al., 2009), and Russian (Witthoft et al., 2003; Winawer et al., 2007). These languages share a common feature that distinguishes them from English: they divide the color blue into two distinct linguistic categories, one depicting lighter blues, and the other depicting darker blues. In the above studies, speakers of those languages were relatively better than English speakers at distinguishing between color samples along the blue color spectrum when the samples' names came from different linguistic categories, even though the task did not require linguistic output.

This kind of implicit linguistic effect is explained by theories arguing that linguistic labels can aid in the discrimination of stimuli that are hard to categorize (Lupyan, 2012) thanks to a pre dicting coding process in which "every level of the hierarchically organized system that constitutes the brain works to predict the activity in the level below" (Lupyan and Clark, 2015, p. 279). In such a predictive framework, the brain's function is to produce a percept that fits the best hypothesis regarding the state of the world that is being conceived (Lupyan and Clark, 2015). That is acquired through an interplay of top-down knowledge about the world and incoming bottom-up sensory information (Bar, 2003). In Lupyan's view, labels work as hubs of perceptual, semantic and contextual information related to specific categories. Their function is to reduce prediction error by enhancing the percep tion of typical categorical features. Therefore, verbal labels can be elicited to foster predictability and support cognition.

Aiming to clarify this issue, several studies include a verbal interference condition. That is, they introduce a concurrent task demanding linguistic resources (e.g., remembering a string of digits). This interference is expected to disrupt categorical effects (advantage for the discrimination of stimuli pertaining to different categories) if linguistic processes are necessary for CP to occur. For instance, Winawer et al. (2007) showed that when an additional task requiring verbal memory was included, the categorical effects found for the Russian participants vanished, suggesting linguistic resources are used by Russian speakers in this seemingly non-linguistic color perception task. The authors also presented a spatial interference condition that did not alter categorical effects, further supporting the view that the a disrup tion of the CP advantages was in fact due to a disruption in lin guistic processing and not to the heavier cognitive load imposed by any interference task.

In the current study, we compared two groups of speakers of the same language that employ different verbal labels for the same color. This comparison is interesting because, unlike previ ous studies where groups of speakers spoke different languages, differences between the current groups should be much subtler, and may reflect cultural variations that affect the frequency of use of such labels.

Similarly to the languages investigated in previous studies (Androulaki et al., 2006; Winawer et al., 2007), in some variants of Spanish, the color blue is associated with two different lingui stic terms: dark blues are *azul* and light blues are *celeste*. However, the Spanish language presents an interesting case, in that dif ferent populations of Spanish speakers differ in the way they implement this distinction. Namely, in some South American countries such as Uruguay, the term *celeste* (light blue) is used on its own. By contrast, in Spain, the term "*celeste*" is used as part of a compound word, i.e., *azul celeste*, making *celeste* a sub category within the larger category of *azul*, or (regular or dark) blue. The word (and color) *celeste* also carries significant cultural weight in Uruguay, given that it is found on national emblems and by extension, national sports team uniforms. A recent study conducted by our group confirmed the use of *celeste* as a sepa rate basic color term (BCT) for light blues in Uruguay. Thirty healthy participants were given 2 min to write down as many color names as they could remember while keeping their eyes closed (Elicited List task: Corbett and Davies, 1997). Following Berlin and Kay's (Berlin and Kay, 1969) work, one would predict that only 11 different color names would be elicited in more than 50% of the lists produced by participants. In this study, Figure 1 Illustration of the stimuli employed in the experimental tasks. Top: color chips ranging from light blue to dark blue. Bottom: middle: example of a triad used in the discrimination task; left: example of a cross-category comparison; right: example of a within-category comparison.

Uruguayan participants consistently produced 12 names, as they included *celeste* as its own color category. In fact, both *azul* and *celeste* were consistently found among the first BCTs reported by Uruguayans (Lillo et al., 2016). For the current experiment, we tested Uruguayan as well as Spanish participanton a color discrimination task we designed using stimuli along the *azul-celeste* boundary. Since cultural as well as linguistic differences have been used to explain Whorfian effects across different populations, the Uruguay-Spain com parison is interesting because the two populations come from different cultures but use the same language and very similar color space partitions. That is, when asked to assign segments of the color spectrum to different color terms, Uruguayans and Spaniards coincide perfectly on all terms except for *celeste*: the space Uruguayans call "*celeste*" falls into the greater category of "*azul*" for Spaniards (Lillo et al., 2016). Given the presence of the 12th BCT for the Uruguayans, we hypothesized that this group would display a relatively stronger categorical advantage than Spaniards.

Material and Methods

Participants A total of 73 individuals participated in this study: 35 were recruited from the Universitat Autònoma de Barcelona, Spain, and 38 were recruited from the Universidad de la República in Montevideo, Uruguay. All of them were native speakers of the Spanish spoken in their country, and 22 of the Spanish participants were also Catalan speakers. Nine participants (2 from Spain and 7 from Uruguay) who produced more than 25% errors and RTs < 200 and >3,000 ms were excluded from the analysis, for a final group of 33 Spaniards (mean age = 25.1, SD = 3; 18 female) and 31 Uruguayans (mean age = 22.5, SD = 3.2; 17 female). Groups did not differ significantly from each other in terms of gender or age [$F(1,62) = 0.802$, $p = 0.374$].

Stimuli We created 20 computer-simulated color chips that ranged from light blue (*azul celeste* in Spain and *celeste* in Uruguay) to dark blue (*azul oscuro* in Spain and *azul* in Uruguay) (Figure 1). Stimuli coordinates (Commission Internationale de l'Eclairage, *Yxy*) ranged from $Y = 29.26$, $x = 0.217$, $y = 0.274$ for stimulus 1 to $Y = 4.18$, $x = 0.182$, $y = 0.167$ for stimulus 20. Stimuli varied primarily in the luminance axis (Y) and the y chromaticity axis, and were selected taking into account previous

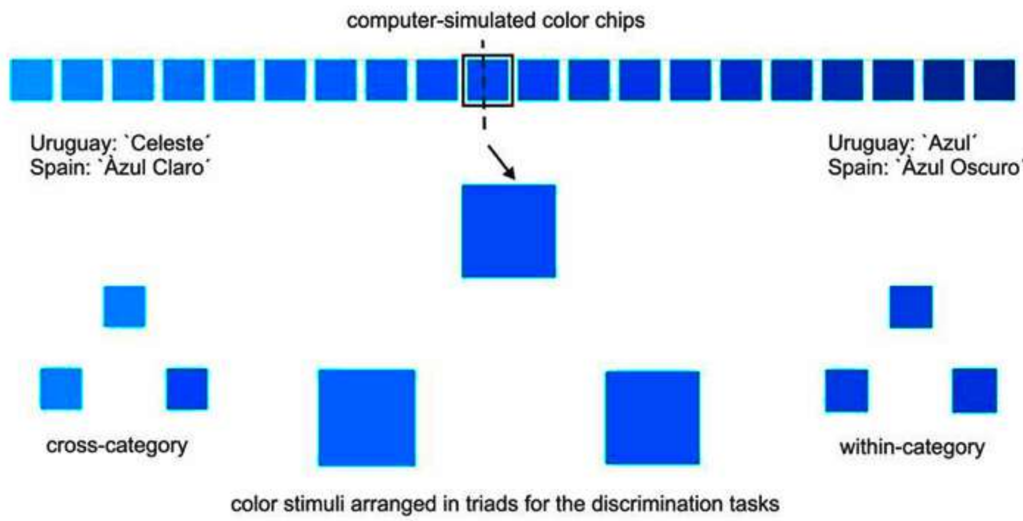


Figure 1. Illustration of the stimuli employed in the experimental tasks. Top: color chips ranging from light blue to dark blue. Bottom: middle: example of a triad used in the discrimination task; left: example of a cross-category comparison; right: example of a within-category comparison.

research on color categories in Spanish (Lillo et al., 2007) as well as cross-linguistic comparisons (Winawer et al., 2007; Roberson et al., 2009). The color squares measured 2.5 cm per side, and subjects viewed the screen from a distance of 60 cm. In addition, there were two categories of deviant stimuli: near and far. “Near” stimuli were colors that were two chips away from the target stimulus while “far” stimuli were four chips away (Figure 2). Discrimination between “near” stimuli was expected to be more difficult than between “far” stimuli.

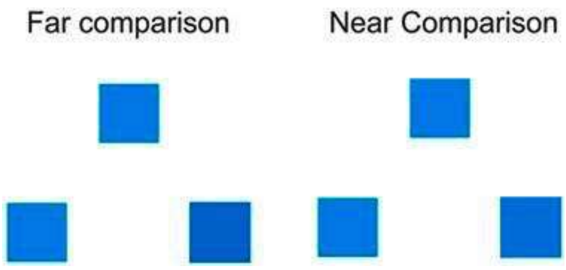


Figure 2. Illustration of the stimuli employed in the experimental tasks. Top: color chips ranging from light blue to dark blue. Bottom: middle: example of a triad used in the discrimination task; left: example of a cross-category comparison; right: example of a within-category comparison.

Procedure Prior to participation, an investigator explained the study to participants, who then signed an informed consent form. All study procedures were conducted with the approval of the Research Ethics Committee of the Department of Psychology at University of the Republic (Uruguay) and the Department of Basic Psychology at the Autonomous University of Barcelona (a separate ethics approval was not required as per the Autonomous University of Barcelona guidelines and as per Spanish regulations) and were in accordance with the Declaration of Helsinki. Participants viewed three color squares arranged in triads (1 above and 2 below) (Figure 1) and were asked to decide which of the two lower squares matched the one on top. The side (right or left) on which the distractor was presented was counterbalanced across trials. Each participant completed three blocks of 136 color discrimination trials: one regular block (Basic Task), one block that also included a secondary spatial interference task, and a third block that included a verbal interference task. Half of the comparisons included “near” stimuli and half included “far” stimuli. The two interference tasks (one verbal and one spatial) were included, following Winawer et al. (2007), to test whether either type of interference affected any observed categorical effects, thus shedding light on the type of processing employed by partici pants during the basic task.

Interference Tasks

(a) Spatial interference: participants viewed a 4 × 4 square grid in which four randomly chosen squares were shaded black (Figure 3) and were instructed to maintain a picture of it in mind until tested. A two-choice test was presented every eight color discrimination trials.

(b) Verbal interference: participants were shown an eight-digit number series (Figure 3) for 3 s every eight color discrimi nation trials and were asked to rehearse it while completing the color discrimination task. Their recall was then tested by having them choose between the original series and a foil that differed by one digit.

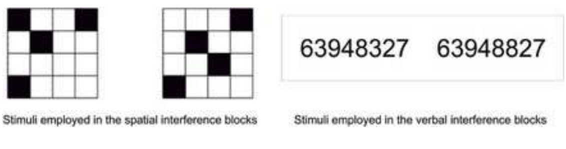


Figure 3. Example of stimuli employed in the spatial interference block (left), and in the verbal interference block (right).

Participants’ Boundaries Following the categorization tasks, participants also completed a *Border detection task* designed to test each individual’s color boundary between dark and light blues. Participants viewed the 20 stimuli (which appeared 10 times and in random order) and pressed a key to indicate whether each color was *celeste* or *azul* (for Uruguayans) and *azul celeste* or *azul oscuro* (for Spaniards). They were asked to make all judgments as quickly and accurately as possible.

Overall, 36% of participants identified Stimulus 10 as the cat egorical boundary, 24% chose Stimulus 9, 20% chose Stimulus 8, 14% chose Stimulus 11, and 6% chose Stimulus 7. All Uruguayans categorized Stimulus 1 as *celeste* (light blue) and stimulus 20 as *azul* (dark blue), while all Spanish participants categorized Stimulus 1 as *azul celeste* (sky blue) or *azul claro* (light blue) and Stimulus 20 as *azul oscuro* (dark blue). Each participant’s score was determined individually by using his/her color boundary to classify the color discrimination trials as either cross-category or within-category. This classification was made individually (i.e., not based on the group average).

Errors and Outliers In order that we only analyzed data from trials in which partici pants were actively following the interference tasks, we system atically discarded all eight color trials preceding each incorrectly answered interference trial (5.74% of trials).

We also eliminated all trials with reaction times below 200 or above 3,000 ms (2.41% of trials across participants). RT analyses were conducted only on accurate responses (87.5%).

Results

We conducted a mixed ANOVA with three within-subject factors (Distance × Interference × Category) and one between-subjects factor (country: Uruguay vs. Spain).

Accuracy Groups did not differ in terms of overall accuracy: Uruguay (M = 86.1, SD = 0.61) vs. Spain (M = 88.3, SD = 0.63), $F(1, 62) = 1.942$, $p = 0.168$, $\eta 2 = 0.030$.

There were two significant main effects: Distance, $F(1,62) = 303.109$, $p < 0.0001$, $\eta 2 = 0.830$, and Category, $F(1,62) = 5.845$, $p = 0.01$, $\eta 2 = 0.086$. When analyzed together, participants were more accurate at distinguishing between far trials (M = 0.94, SD = 0.04) than between near trials (M = 0.80, SD = 0.09), and between cross-category trials (M = 0.87, SD = 0.07) than between within-category trials (M = 0.86, SD = 0.06). There were also three significant interactions: Interference × Country, Distance × Country and, most interestingly, Category × Country.

Interference × Country, $F(1, 62) = 3.219$, $p = 0.043$, $\eta 2 = 0.049$. *Post hoc* analyses showed that the interference factor was not significant when analyzed separately for each group, and that the difference between groups was significant only in the verbal interference condition, $F(1,62) = 2.304$, $p = 0.025$, $d = 0.4$. Distance × Country, $F(1, 62) = 4.252$, $p = 0.043$, $\eta 2 = 0.064$. Uruguayans had relatively greater difficulty discriminating between near stimuli (near: M = 0.78, SD = 0.13; far: M = 0.94, SD = 0.06) than did Spaniards (near: M = 0.82, SD = 0.13; far: M = 0.95, SD = 0.06).

Post hoc analyses (separate one-way ANOVAs for each group) showed that distance effects were significant for both countries, Uruguay. $F(1, 30) = 157.375$, $p < 0.0001$, $\eta 2 = 0.840$., Spain: $F(1, 32) = 145.353$, $p < 0.0001$, $\eta 2 = 0.820$. Moreover, pairwise comparisons showed that neither near nor far cases showed differences between countries ($p > 0.05$).

Category × Country, $F(1,62) = 2.123$, $p = 0.19$, $\eta 2 = 0.086$. Uruguayans showed an advantage for cross-category trials com pared to within category trials (M = 0.87, SD = 0.07 vs. M = 0.85, SD = 0.07); *post hoc* analyses: $t(1,30) = 3.268$, $p = 0.003$, $d = 0.29$. Spaniards, on the other hand, did not show this advantage (within: M = 0.88, SD = 0.06, cross: M = 0.88, SD = 0.07), $p > 0.05$ (see Figure 4). All other effects and interactions were not significant (all $p > 0.05$).

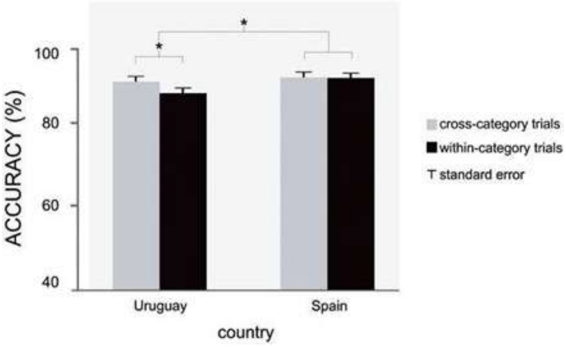


Figure 4. Mean differences in accuracy between cross- and within-category stimuli by country. Error bars represent SEM.

RT Overall, Uruguayans were significantly slower than Spaniards, $F(1,62) = 8.196$, $p = 0.006$, $\eta 2 = 0.117$ (M = 1043 ms, SD = 278 ms vs. M = 900 ms, SD = 287 ms). There were also significant main effects of Distance, $F(1,62) = 267.638$, $p < 0.0001$, $\eta 2 = 0.812$, and Category, $F(1,62) = 27.331$, $p < 0.0001$, $\eta 2 = 0.306$.

In line with the accuracy results, participants were faster at discriminating between far trials (M = 862 ms, SD = 175 ms) than near ones (M = 1,081 ms, SD = 235 ms), and on cross-category (M = 952 ms, SD = 206) compared to within-category trials (M = 991 ms, SD = 198 ms) (see Figure 5).

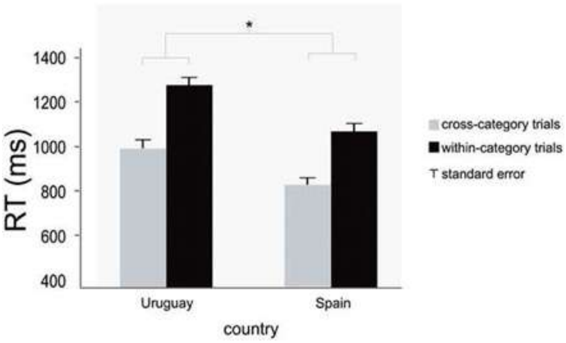


Figure 5. Mean response times (ms) between cross- and within-category stimuli by country. Error bars represent SEM.

The first-order interaction of Interference × Country was significant, $F(1,61) = 3.517$, $p = 0.033$, $\eta 2 = 0.054$. Sessions with spatial interference, in which Uruguayans performed best, resulted in the Spanish group's slowest responses (Spain: Basic: M = 889, SD = 336; Spatial: M = 941, SD = 328; Verbal: M = 869, SD = 343; Uruguay: Basic: M = 1087, SD = 347: Spatial: M = 994, SD = 342; Verbal: M = 1048, SD = 354).

Post hoc analyses showed that differences across sessions were not significant within countries, but results comparing Spain and Uruguay were different for two of the three interference conditions. Differences between groups were significant in the Basic (no interference) session, $t(1,62) = 3.271$, $p = 0.002$, $d = 0.58$, and in the Verbal interference session, $t(1,62) = 2.895$, $p = 0.005$, $d = 0.51$.

Distance × Category, $F(1,62) = 3.769$, $p = 0.019$, $\eta 2 = 0.085$. A category advantage (difference between cross- and within category trials) was stronger for far (Mdifference = 54 ms) than for near color comparisons (Mdifference = 25 ms).

Nevertheless, *post hoc* analyses reflected that both differences were significant: Far, $F(1,63) = 3.769$, $p = 0.003$, $\eta 2 = 0.129$; Near, $F(1,63) = 3.769$, $p = 0.000$, $\eta 2 = 0.257$. Additionally, categorical effects were significant at both distance conditions.

Cross-category: $F(1,62) = 27.811$, $p = 0.000$, $\eta 2 = 0.310$; within category: $F(1,62) = 62.927$, $p = 0.000$, $\eta 2 = 0.504$.

While the Category × Country interaction was not significant ($p = 0.090$), the three-way Country × Distance × Category inter action was, $F(1,62) = 6.596$, $p = 0.013$, $\eta 2 = 0.096$. Uruguayans showed a stronger categorical effect on near trials than on far trials.

Separate two-way ANOVAs conducted for each group showed that the interaction between distance and category was significant for Uruguayans, $F(1, 30) = 11.041$, $p = 002$, $\eta 2 = 0.269$, but not for Spaniards, $F(1, 32) = 0.635$, $p = 0.902$. $\eta 2 = 0.00$. For the Uruguayan group, RTs were faster for near cross-category trials than near within-category trials (M = 1112 ms, SD = 238 ms vs. M = 1193 ms, SD = 231 ms); *post hoc* analyses were significant: $t(1, 30) = 5.312$, $p < 0.0001$, $d = 0.34$, while far cross-category trials did not differ significantly from far within-category trials (M = 922 ms, SD = 194 ms: vs. M = 944 ms, SD = 198 ms; *post hoc* analyses: $p > 0.05$) (see Figures 5 and 6).

Post hoc analyses also showed that categorical differences between countries were significant for near trials, $F(1,62) = 6.852$, $p = 0.011$, $\eta 2 = 0.100$, but not for far ones, $p > 0.05$. verbal interference condition, and slower in the no interference condition. Also, Uruguayans were less accurate than Spaniards at discriminating between near stimuli. The Uruguayan group showed more categorical effects in terms of accuracy, while both groups showed stronger categorical effects for near cases in terms of RT (with Uruguayans displaying significantly stronger effects). Finally, there was a non-significant trend for differences in the effects of verbal interference on categorical effects

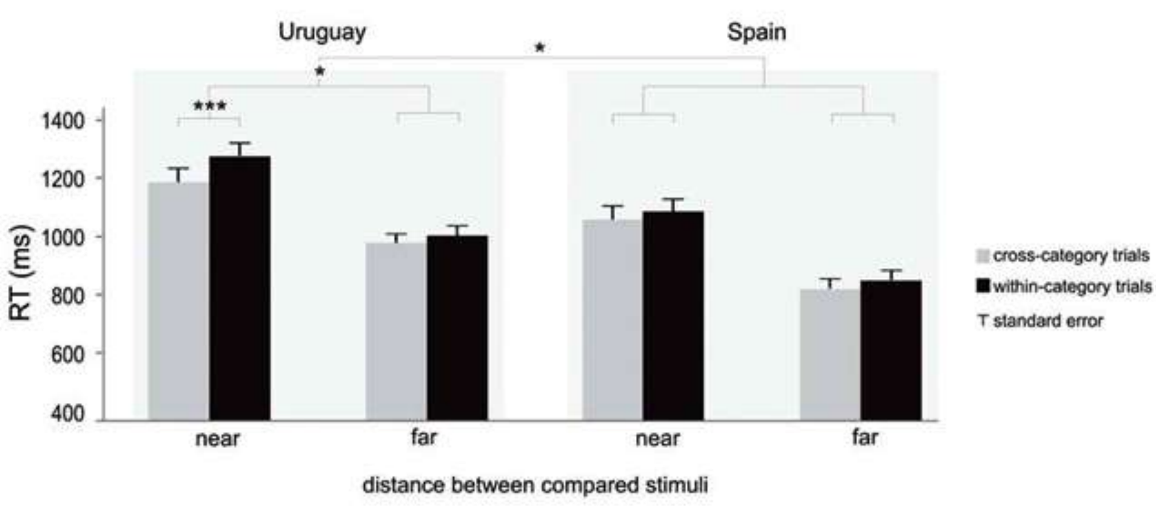


Figure 6. Mean response times (ms) between cross- and within-category stimuli by country and distance. Error bars represent SEM.

between groups for RT.

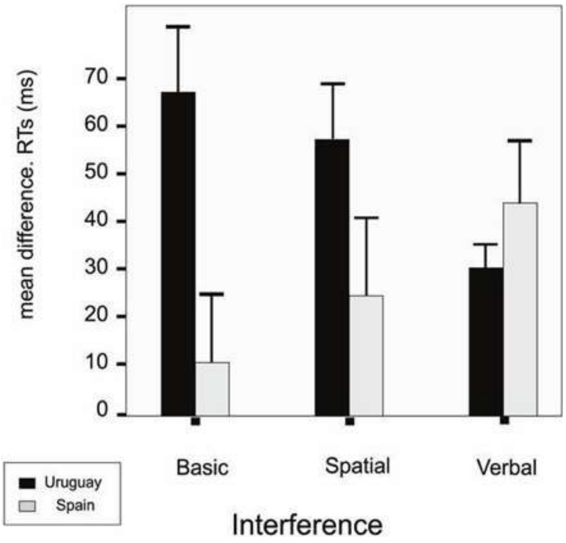


Figure 7. Category advantage (difference between cross-category and within category RT means) for the three interference conditions by group. Error bars represent SEM.

Discussion

The current study supports the Whorfian notion that language can influence color perception and is unique in that we were able to show differences in categorical effects in two groups of participants who speak the same language. Specifically, we found that Uruguayans, who have distinct color terms for light and dark blue, were more sensitive to color boundaries than Spaniards, who use a single color term for dark blue and two different compound terms for light blue. We also observed that a less frequent non-BCT—azul celeste—yielded some categorical facilitation. In this study, while both groups presented categorical effects in RT, the effect was strongest for Uruguayans on the more difficult "near" trials. Furthermore, only the Uruguayans were significantly more accurate at cross-category comparisons.

Interestingly, a non-significant difference was observed for categorical effects between countries in the different interference conditions (country by category by interference, $p = 0.059$). We calculated the differences between cross- and within-category trials to obtain a categorical effect score. Categorical effect size was greater for Uruguayans (68 ms) than Spaniards (10 ms) in the basic condition [pos hoc: $F(1,62) = 6.089$, $p = 0.016$, $e = 0.089$], more similar between groups in the spatial condition [56 vs. 24; $F(1,62) = 1.513$, $p = 0.223$, $e = 0.024$] and almost equal between groups in the verbal interference condition [30 vs. 43; $F(1,62) = 0.407$, $p = 0.526$, $e = 0.007$] (see Figure 7).

In sum, participants were faster and more accurate when discriminating between far stimuli than near stimuli and when stimuli pertained to different categories. Uruguayans were slower than Spaniards overall, less accurate and slower in the BCT—*azul celeste*—yielded some categorical facilitation. In this study, while both groups presented categorical effects in RT, the effect was strongest for Uruguayans on the more difficult "near" trials. Furthermore, only the Uruguayans were significantly more accurate at cross-category comparisons.

In contrast to previous studies where the color categories employed by the two populations clearly distinguished between dark and light blues (e.g., Russian and American participants in Winawer et al., 2007), one of the compound terms for light blue used by Spaniards (*azul celeste*) contains the monolexemic term (*celeste*) used by Uruguayans. From Lillo et al. (2016), we know that Spaniards do not consider "*celeste*" or "*azul celeste*" as a 12th BCT, as Uruguayans do, which may explain the weaker categorical effects observed among Spaniards relative to Uruguayans. Furthermore, as mentioned above, "*celeste*" is particularly salient in Uruguayan Spanish for cultural reasons, and may therefore appear more frequently for this population. According to several authors, the degree of exposure to color categories correlates with the strength of categorical effects in color discrimination tasks (Witthoft et al., 2003; Thierry et al., 2009; Athanasopoulos et al., 2011). An interesting future study would be to test category effects with a monolexemic color term whose frequency of use differs between two populations that speak the same language.

Importantly, several studies have shown that categorical effects on perception can be elicited by newly learned categories (Zhou et al., 2010; Clifford et al., 2012). In Zhou et al. (2010), participants who learned two new categories depicting light and dark shades of blue showed a categorical advantage compared with a control group, suggesting that the introduction of a novel verbal label can affect CP.

In Winawer et al. (2007), verbal interference disrupted CP for Russian but not for English speakers, suggesting a key role of language in CP (Roberson and Davidoff, 2000; Gilbert et al., 2006; Winawer et al., 2007). The results of the present study suggest that category saliency may also be affected by cultural factors.

Although the effect did not reach significance, we also observed that verbal interference diminished the categorical effect in Uruguayans and increased it in Spaniards (see Figure 7), which suggests CP effects are affected by linguistic input. Interestingly, Spaniards showed greater CP during the verbal interference block, suggesting the recruitment of the verbal label "*azul*" was inhibited. As shown by the *Stroop* effect (Stroop, 1935), automatic elicitation of a verbal label can interfere with color discrimina tion. Arguably, the discrimination between stimuli representing dark and light blues would benefit from the inhibition of the verbal label "*azul*" linked to the Spaniards' main blue category. Thus, further work is needed to clarify this issue. If replicated, it would be an unusual finding that has not been reported for English speakers in previous cross-cultural studies.

It should be noted that because part of our study was con ducted in Barcelona, some of our Spanish participants also spoke Catalan, which uses "*blau cel*" as a term for light blue. We have not studied "*blau cel*" or Catalan speakers specifically, so we cannot say whether this term is more similar to any of the terms used by Spaniards in Spanish or by Uruguayans. In order to exclude this variable as a possible confound, we conducted an additional ANOVA comparing the subset of Catalan-speaking Spaniard ($n = 18$) to the non-Catalan-speaking Spaniards ($n = 15$) and found that groups did not differ on any of the variables or interactions of interest.

A recent interpretation of Whorfian effects (proposed more than 100 years ago by William James; James, 1890) is called the *Label feedback hypothesis* (Lupyan, 2008, 2012), which proposes that labels (i.e., words) are automatically recovered to solve dif ficult discrimination cases and are recruited unconsciously when an object is perceived in order to highlight characteristic features and thus assist in the categorization process.

Furthermore, recent studies have revealed that neural net works of color perception show strong connections between basic visual areas V1 and V4 and inferotemporal and nearby regions associated with categorization (Walsh, 1999; Roe et al., 2012; Gilbert and Li, 2013; Simanova et al., 2015; Winawer and Witthoft, 2015). Moreover, an fMRI study showed activation of language regions during color perception, supporting the notion of an interaction between higher level cognition and perceptual processes (Siok et al., 2009; Brouwer and Heeger, 2013).

In the present study, perceptual processes seemed to ben efit from the words' referential attributes, but the effect differed between Spanish-speaking groups. This suggests that the interplay between categorization and perception only partially depends on a particular language's structure (Ozgen and Davies, 2002; Harnad, 2005; Lupyan et al., 2007; Collins and Olson, 2014).

An alternative interpretation is that perception could be driven by cultural—and not just linguistic—influences. In fact, cultural differences in speakers of the same language may even be the driving force behind the creation of different linguistic terms. The Emergence Hypothesis for BCTs (Kay and Maffi, 1999) proposes an explanation for how BCTs have evolved in different cultures. Kay and McDaniel (1978) suggest that derived categories are a fuzzy set of intersections among primary terms. According to this view, the emergence of a new category denoting a light shade of blue would be the result of the intersection between the blue and white categories, as Androulaki et al. (2006) proposed for Greek. Exactly why a language would add a new BCT is not clear. Casson (1997) proposed that a society's technological development will increase the importance of color as a distinguishing property of objects. Paramei (2005) and Steels and Belpaeme (2005) agree that cultural and social factors are key in the development of color lexicons. Such constraints imply that color names map onto color appearances in a culturally modal pattern (Frumkina, 1999; Jameson, 2005) and, in certain languages, could emerge as culturally basic.

Probably the main debate in linguistic relativity is whether CP occurs early on (during stimulus perception; Notman et al., 2005; Lupyan, 2012) or at the time a response is given (affecting post-perceptual processes; e.g., Pinker, 1995; Li and Gleitman, 2002). This question has been investigated using ERP, with studies showing early (Fonteneau and Davidoff, 2007; Thierry et al., 2009; Clifford et al., 2010; Mo et al., 2011; Forder et al., 2017), post perceptual (Clifford et al., 2012; He et al., 2014; Witzel and Gegenfurtner, 2016) and both effects (Holmes et al., 2009). This suggests that a strictly linguistic theory of CP is at best incomplete.

One unexpected result in the current study was that Uruguayans were both most accurate and fastest at the spatial interference block, relative to the other two blocks. One possible interpretation for this is that unlike verbal interference, spatial interference had a minimal effect on performance on a task where verbal aspects were critical, and that the added challenge resulted in higher accuracy. This would not, however, explain why that interference block would result in better accuracy than the block with no interference. We do not have enough data to answer this question at the moment but will investigate it in future studies.

Another interesting but not totally unexpected finding was that overall, Uruguayans gave slower responses than Spaniards. As observed by previous investigators, this may reflect differences in groups' experience as study participants (Witthoft et al., 2003; Winawer et al., 2007; Witzel and Gegenfurtner, 2015). In the present study, while both groups were recruited within university psychology departments, the Spanish group was generally more familiar with psychophysical experiments than the Uruguayan group. In order to ensure that categorical effects across groups were not related to overall RT, additional analyses were performed on the subset (50%) of Uruguayans with the fastest responses. Results confirmed the trends observed for the whole group. To conclude, color terms (both monolexemic and compound) carry different degrees of enhanced frequency and saliency within a linguistic community, which in turn depend on social, cultural, and historical factors (see Berlin and Kay, 1969; Casson, 1997; Kay and Maffi, 1999; Paramei, 2005, but also see Saunders, 2000). The present work shows that these differences can lead to different CP effects across groups that speak the same language.

Ethics Statement

This study was carried out in accordance with the recommendations of University of the Republic and Autonomous University of Barcelona ethics committees with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of the Republic ethics committee.

Author contributions

AA and FG-P conceived the study which was designed with the collaboration of IR and AM. IR and FG-P carried out the experi ments and the analyses were conducted by AA, IR, and FG-P. All the authors contributed to the writing of the article.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

FG-P received support from PRODIC (Programa de Desarrollo Académico de la Información y la Comunicación, FIC-UDELAR). FG-P and AM received support from CICEA (Interdisciplinary Cognition Center for Teaching and Learning-UDELAR).

References

Androulaki, A., Gômez-Pestaña, N., Mitsakis, C., Jover, J. L., Coventry, K., and Davies, I. (2006). Basic colour terms in modern Greek: twelve terms including two blues. *J. Greek Linguist. 7*, 3–47. https://doi:10.1075/jgl.7.03

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: the case of Greek blues*. *Bilingualism 12*, 83–95. https://doi:10.1017/S136672890800388X

Athanasopoulos, P., Damjanovic, L., Krajciova, A., and Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: the case of Japanese blues. *Bilingualism 14*, 9–17. https://doi:10.1017/S1366728909990046

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J. R., and Thierry, G. (2010). Perceptual shift in bilingualism: brain potentials reveal plasticity in pre attentive colour perception. *Cognition 116*, 437–443. https://doi:10.1016/j.cognition.2010.05.016

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci. 15*, 600–609. https://doi:10.1162/089892903321662976

Berlin, B., and Kay, P. (1969). Basic color terms: their universality and evolution. *David Human Series Philos. Cogn. Sci. Reissues 19*, 178.

Bimler, D., and Uusküla, M. (2014). "Clothed in triple blues": sorting out the Italian blues. *J. Opt. Soc. Am. A* 31, A332–A340. doi:10.1364/JOSAA.31.00A332 Bloom, P., and Keil, F. C. (2001). Thinking through language. *Mind Lang.* 16, 351–367. https://doi:10.1111/1468-0017.00175

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn. Psychol. 43*, 1–22. https://doi:10.1006/cogp.2001.0748

Brouwer, G. J., and Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *J. Neurosci.* 33, 15454–15465. https://doi:10.1523/JNEUROSCI.2472-13.2013

Brown, R. (1976). Reference in memorial tribute to Eric Lenneberg. *Cognition 4*, 125–153. https://doi:10.1016/0010-0277(76)90001-9

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Lang. Learn. 58*, 63–79. https://doi:10.1111/j.1467-9922.2008.00462.x

Casson, R. W. (1997). "Color shift: evolution of English color terms from brightness to hue," in *Color Categories in Thought and Language*, eds C. L. Hardin and L. Maffi (Cambridge: Cambridge University Press), 225–239.

Clifford, A., Franklin, A., Holmes, A., Drivonikou, V. G., Özgen, E., and Davies, I. R. (2012). Neural correlates of acquired color category effects. *Brain Cogn. 80*, 126–143. *https://doi:10.1016/j.bandc.2012.04.011*

Clifford, A., Holmes, A., Davies, I. R., and Franklin, A. (2010). Color categories affect pre-attentive color perception. *Biol. Psychol. 85*, 275–282. https://doi:10.1016/j. biopsycho.2010.07.014

Collins, J. A., and Olson, I. R. (2014). Knowledge is power: how conceptual knowl edge transforms visual cognition. *Psychon. Bull. Rev. 29*, 843–860. https://doi:10.3758/s13423-013-0564-3

Corbett, G. G., and Davies, I. R. L. (1997). "Establishing basic color terms: measures and techniques," in *Color Categories in Thought and Language*, eds C. L. Hardin and L. Maffi (Cambridge: Cambridge University Press), 197–223.

Fonteneau, E., and Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport 18*, 1323–1327. https://doi:10.1097/WNR.0b013e3282c48c33

Forder, L., He, X., and Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS ONE* 12:e0178097. https://doi:10.1371/journal.pone.0178097

Frumkina, R. M. (1999). What does my eye tell your mind? *Behav. Brain Sci.* 22, 951–952. https://doi:10.1017/S0140525X99302211

Gilbert, A. L., Regier, T., Kay, P., and Ivry, R. B. (2006). Whorf hypothesis is sup ported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U.S.A.* 103, 489–494. https://doi:10.1073/pnas.0509868103

Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. https://doi:

10.1038/nrn3476
Harnad, S. (2005). "To cognize is to categorize: cognition is categorization," in *Handbook of Categorization in Cognitive Science*, eds H. Cohen and C. Lefebvre (Oxford: Elsevier Science), 19–43.
He, X., Witzel, C., Forder, L., Clifford, A., and Franklin, A. (2014). Color cate gories only affect post-perceptual processes when same-and different-category colors are equally discriminable. *JOSA A* 31, A322–A331. https://doi:10.1364/ JOSAA.31.00A322
Holmes, A., Franklin, A., Clifford, A., and Davies, I. (2009). Neurophysiological evidence for categorical perception of color. *Brain Cogn.* 69, 426–434. https://doi:10.1016/j.bandc.2008.09.003
James, W. (1890). *The Principles of Psychology*, Vol. I. New York: Henry Holt. Jameson, K. A. (2005). Culture and cognition: what is universal about the representation of color experience? *J. Cogn. Cult.* 5, 293–348. https://doi:10.1163/156853705774648527
Kay, P., and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *Am. Anthropol.* 86, 65–79. doi:10.1525/aa.1984.86.1.02a00050 Kay, P., and Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *Am. Anthropol.* 101, 1–32. doi:10.1525/aa.1999.101. 4.743
Kay, P., and McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language* 54, 610–646. https://doi:10.2307/412789
Levelt, W. (2014). *A History of Psycholinguistics: The Pre-Chomskyan Era*. Oxford: Oxford University Press.
Li, P., and Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition* 83, 265–294. https://doi:10.1016/S0010-0277(02)00009-4
Lillo, J., Moreira, H., Vitini, I., and Martin, J. (2007). Locating basic Spanish colour categories in CIE L*u*v* space: identification, lightness segregation and corre spondence with English equivalents. *Psicologica* 28, 21–24.
Lillo, J., Prado-León, L., Gonzalez, F., Alvaro, L., Moreira, H., and Melnikova, A. (2016). *Spanish Basic Color Categories Are 11 or 12, Depends on the Dialect, Presented at Progress in Colour Studies Conference, London, 2016*. London: University College of London.
Lucy, J., and Shweder, R. (1979). Whorf and his critics: linguistic and nonlinguistic influences on color memory. *Am. Anthropol.* 81, 581–615. https://doi:10.2307/675777
Lucy, J. A. (1997). Linguistic relativity. *Annu. Rev. Anthropol.* 26, 291–312. https://doi:10.1146/annurev.anthro.26.1.291
Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108, 566–577. https://doi:10.1016/j. cognition.2008.03.009
Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3:54. https://doi:10.3389/fpsyg.2012.00054
Lupyan, G., and Clark, A. (2015). Words and the world: predictive coding and the language-perception-cognition interface. *Curr. Direct. Psychol. Sci.* 24, 279–284. https://doi:10.1177/0963721415570732
Lupyan, G., Rakison, D. H., and McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychol. Sci.* 18, 1077–1083. https://doi:10.1111/j.1467-9280.2007.02028.x
Mo, L., Xu, G., Kay, P., and Tan, L. H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14026–14030. https://doi:10.1073/pnas.1111860108
Notman, L. A., Sowden, P. T., and Özgen, E. (2005). The nature of learned cate gorical perception effects: a psychophysical approach. *Cognition* 95, B1–B14. https://doi:10.1016/j.cognition.2004.07.002
Ozgen, E., and Davies, I. R. L. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol.* 131, 477–493. https://doi:10.1037/0096-3445.131.4.477
Paramei, G. V. (2005). Singing the Russian blues: an argument for culturally basic color terms. *Cross-Cult. Res.* 39, 10–38. https://doi:10.1177/1069397104267888
Pinker, S. (1995). Language acquisition. *Language* 1, 135–182. Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365. https://doi:10.1017/S0140525X99002022
Roberson, D., and Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cognit.* 28, 977–986. https://doi:10.3758/BF03209345
Roberson, D., Hanley, J. R., and Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition* 112, 482–487. https://doi:10.1016/j. cognition.2009.06.008
Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., et al. (2012). Toward a unified theory of visual area V4. *Neuron*. 74, 12–29. https://doi:10.1016/j.neuron.2012.03.011
Saunders, B. (2000). Revisiting basic color terms. *J. R. Anthropol. Inst.* 6, 81–99. https://doi:10.1002/col.5080170514
Steels, L., and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *Behav. Brain Sci*. 28, 469–488.
Simanova, I., Francken, J. C., de Lange, F. P., and Bekkering, H. (2015). Linguistic pri ors shape categorical perception. *Lang. Cogn. Neurosci.* 3798, 1–7. https://doi:10.1080/23273798.2015.1072638
Siok, W. T., Kay, P., Wang, W. S., Chan, A. H., Chen, L., Luke, K. K., et al. (2009). Language regions of brain are operative in color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8140–8145. https://doi:10.1073/pnas.0903627106
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643. https://doi:10.1037/h0054651
Thierry, G. (2016). Neurolinguistic relativity: how language flexes human perception and cognition. *Lang. Learn.* 66, 690–713. https://doi:10.1111/lang.12186
Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., and Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4567–4570. https://doi:10.1073/pnas.0811155106
Vygotsky, L. S. (1987). Thinking and speech. *The Collected Works of LS Vygotsky*, Vol. 1, New York: Plenum Press, 113–114.
Walsh, V. (1999). How does the cortex construct color? *Proc. Natl. Acad. Sci. U.S.A.* 96, 13594–13596. https://doi:10.1073/pnas.96.24.13594
Whorf, B. L. (1956). *Language, Thought, and Reality*. Cambridge, MA: MIT Press. Winawer, J., and Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Vis. Neurosci.* 32, E020. https://doi:10.1017/S0952523815000176
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785. https://doi:10.1073/pnas.0701644104
Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A., and Boroditsky, L. (2003). "Effects of language on color discriminability," in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 1247–1252.
Witzel, C., and Gegenfurtner, K. R. (2015). Categorical facilitation with equally discriminable colors. *J. Vis.* 15, 22. https://doi:10.1167/15.8.22
Witzel, C., and Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *J. Exp. Psychol.* 42, 540–570. https://doi:10.1037/xhp0000154
Zhong, W., Li, Y., Huang, Y., Li, H., and Mo, L. (2017). Is the lateralized categorical perception of color a situational effect of language on color perception? *Cogn. Sci.* https://doi:10.1111/cogs.12493
Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., and Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9974–9978. https://doi:10.1073/pnas.1005669107
Zlatev, J., and Blomberg, J. (2015). Language may indeed influence thought. *Front. Psychol.* 6:1631. https://doi:10.3389/fpsyg.2015.01631

Matemáticas monstruosas
Monster math

Se trata de la portada del libro de ejercicios que utilizamos en la intervención realizada en escuelas en el proyecto desarrollado en conjunto con Justin Halberda (Universidad Johns Hopkins) en 2016. El proyecto buscaba estimular las bases cognitivas de la matemática temprana a través de ejercicios de estimación de magnitudes. Este libro de actividades, acompañaba un juego para tablet también inspirado en monstruos que plantean diferentes retos de estimación de magnitudes a los niños. El juego, desarrollado para Android, puede descargarse aquí

This is the cover of the exercise book that we used in the intervention carried out in schools in the project developed together with Justin Halberda (Johns Hopkins University) in 2016. The project sought to stimulate the cognitive foundations of early mathematics through magnitude estimation exercises. This activity book, accompanied a tablet game also inspired by monsters that pose different magnitude estimation challenges to children. The game, developed for Android, is available for download here.

School Readiness losses during the COVID-19 outbreak. A comparison of two cohorts of young children

Meliza González[1], Tianna Loose[1], Maite Liz[1], Mónica Pérez[1], Juan I. Rodríguez-Vinçon[1], Clementina Tomás[1] and Alejandro Vásquez-Echeverría[1]

Tianna Loose https://orcid.org/0000-0002-2656-2760, Clementina Tomas-Llerena https://orcid.org/0000-0002-5125-3935, Alejandro Vásquez-Echeverría https://orcid.org/0000-0002-9538-4857

1 Facultad de Psicología y Centro de Investigación en Cognición para la Enseñanza y el Aprendizaje. Universidad de la República, Uruguay

Correspondence concerning this manuscript should be addressed to Alejandro Vásquez-Echeverría, Facultad de Psicología, Tristán Narvaja 1674, Montevideo 11200, Uruguay. Email: **alejandro.vasquez@pedeciba.edu.uy**

This is a free version -with the approval of the authors- of the pre-print of the manuscript González, M., Loose, T., Liz, M., Pérez, M., Rodríguez−Vinçon, J. I., Tomás−Llerena, C., & Vásquez−Echeverría, A. (2022). School readiness losses during the COVID−19 outbreak. A comparison of two cohorts of young children. *Child Development 93* (4), 910-924. Available at **https://doi.org/10.1111/cdev.13738**

Abstract

The COVID-19 context has created the most severe disruption to education systems in recent history. Its impact on child development was estimated comparing two cohorts of 4 to 6-year-old Uruguayan children: control (n = 34,355, 48.87 % girls) and COVID cohort (n = 30,158, 48.95% girls) assessed between 2018 and 2020 in three waves, by a routinely administered school readiness instrument in public preschools. Ethnicity information is not available. For the COVID cohort, losses were observed in Motor and Cognitive development, Attitudes towards learning, and Internalizing behavior (range β = 0.27 - 0.13). Losses were less pronounced among children from higher socioeconomic schools. These results extend the literature on the consequences of the pandemic on learning and early child development.

Keywords: school readiness, COVID-19, school closure, early child development, preschool

Introduction

Early childhood development and COVID-19

The COVID-19 pandemic is a major socio-historical event that continues to severely disrupt people's lives (Benner & Mistry, 2020). In the past, researchers have shown that other large scale events (e.g. the SARS and N1H1 epidemics, the 2008 recession, the great depression, and World War II) were powerful enough to divert children from normative developmental trajectories with long-lasting effects (Almeida & Wong, 2009; Benner & Mistry, 2020). For example, children who received less education during WWII experienced adverse effects 40 years later (Ichino & Winter−Ebmer, 2004). The impact of such societal events is even more potent at ages where typical developmental turning points occur, such as school entry at age four or the transition from childhood into adolescence (Almeida & Wong, 2009; Benner & Mistry, 2020). Our study focuses on developmental losses among children who transitioned into school entry during the COVID-19 pandemic.

In the context of COVID-19, school closures are likely the most relevant factor disrupting the lives of young children, a factor traditionally responsible for learning losses that has affected 94% of learners worldwide (United Nations, 2020). However, it is impossible to isolate the impact of school closures on learning or developmental losses (Bacher-Hicks & Goodman, 2021). Instead, we should consider the pandemic as a myriad of interactive factors contributing to hardships among young learners (Bacher-Hicks & Goodman, 2021) related to the impact of the pandemic on children's families (e.g., job loss, financial losses, remote working, illness, death, stress, mental health and improvised parenting practices), teachers (e.g., stress, the sudden switch to online learning), social lives (e.g., loss of social contact) and access to services (e.g., daycare, delayed healthcare visits) (Ananat & Gassman-Pines, 2020; Bacher-Hicks & Goodman, 2021; Rothstein, 2020). Nevertheless, even without the added burden of societal catastrophes, studies evidenced that routine school closures during the summer months provoke learning losses (e.g., Alexander et al., 2007). As a whole, we expect that the current pandemic context would negatively impact young children's school readiness, but there is no research on the topic to our knowledge. In light of this large gap, the effect of the pandemic on early childhood development needs to be a priority for researchers (Benner & Mistry, 2020; Yoshikawa et al., 2020).

Recent studies suggest that the pandemic has a negative impact on school performance among older children and adolescents. When elementary schools closed for eight weeks in the Netherlands, children lost the equivalent of 20% of what would be achieved during a typical school year (Engzell et al., 2021). Another study examined achievements in reading and math among three million elementary school children. When comparing children who were exposed to the pandemic to a reference group that was not, the authors found that scores in math dropped five to 10 percent but reading scores did not significantly differ (Kuhfeld et al., 2020). Among children in Belgium, scores in math (SD = 0.19) and Flemish language (SD = 0.29) dropped more in a cohort of children exposed to the pandemic than in a reference cohort (Maldonado & De Witte, 2020). On the basis of the extant literature, we expect that the pandemic negatively impacted school readiness among young children. This is important because school readiness is associated with achievement during elementary school (Duncan et al., 2007, 2020) and later on (Watts et al., 2014). Other studies evidenced that isolation and interpersonal relationships impacted areas of children's mental health, wellbeing (Araújo et al., 2020; Guerrero, 2021; Liu et al., 2021; Loades et al., 2020; Spiteri, 2021) and physical activity (Gobbi et al., 2020). This could threaten the United Nations (2021) Sustainable Development Goal 4 promoting "*inclusive and equitable quality education and promote lifelong learning opportunities for all*", especially among children with underprivileged backgrounds.

Widening achievement gaps and inequity

Before the pandemic, large achievement gaps distinguished learners according to household income (e.g., Chmielewski & Reardon, 2016; Papay et al., 2020). These gaps are exacerbated because of the differential impact the pandemic has on children according to factors related to their socio-economic status. For example, low-income parents are more likely to be essential workers, with the risk of increased COVID-19 transmission, and not be able to care for their children during confinement periods. Also, low-income households are less likely to have a quality internet connection which is necessary for online learning. Financial strain also makes it less likely that parents can afford private tutoring to compensate for school closures (Bailey et al., 2021). Furthermore, children from lower-income households are more likely to experience overcrowding in small houses, food insecurity, unstable home life, and mental health issues, including trauma, anxiety, and depression (Masonbrink & Hurley, 2020). These children suffer more from the impact of school closures, thus increasing achievement disparities (Atteberry & McEachin, 2021; Steward et al., 2018).

These disparities translate directly to inequity in learning losses during the COVID-19 pandemic. When 200 education researchers were asked to project the extent to which the achievement gap will widen among school-age children in the year following the pandemic, respondents estimated that scores

Table 1. Timeline of data collection and COVID mitigation measures in the school context

Date	Semester	COVID measures	Control cohort	COVID cohort
Year 2018				
March-July	semester 1	N/A	age 4, wave 1	
Aug.-Dec.	semester 2	N/A	age 4, wave 2	
Year 2019				
Jan.-Feb. 2019	vacation	N/A		
March-July 2019	semester 1	N/A		age 4, wave 1
Aug.-Dec. 2019	semester 2	N/A	age 5, wave 3	age 4, wave 2
Year 2020				
Jan.-Feb.	vacation	none		
March 1-15	semester 1	schools open		
March 15-June 15	semester 1	schools closed*		
June 15-July	semester 1	partial opening		
Aug-Oct. 14	semester 2	partial opening		
Oct. 15-Dec	semester 2	schools open		age 5, wave 3

Note. * Except for schools in rural areas that reopened on April 24th. N/A = Not applicable.

in reading would widen from 1 standard deviation point to 1.25, and that the standard deviation in math scores would widen to 1.30. Furthermore, respondents projected that these alarming disparities would last two years after the onset of the pandemic (Bailey et al., 2021). Researchers in the Netherlands found that the negative impact on learning was 55% larger among students who had less educated parents (Engzell et al., 2021). In Chetty et al.'s study (2020), achievement fell by 30% for higher-income students and then quickly bounced back after schools reopened, whereas achievement fell by 50% among lower-income students and stayed low throughout the school year. In the context of COVID-19, inequity in school readiness losses among underprivileged preschoolers is expected (e.g., Benner & Mistry, 2020; Yoshikawa et al., 2020), but has never been studied to our knowledge. Quantifying these disparities can help inform public policy and allocate the necessary resources to the most vulnerable students.

The current study

In Uruguay and elsewhere the unprecedented COVID-19 pandemic led to the suspension of face-to-face classes. Nonetheless, deaths per capita in Uruguay were among the lowest in South America by the end of 2020, which, unlike in other countries, allowed educational authorities to restore face-to-face activities during the last period of the year, including the administration of a nationwide school readiness assessment that has been implemented regularly since 2018 within public schools. This scenario provided the setting for a natural experiment on the impact of the COVID context (e.g., school closures, economic crisis) on child development. To our knowledge, no research has quantified the impact of the pandemic on pre-schoolers' development. Though all children may be suffering developmentally in the context of COVID-19, those from less privileged communities would be the hardest hit.

Table 2. Children assessed with INDI and reasons for sample exclusion, by cohort

	Control cohort	COVID cohort
Children with at least one assessment at wave 2 and 3	39359	39329
Attrition wave 2 and 3	265	432
Attrition wave 2	3601	2195
Attrition wave 3	1138	6544
Total attrition	5004	9171
Total children assessed at wave 2 and 3 (final sample in DID analyses)	34355	30158
Total children assessed at wave 1, 2 and 3	33255	29823

Table 3. Distributions of age, sex, school quintile and school typology for each cohort

	Control cohort	COVID cohort	Statistic	
Mean age in months (SD)				
Wave 1	53.57 (3.47)	53.78 (3.52)	-7.96	***
Wave 2	60.18 (3.49)	59.69 (3.52)	17.91	***
Wave 1	71.65 (3.49)	71.88 (3.52)	-8.16	***
Girls (%)	48.87	48.95	-0.2	
School quintile (%)				
Q1	17.9	17.5	0.03	
Q2	20.5	20.4	3.32	**
Q3	18.9	18.2	5.04	*
Q4	18.6	20	7.16	**
Q5	22.3	22.3	0.26	
No data	1.8	1.6	11.5	***
Type of school (%)				
Simple (4 hours a day)	39.47	38.05	13.65	***
Full day (8 hours a day)	21.96	21.74	0.47	
Extended (4 hours + lunch)	2.54	3.42	42.87	***
APRENDER	23.23	23.54	0.87	
Other (e.g., teacher traning, rural)	12.79	13.25	2.93	*
N	34355	30158		

* p < .05; ** p < .01; p < .001

Accordingly, this study aims to provide insights into the heterogeneity of this impact across different socioeconomic settings. Even if this study did not undergo pre-registration, on the basis of the extant literature we expected to find moderate-to-small losses in school readiness among the COVID Cohort, and that these losses would be larger among children in schools from lower-SES districts. Thus, this study can be considered confirmatory.

The context of the current study

The impact of the pandemic in Uruguay Reactions to the COVID-19 pandemic differed across countries and our study was conducted in Uruguay. Uruguay is situated in South America and has a small-density population of 3.5 million inhabitants and moderate-to-high life expectancy at birth (77.1 years, 44th in global rank; World Health Organization, 2021). In reaction to the pandemic, the Uruguayan government communicated recommendations to citizens (guidelines for hand sanitization, masks, social distancing, self-isolation, etc.) and framed public campaigns with messages to act responsibly. Though the government never legally imposed a lockdown, other restrictions were implemented including on social gatherings and international travel. In the first months of the pandemic, many restaurants and shops were closed and were allowed to reopen progressively around June 2020. Public offices reopened in April or May for strictly necessary purposes, but most of the work shifted to distance. In that context, unemployment rose approximately 1 percentage point from March to October. There was a clear impact on the economy with GDP falling 5.9% in 2020 (BCU, 2021). The active cases per capita in Uruguay remained low until the end of December 2020 when numbers started to rise.

Timeline of school closures in Uruguay The World Health Organization declared that the COVID-19 outbreak was a public health emergency of international concern on January 30, 2020. China was the first country to close schools in mid-February, and by mid-March, about 135 countries were affected by school closures (UNESCO, 2020). Uruguay implemented a certain degree of school attendance during 2020 because of the good epidemiological situation. In Uruguay, the school year runs from March to mid-December rather than from September to June. Schools were open as usual for the first two weeks of March 2020 and then faced the first nationwide school closures. Distance learning was implemented using home devices and virtual educational platforms from "Plan Ceibal". Plan Ceibal is a governmental socio-educational project in Uruguay, widely accepted by families and education communities, that supports education through technology. It was created in 2007 and inspired by the One Laptop Per Child project (One Laptop Per Child, n.d.). Nonetheless, the coverage of Plan Ceibal within the preschool system is limited (e.g., devices are not delivered to preschoolers), making distance learning particularly challenging for very young learners.

Under these conditions, classes were held entirely online for three months (mid-March through mid-June), with the exception of a small number of children attending schools in the low-density populated rural areas, which reopened on April 22th. Afterwards children were able to return gradually to face-to-face classes on a voluntary basis. On June 15, 2020, non-mandatory in-person classes were held and most children attended school twice or three times a week in compliance with sanitary measures (e.g. reduced class size and increased social distancing). Home-schooling was used to compensate for the reduced hours. This scenario carried on from June 15 to August 2020 when schools were able to hold in-person activities depending on the number of children enrolled and the physical size of the school building. On October 13, 2020, face-to-face education was fully restored and continued until the school year ended (Presidencia de la República, 2021). A typical school year in Uruguay lasts 185 days (National Administration of Education (ANEP), 2021). In 2020, schools were open on average 80.6 days, but variability was high (56 - 113 days for percentiles 10 and 90, respectively). In Uruguay, children in age 5 classrooms attended school 53.1 days on average (ANEP, 2021), and 104 school days were lost. These numbers are by far better than those of Latin America and the Caribbean, which had an average of 158 days of school closures, and South Asia with 146 days (Unicef, 2021).

In Uruguay, education is mandatory starting at age 4, and coverage for preschool and primary school is almost universal. By 2018, 94% of four- and 99% of five-year-olds attended preschools (INEEd, 2019). Education providers can be either public, which means government-funded and managed schools that provide free education, or private, which rarely are free but can sometimes provide scholarships (e.g., some Catholic schools in low-income districts). Traditionally, the main provider of education has been the national government, with a strong network of public schools including in low-populated rural areas. In 2020, the public preschool and primary school system reached coverage of 81.7% of all children (ANEP-DGEIP, 2021). Teachers should have an official tertiary degree. Until 2017 there was no differentiation between primary and preschool teachers' education, in a four-year curriculum focused on primary education content. As of 2017, a new four-year degree was created for early education with a curriculum aligned with the idea of stimulating child development rather than predominantly teaching academic content (ANEP-CFE, 2016).

Methods

Participants and procedure We used longitudinal cohort data collected through the Uruguayan School Readiness-Child Development Inventory (INDI, by its Spanish acronym) among children attending public schools in Uruguay. Children's information (age and gender) and information about schools' SES (quintile) were also provided from an anonymized administrative dataset in agreement with the educational authority (CODICEN). This source of data does not collect information about ethnicity. However, according to the last Uruguayan census, 93.9% of individuals self-reported European ancestry, 8.1% reported being Afro-Uruguayan and 5.1% reported Indigenous ancestry (Cabella et al., 2013). Note that self-reporting multiple ethnic identifications was allowed.

We included two cohorts of children aged 4 to 6 years old enrolled in Uruguayan public preschool in age 4 and 5 classrooms. The 2019-2020 cohort of children was naturally exposed to the COVID-19 pandemic at age 5 (mainly focused on school closures) and will be referred to as the "COVID cohort". The 2018-2019 cohort was used as a reference group or "control cohort" since the children were not exposed to the pandemic at age 5. Table 1 shows a timeline of the Uruguayan school year, COVID-19 mitigation measures impacting school closures, and features the three waves of data collection in each cohort. Wave 1 of data collection was carried out among 4-year-old children in the first semester of the school year (age 4 classroom). Wave 2 data was collected in the second semester (age 4 classroom). Note that at wave 1 and wave 2 neither the control cohort nor the COVID cohort was exposed to the pandemic. Wave 3 was collected one year after wave 2 (age 5 classrooms) at a time where the COVID cohort had been naturally exposed to the pandemic (year 2020), whereas the control cohort had not (year 2019).

Nearly 40,000 children per cohort were administered at least one assessment at waves 1, 2, or 3. However, we discarded participants from analyses if they were missing at least one assessment at waves 2 or 3. This led us to discard 13% of the control cohort and 23% of the COVID cohort, leading to a final sample of 34,355 and 30,158 participants, respectively (Table 2). Within the COVID cohort, attrition was due to children who lacked an assessment at wave 2 (24%), wave 3 (71%), or both

Table 4. Means (and standard deviations) of INDI's scores in all domains and subscales for each wave and cohort

	Wave 1				Wave 2				Wave 3			
Cohort	Control		COVID		Control		COVID		Control		COVID	
Language	23.44	(10.47)	24.25	(10.46)	34.92	(13.90)	35.10	(13.58)	44.75	(14.25)	42.23	(14.60)
Logical-mathematical	17.3	(8.35)	18.19	(8.33)	28.44	(10.97)	28.64	(10.59)	37.7	(9.99)	36.11	(10.59)
Self-projection	25.25	(9.05)	26.36	(8.69)	32.6	(9.35)	33.05	(8.89)	37.83	(8.54)	36.33	(8.78)
Executive functioning	31.32	(8.26)	31.87	(8.57)	35.03	(7.66)	35.93	(8.05)	38.3	(7.61)	38.65	(7.15)
Internalizing	10.18	(4.89)	10.02	(4.77)	9.09	(4.35)	8.83	(4.19)	8.3	(3.93)	8.48	(4.12)
Externalizing	14.97	(8.72)	14.85	(8.52)	14.57	(8.52)	14.20	(8.30)	13.91	(8.24)	11.76	(6.25)
Prosocial	24.15	(6.48)	24.53	(6.23)	27.41	(6.19)	27.83	(6.02)	29.2	(5.83)	29.4	(5.50)
Cognitive development	59.04	(19.65)	61.25	(19.47)	81.79	(23.85)	82.76	(23.29)	99.97	(23.20)	96.42	(23.80)
Motor development	20.16	(5.32)	20.66	(4.97)	24.75	(4.89)	24.95	(4.61)	27.57	(3.90)	26.73	(4.23)
Attitudes t/ learning	27	(7.15)	27.31	(7.02)	30.52	(7.15)	30.91	(6.95)	32.74	(6.86)	32.43	(6.76)

wave 2 and 3 (i.e., only wave 1 assessment, 5%). There was more missing data in the COVID cohort, notably at wave 3, because, within the context of the health emergency, teachers were given the option to opt-out of evaluating a given child, and more children did not attend school.

We analyzed the impact of attrition on mean INDI scores. We found that children who did not complete wave 3 had slightly lower average INDI scores at waves 1 and 2. On the other hand, those who were excluded because of a lack of assessment in wave 2, most often had higher means than those who were evaluated. The differences in both cases are small, and do not affect the general average, i.e., the means in waves 2 and 3 for the total cohort and valid cases are not statistically significantly different (see Supplementary Material 1).

Across waves and cohorts, data was reported from 1554 schools with an average of 26.24 children per center. Data on gender, age, socioeconomic school quintiles, and school typologies for the two cohorts are provided in Table 3. In order to assess the equivalency of cohorts across these characteristics, we used Student's *t* tests (numerical variables) and Pearson's chi-squared test (categorical variables). We found statistically significant differences in age between the COVID and Control cohort. This was due to small differences in the period of assessment across waves and years. No sex differences were found in the distribution by cohort. The cohorts significantly differed on school quintiles and by school typology.

Table 5. Cohorts comparison according to DiD estimates without controls and controlling for age, gender, and socio-economic status (main specification)

	Unstandardized estimates (SE)		Standardized estimates (SE)	
	Bivariate	Regression -controls added	Bivariate	Regression -controls added
Language	-2.69*** (-0.46)	-3.31*** (-0.45)	-0.18*** (0.03)	-0.22*** (0.03)
Logical-mathematical	-1.79*** (-0.30)	-2.34*** (-0.29)	-0.16*** (0.02)	-0.21*** (0.02)
Self-projection	-1.95*** (-0.29)	-2.32*** (-0.29)	-0.21*** (0.03)	-0.25*** (0.03)
Executive functioning	-0.55** (-0.23)	-0.74*** (-0.23)	-0.07** (0.03)	-0.10*** (0.02)
Prosocial behaviors	-0.23 (-0.21)	-0.33 (-0.20)	-0.04 (0.03)	-0.06 (0.03)
Internalizing behaviors	0.44*** (-0.13)	0.52*** (-0.13)	0.11*** (0.03)	0.13*** (0.03)
Externalizing behaviors	-1.78*** (-0.20)	-1.76*** (-0.19)	-0.22*** (0.02)	-0.22*** (0.02)
Cognitive development	-4.51*** (-0.72)	-5.72*** (-0.69)	-0.18*** (0.02)	-0.23*** (0.02)
Motor development	-1.05*** (-0.14)	-1.24*** (-0.14)	-0.23*** (0.02)	-0.27*** (0.03)
Attitudes t/ learning	-0.69*** (-0.21)	-0.98*** (-0.21)	-0.10*** (0.03)	-0.14*** (0.02)

Note. Negative values indicate skills loss in the COVID cohort, except for Internalizing and Externalizing behaviors in which negative values indicate improvement. SE = Standard error. ***p < .001, ** p < .01

Measures

Socio-demographic information was collected using an administrative source including age, gender, and the schools' socioeconomic status (SES) quintile (1 = lowest, 5 = highest). ANEP assigns quintiles to schools based on the average socioeconomic information of the enrolled children's families. Thus, it is possible but unlikely that children coming from low-income families are enrolled in a high-SES rating school and vice-versa.

The School Readiness - Child Development Inventory (INDI) was used. INDI is a norm-referenced, teacher-reported school readiness assessment carried out in the educational context. It is implemented by the authorities to improve school readiness practices and better target educational and health protection measures for vulnerable children. Teachers receive a 2-hour annual training on administration and the use of the automatic reports system for educational interventions. Further reading of the manuals is strongly recommended as they provide specific details for scoring and interpreting reports. Teachers can also attend optional courses throughout the school year. To administer INDI, teachers observe the child's frequency of specific behaviors in certain developmental milestones, during a typical school day within a period of three to four weeks. Most items are observational, but some require individualized testing for scoring (e.g., "Holds the pencil properly") with specifically designed activities and response options. Teachers score indicators on a digital platform using a 6-point frequency Likert scale ranging from "never" to "always". Teachers can also indicate any reasons that prevent evaluation (e.g., frequent absences, severe developmental difficulties) (see Supplementary Material 2). Detailed information about scoring is available on the digital platform where teachers complete the INDI and in the administration manual.

INDI accounts for four developmental domains, some of which include subscales (Vásquez-Echeverría et al., 2021): (a) Cognitive development is composed of Language (including oral comprehension and production, phonological awareness, and early literacy skills, e.g., *Understands a short story;* 7 items), Logical-mathematical skills (early numeracy skills such as verbal counting, recognizing number symbols, shapes, and manipulating quantities, e.g., *Recognizes numbers between 1 and 10,* 6 items), Self-projection (including perspective-taking abilities, as in the case of episodic memory, foresight and theory of mind; Buckner and Carroll, 2007, e.g., *Anticipates what he will need in the future,* 6 items) and Executive functioning (including mainly attentional and self-regulation skills, e.g., *Is able to wait for turns,* 6 items). Cognitive subscales scores are summed up to yield a Cognitive development total score; (b) Motor development (mostly composed of content tapping fine-motor skills such as pencil-grip and gross-motor skills such as locomotion, e.g., *Holds the pencil properly,* 6 items); (c) three orthogonal subscales compose Socioemotional development: Internalizing behaviors (behavior problems such as anxiety, sadness or behavioral inhibition, e.g., S*pends time alone, isolated from the group,* 5 items), Externalizing behaviors (behavior problems such as verbal or physical aggression, defiant attitudes towards teachers and peers, e.g., *Verbally assaults his/her peers,* 4 items) and Prosocial behaviors (cooperative and empathic attitudes, e.g., *Shares toys and materials,* 5 items). No composite socioemotional score is provided, and (d) Attitudes towards learning (including school adaptation, motivation and creativity, e.g., *Shows curiosity and interest in class;* 6 items). The INDI version used in the current study is adapted to children aged 49 to 79 months old (age 4 and age 5 classrooms) and totals 52 items. The INDI Automatic Reports System provides feedback about child performance on each domain and subscale comparing individual results with a normative sample using four categories: risk, monitoring zone, on track, and outstanding. Previous studies suggested that INDI scores have good-to-excellent internal consistency (α range .73 - .93), test-retest stability (*r* range .80 - .99), and inter-rater reliability (ICC range .71 to . 95) coefficients, as well as convergent validity estimates (Vásquez-Echeverría, 2020). INDI was administered to preschoolers nationwide from 2018 to 2020. Though the administration of INDI is not compulsory, it was administered to 94.8% of preschoolers in 2018, 96.1% in 2019, and 70.3% in 2020. More information about the INDI assessment system can be found at https://indi.psico.edu.uy/en.

Data Analysis To study the impact of the COVID-19 pandemic on developmental markers, we compared levels of development across the two cohorts of children (COVID cohort and control cohort). This counterfactual scenario allowed us to infer what we would expect if the children had not been exposed to the pandemic.

We estimated the health emergency's impact on developmental markers using differences-in-differences analysis (DiD) (Angrist & Pischke, 2009). This regression-based approach was used to compare developmental changes of the two cohorts across the three waves of data collection while reducing bias from unobserved variables. DiD analysis operates on the assumption that scores would evolve in the same way across the two cohorts of data (parallel trends), had the treatment or event (e.g. pandemic) not occurred. This quasi-experimental design is appropriate for testing the impact of a situation that would otherwise be impossible or unethical to provoke, such as a pandemic exposure (Wing et al., 2018). The difference between trends of the two cohorts is then quantified as a marker of the pandemic's impact.

The impact estimator is the difference in scores between waves 3 and 2 of the COVID cohort, minus the differences in the scores between waves 3 and 2 of the Control cohort. The model's equation for a given child, including control variables can be defined as in equation 1:

$$y_{it} = \beta_0 + \beta_1 C_i + \beta_2 X_{it} + \delta_0 T_{it} + DiD\ T_{it}*C_i + u_{it}\ (1)$$

where i =1, *n* and *t* = 2 ,3; *y* is the outcome variables for each child *i* at wave *t*, C_i is a dummy that indicates if the child belongs to the COVID cohort, T_{it} is a dummy that indicates if the observation *y* of child *i* correspond to wave *t* = 3. The coefficient of interest, DiD, multiplies the interaction term C* T; which is equivalent to a dummy variable that indicates if the observations are from the COVID cohort at wave 3. *X* stands for the control variables that may vary across measurement occasions *t*, and *u* is the

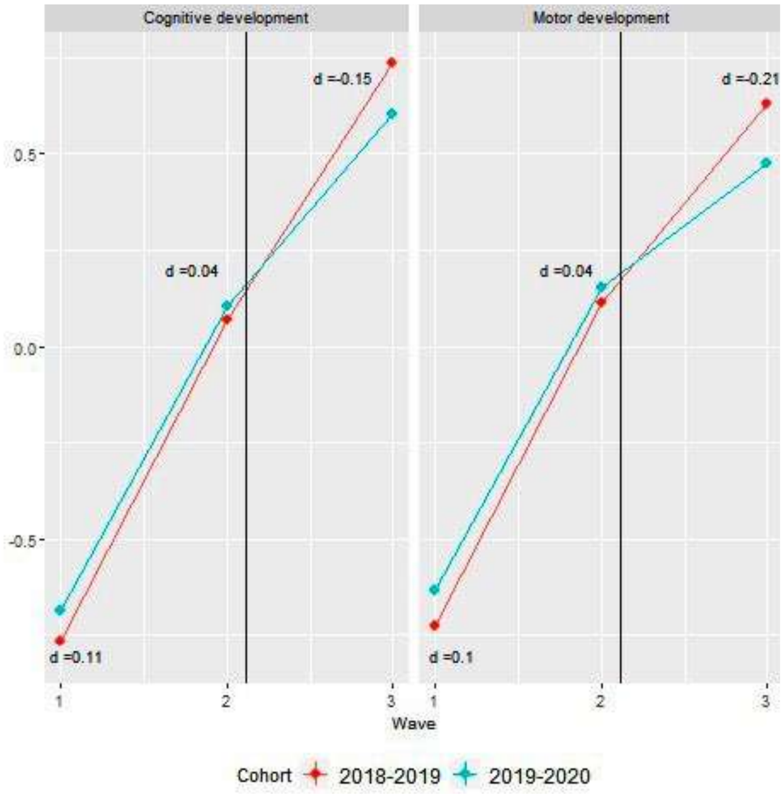


Figure 1. Standardized means for the COVID and Control cohorts (and effect sizes of the differences) at each wave, for cognitive and motor development scores

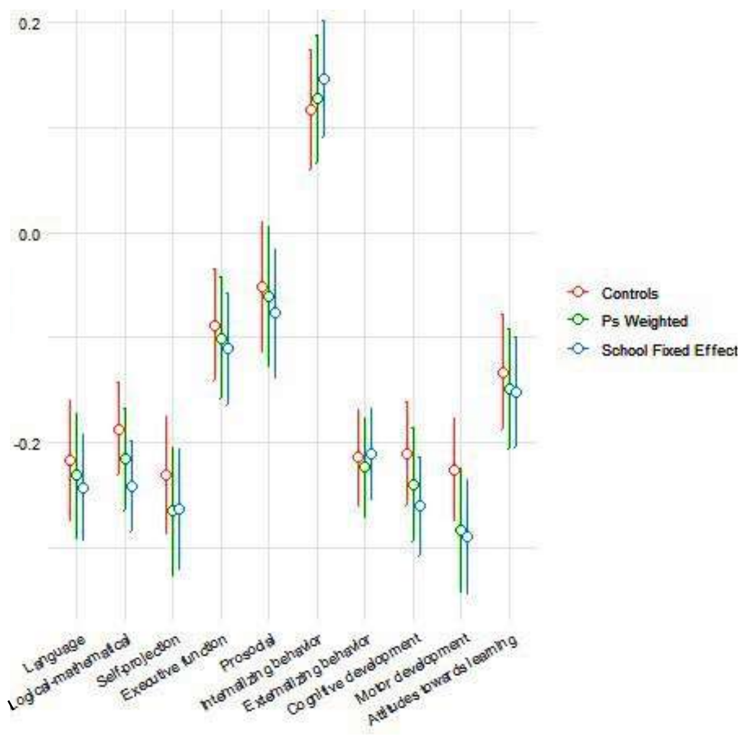


Figure 2. DiD estimates of impact of COVID-19 on INDI domains and subscales. Red dots represent the main specification (estimations with controls). Green dots represent the estimations using propensity matching and the blue dots represent the estimation using fixed effects at the school level. Negative coefficients represent a reduction in developmental performance according to the expected trend, except for Internalizing and Externalizing behaviors where negative scores indicate an improvement compared to the expected performance.

error term. Our model controlled for children's age-in-months, sex, the school SES quintile, and cluster robust standard errors estimate at the group (class) level. In these analyses, we used standardized estimates to convey effect sizes following Ferguson's (2009) criteria (i.e., .2, .5 and .8 as small, moderate, and strong effects, respectively).

In order to reduce possible bias in DiD estimations and check the robustness of results of the main model, we conducted analyses under different specifications: (a) model without controls; (b) balanced model with propensity score matching, to control for the imbalance of cohorts samples by school type, school SES quintile, children's age-in-months and sex; (c) fixed-effect models, to control for the scores' nestedness and non-random variance within schools. For this model, only schools with at least 10 participants were selected.

We examined heterogeneous effects by the school SES quintile, sex, and previous development by using tertiles of development at wave 2. To explore the variability of impact estimates between schools, a mixed-effects model (where impact estimates will be allowed to vary across schools) was conducted.

Additionally, we provided descriptive statistics of the evolution of developmental profiles with intra- and inter-individual comparisons. Development was categorized as *high risk*, *monitoring zone*, or *on-track* based on INDI normative data corrected for age in months. We expected to find a higher proportion of children with high risk or in the monitoring zone at wave 3 in the COVID cohort in comparison to the control cohort.

Results

Preliminary analyses We compared the mean scores for INDI domains and subscales at each of the three measurement waves. Descriptive statistics for each wave and cohort are presented in Table 4. With a visual inspection of the means at waves 1 and 2, we observed a similar trend before measurements in age 5 classrooms. To further check this assumption, we performed a DiD model following equation (1) with t = 1, 2, 3. DiD estimates obtained for the pre-pandemic interaction were all non-significant, except for Executive functioning at the $p < .01$ threshold (= 0.06, p = .002). Internalizing behavior (= -0.04, p = .04) and Attitudes towards learning (= 0.05, p = .02) were significant at the $p > .05$ threshold, with coefficients close to zero.

In order to compare differences between cohorts, we expressed the effect size of differences with Cohen's d (see Figure 1 for cognitive and motor scores, and Supplementary Material 3). The COVID cohort scored slightly higher than the control cohort at waves 1 and 2. At wave 3 (when the COVID cohort was exposed to the pandemic) lower scores were observed in the COVID cohort, except for internalizing behaviors. Furthermore, effect sizes were most often larger at wave 3 than at waves 1 or 2, especially for externalizing behaviors (d = .29) and motor skills (d = .21). Effect sizes in the .10 to .20 range were observed for Language, Logical-mathematical skills, Self-projection, and Cognitive development such that the COVID cohort scored lower than the control cohort.

Differences-in-differences modeling

General estimation We evaluated the impact of the COVID-19 pandemic on childhood development using the DiD approach. Regression analyses were carried out following our main specification (controlling for age, gender, and the school's SES quintile) and without those controls (see Table 5). Analyses showed that the COVID cohort had statistically significant losses, with standardized DiD estimates above .20 in all INDI cognitive subscales: Language, Logical-mathematical skills, and Self-projection, except for Executive functioning (0.10). Concerning dimensions, Motor development and Cognitive development showed losses above 0.20, followed by Attitudes towards learning (0.14). Internalizing behaviors increased (0.13), whereas Externalizing behaviors decreased (0.22), suggesting that the COVID cohort exhibited less aggressive and frustrated behaviors than the control cohort. No significant differences were observed in Prosocial behaviors. These results held even after controlling for age, gender, and the school SES quintile.

To check the robustness of the DiD modelling, we used propensity score matching to control for the imbalance of cohorts in the main demographic and socioeconomic variables, and fixed effects to account for nested data across schools. Balance of treatment and control groups were used as covariates (Supplementary Material 4). We found that estimates were robust to different specifications (see Figure 2 for impact estimates). In order to enhance statistical parsimony, the main specification (model with controls) was selected for further analyses and interpretation. Fixed effects model selection would also require excluding some schools with fewer children from the analysis.

Figure 2. DiD estimates of impact of COVID-19 on INDI domains and subscales. Red dots represent the main specification (estimations with controls). Green dots represent the estimations using propensity matching and the blue dots represent the estimation using fixed effects at the school level. Negative coefficients represent a reduction in developmental performance according to the expected trend, except for Internalizing and Externalizing behaviors where negative scores indicate an improvement compared to the expected performance.

Impact estimates across schools were calculated using a linear mixed model. Figure 3 shows that losses varied substantially across schools in Cognitive and Motor development, and in some schools there were even gains. This pattern is also observed in cognitive and socioemotional subscales and in Attitudes towards learning (see Supplementary Material 5). Schools with gains were distributed across all school SES quintiles. Gains in Cognitive development were more frequent in higher-SES schools, but these differences were not statistically significant.

Next, considering the moderate to high correlation between subscale scores, we selected INDI scores with negative impact estimates above the minimum practical significant effect of .20 to conduct a Multivariate Analysis of Variance (MANOVA) (Supplementary Material 6). Dependent variables included the standardized differences of the scores between waves 3 and 2 for Language, Logical-Mathematical skills, Self-projection, and Motor development. The cohorts were used as independent factors. The MANOVA results revealed significant differences across the two cohorts on all the dependent variables (F (4, 50951) = 279.12, p < .001).

Estimation by school SES, previous developmental level, and sex We conducted analyses to investigate if the school's SES moderated developmental trajectories. Schools were categorized into five quintiles, with higher quintiles indicating the more privileged school districts. In the COVID cohort, Language and Logical-mathematical skills showed losses in the four lowest SES quintiles, whereas very small and non-significant declines were observed in the highest SES districts (quintile 5). Losses in Self-projection, Motor development, and Cognitive development were observed across all quintiles. Though larger impacts were observed among children attending centers from quintiles 2 and 3, followed by quintiles 1 and 4, and lastly quintile 5, few of these differences were statistically significant (see Figure 4). More specifically, significant differences were only observed between quintile 3 and 5 in Language, Logical-mathematical skills, and Cognitive development.

Interestingly, socioeconomic gaps in school readiness increased. For example, the effect size of the differences between Q1 and Q5 in the COVID cohort between T2 and T3 increased from d = 0.33 to d = 0.45 in Language and from d = 0.50 to d = 0.55 in Logical-mathematical subscale, and from d = 0.21 to d = 0.23 in Motor development. Descriptive statistics by quintile can be found in Supplementary Material 7.

Next, considering the moderate to high correlation between subscale scores, we selected INDI scores with negative impact estimates above the minimum practical significant effect of .20 to conduct a Multivariate Analysis of Variance (MANOVA) (Supplementary Material 6). Dependent variables included the standardized differences of the scores between waves 3 and 2 for Language, Logical-Mathematical skills, Self-projection, and Motor development. The cohorts were used as independent factors. The MANOVA results revealed significant differences across the two cohorts on all the dependent variables (F (4, 50951) = 279.12, p < .001).

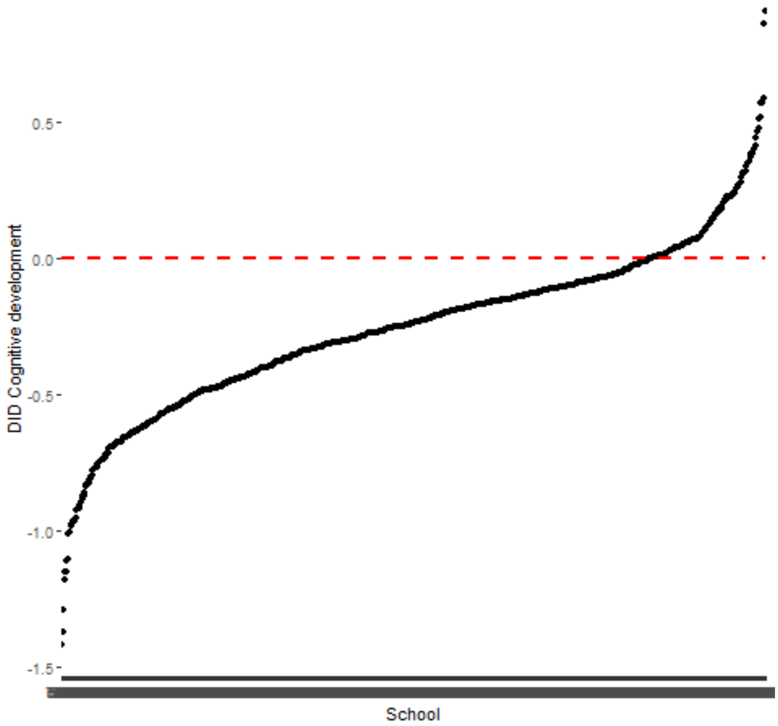
Estimation by school SES, previous developmental level, and sex We conducted analyses to investigate if the school's SES moderated developmental trajectories. Schools were

categorized into five quintiles, with higher quintiles indicating the more privileged school districts. In the COVID cohort, Language and Logical-mathematical skills showed losses in the four lowest SES quintiles, whereas very small and non-significant declines were observed in the highest SES districts (quintile 5). Losses in Self-projection, Motor development, and Cognitive development were observed across all quintiles. Though larger impacts were observed among children attending centers from quintiles 2 and 3, followed by quintiles 1 and 4, and lastly quintile 5, few of these differences were statistically significant (see Figure 4). More specifically, significant differences were only observed between quintile 3 and 5 in Language, Logical-mathematical skills, and Cognitive development. Interestingly, socioeconomic gaps in school readiness increased. For example, the effect size of the differences between Q1 and Q5 in the COVID cohort between T2 and T3 increased from *d* = 0.33 to *d* = 0.45 in Language and from *d* = 0.50 to *d* = 0.55 in Logical-mathematical subscale, and from *d* = 0.21 to *d* = 0.23 in Motor development. Descriptive statistics by quintile can be found in Supplementary Material 7.

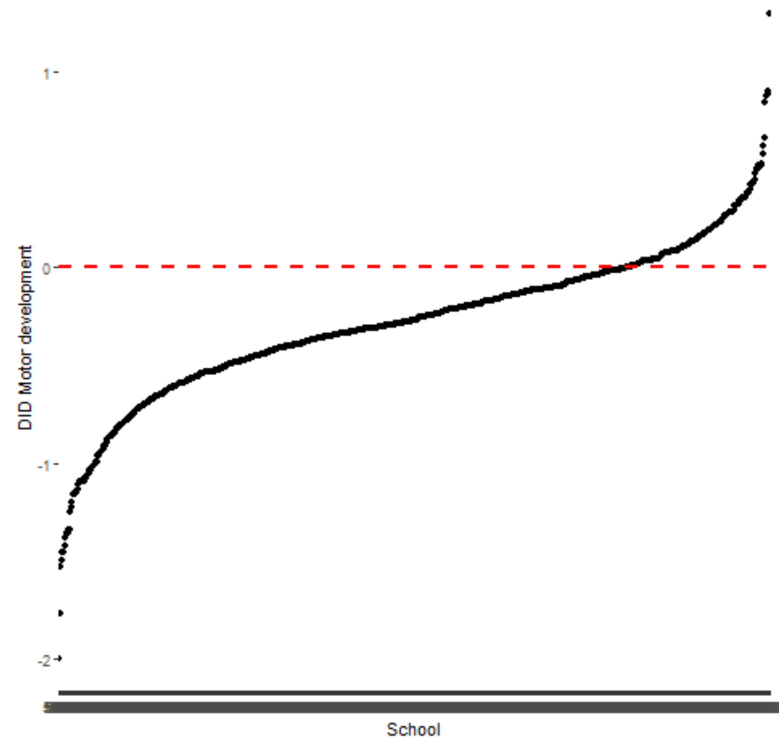
Next, we compared impacts according to previous development (categorized by tertiles of performance at wave 2) and sex (Supplementary Material 8). Children with lower levels of development at age 4 suffered the most from school closures in all dimensions, except for Executive functioning. Nonetheless, statistically significant differences between tertiles were observed only in Logical-mathematical skills and Motor development, whereas all other differences were non-significant. No sex differences were observed except for a statistically significant effect of Externalizing behaviors (boys had more reduction). Descriptive statistics by sex and prior developmental scores can be found in Supplementary Material 7.

Analyses by developmental profiles We further estimated developmental profiles for each INDI score at waves 1, 2, and 3 for the COVID and control cohorts (Supplementary Material 9). When comparing waves 1 and 2, the COVID cohort had a significantly higher percentage of children meeting global developmental expectations (i.e., outstanding and on-track profiles) than the control cohort. At wave 3, the percentage of children who met developmental expectations was significantly lower in the COVID cohort, except for Externalizing behaviors, which was significantly higher.

Next, we carried out intra-individual analyses for children to track their trajectory from wave 2 to 3 (1-year interval). The trajectory for each child was categorized as favorable, unfavorable, or stable for each domain and subscale, based on the normative data used for the INDI reports system. We found that in the COVID cohort the percentage of children with an unfavorable development was significantly higher whereas the



(A)



(B)

Figure 3. Variability of impact estimates at the preschool level for Cognitive (A) and Motor development (B). Negative values (below the red line) represent the preschools where the COVID-cohort showed losses.

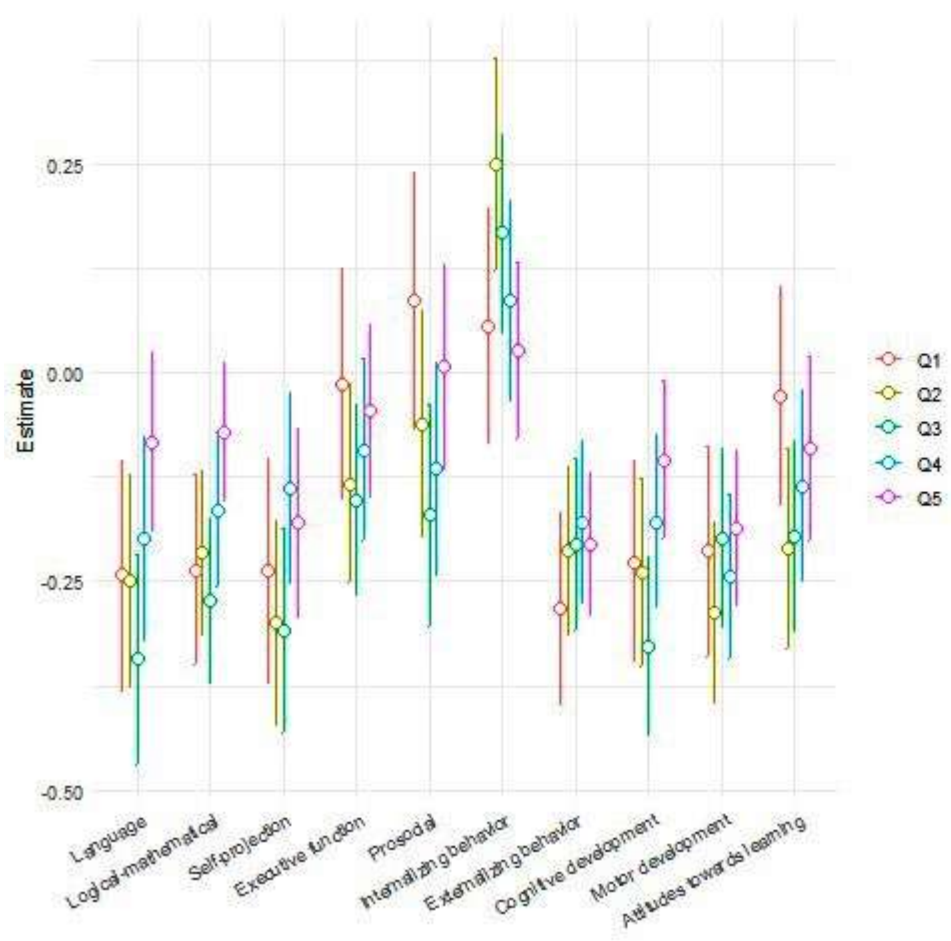


Figure 4. DiD estimates of the impact of COVID-19 on INDI domains and subscales by school quintile. Statistical controls include sex and age-in-months.

percentage of favorable development was significantly lower. This applied to all domains and subscales, except for Externalizing behaviors where we found a statistically significant effect in the opposite direction (see Supplementary Material 10).

Discussion

We investigated the impact of the COVID-19 pandemic on school readiness among preschoolers using two large representative cohorts of Uruguayan children and three waves of data collection. We added novelty to the extant literature in the following ways. Firstly, because we compared a cohort of students who attended preschools through the pandemic to a control cohort who was not exposed to restrictions. Secondly, we used teachers' assessments carried out in the school setting rather than parental reports or online educational platforms. Thirdly, we evaluated cognitive, motor, and socioemotional development simultaneously. Fourthly, we focused on the development of preschool-age children whereas most research to date has focused on other ages or education levels.

Impact of the pandemic on child development

We found that children who were in preschool during the pandemic experienced negative outcomes in multiple developmental domains. This converges with extant literature documenting learning losses in the context of the COVID-19 pandemic among school-age children (Engzell et al., 2021; Kuhfeld et al., 2020; Maldonado & De Witte, 2020), but we were the first to our knowledge to document the impact on school readiness among preschoolers. Though we cannot attribute developmental losses to school closures specifically, our findings also converge with studies documenting detrimental effects on learning of school closures during summer recess (Alexander et al., 2007; Bao et al., 2020; Gershenson, 2013) or emergencies leading to school closures (for a review see Araújo et al., 2020). We found the cognitive and motor development of children was suffering the most, followed by their attitudes towards learning. Within cognitive functioning, Self-projection, Language and Logical-mathematical skills showed the largest losses but effect sizes were small. Nevertheless, these losses could pose a threat to educational achievement later in life. For instance, literature on school readiness highlights that early math and literacy performance are strong predictors of academic outcomes in primary school (Duncan et al., 2007, 2020). Indeed, health and economic crises may have repercussions on the life course years later (Benner & Mistry, 2020). Furthermore, these areas of functioning may be the most difficult to compensate for at home because they require teaching expertise, age-appropriate activities, and materials, and are highly dependent on the quality of stimulation from caregivers (e.g. Anders et al., 2012). Time-structured activities and events were reduced during the pandemic, which could impact the development of self-projection. Similarly, stay-at-home measures led to a drastic decrease in physical activity (Gobbi et al., 2020), which may explain the underdevelopment in motor skills. As there have been few to no studies on the impact of the pandemic on these developmental areas, it would be relevant to develop further research on the topic among preschool-age children.

Most surveys in early childhood have focused on the pandemic's impact on socio-emotional development (Guerrero, 2021). Likewise, we accounted for internalizing, externalizing, and prosocial behaviors. Our study evidenced that the COVID-19 context had a small but significant negative impact on anxious and avoidant internalizing behaviors. These findings converge with other studies carried out among children (Liu et al., 2021). Avoidant and anxious behaviors in children could be an indirect result of increased parental stress, deregulation of daily life (Hiraoka & Tomoda, 2020; Tso et al., 2020), or increased teacher stress (Bacher-Hicks & Goodman, 2021). Unexpectedly, the COVID cohort exhibited less externalizing behaviors when compared to the control group. This might be explained in part by the fact that most often socio-emotional assessments and reports of behavioral problems were based on parental observations during confinement periods (Glynn et al., 2021; Guerrero, 2021; Liu et al., 2021), rather than on teacher evaluations within school settings. Alternatively, the school reopening conditions in Uruguay could explain the lower prevalence of externalizing behaviors because they implied a lower child-teacher ratio and increased supervision of social interactions (NICHD Early Child Care Research Network, 2004; van Verseveld et al., 2019). Similarly, a cross-sectional study carried out among Canadian children and adolescents found that youth reported lower levels of bullying and aggression during COVID-19 in comparison to pre-pandemic levels (Vaillancourt et al., 2021).

We found that the impact of the pandemic on school readiness losses varied considerably across preschool centers. This variation converges with literature on learning losses during summer recesses (Atteberry & McEachin, 2021) and school closures due to the COVID-19 pandemic (Engzell et al., 2021). In the current context, aspects of online learning during closures (e.g. frequency of videoconferencing, parental adherence to distance learning, and availability and quality of bandwidth), and of schools reopening (e.g. space constraints to comply with social distancing protocols, safety-protocols due to possible or confirmed COVID-19 cases) differed across schools to some extent. Also, the quality of developmental stimulation at home could vary across families and school neighborhoods. Further research would be necessary in order to pinpoint the factors within each school that could explain why there was such variation. Another interesting and unexplored line of research could focus on children who benefited from the pandemic. For example, the COVID-19 context may have provided a higher-quality developmental environment for some children because it could increase the quality and frequency of interaction between family members, improve the socioemotional climate and provide a closer follow-up of educational activities, especially among higher-SES families. However, our study showed little evidence in favor of a SES-based explanation of gains.

Impact of the pandemic across socioeconomic status and by previous development

After observing differences in trajectories across the two cohorts, we took an interest in the impact of schools' SES and found significant differences. Specifically, we found that developmental losses for children in the highest quintile were less pronounced, especially in language and math skills. Likewise, previous studies revealed higher learning losses among primary school-age children with lower SES in the context of COVID-19 (Engzell et al., 2021; Maldonado & Witte, 2020). The effect of lower SES on developmental losses is well documented and there is an array of explanations of why underprivileged children would be more negatively impacted by the pandemic and school closures. Firstly, routine school closures during summer recess have more of an impact on children of lower-SES (Alexander et al., 2007), suggesting that the effect is impacted by school closures specifically. In the context of the pandemic, a high-quality home environment or family stimulation (e.g., literacy practices at home, age-appropriate toys and materials) may absorb some of the negative effects among children from privileged families (Bao et al., 2020; Anders et al., 2012). In contrast, low-income families would adhere less to online education because they lack equipment or motivation (Kruszewska et al., 2020). The pandemic has put an additional financial strain on families, especially among those with lower SES (Bailey et al., 2021; Clark et al., 2020). For the most disadvantaged, food insecurity is a major threat to childhood development in Latin America (Guerrero, 2021) and on an international scale (Sharma et al., 2020). A higher risk of psychosocial problems among children of low-income families is well documented (e.g., Tso et al., 2020) and could be exacerbated in the context of the pandemic.

Relatedly, children who were struggling the most when in age 4 classrooms displayed larger developmental losses thus increasing the achievement gap. A similar finding was reported in previous studies (Engzell et al., 2021; Kuhfeld et al., 2021). Prior to the pandemic, the home environment of these children could have lacked appropriate stimulation, and attending school would be particularly critical to compensate. Furthermore, once the pandemic was underway, their families may have been among those who suffered the most (e.g., food insecurity, economic hardships, parental stress).

Strengths, limitations, and future directions

Major strengths of our study included the novelty of studying young children, the large sample size, the presence of a control group not exposed to COVID-19, and our use of teacher reports even in this pandemic context. However, multiple limitations should be noted. Firstly, we cannot be certain of what aspects of the COVID-19 pandemic specifically impacted school readiness (e.g., school closures, parental and teacher stress, access to computers, etc.). We accounted for the school SES, but we were not able to measure family SES or income, nor account for the economic and psychosocial hardships of families during the pandemic (e.g., unemployment, stress levels). There was more attrition in the COVID cohort than in the control cohort that may bias (underestimate) the impact of the pandemic on school readiness, as more vulnerable children were more likely to have dropped out. We did not account for the family's adherence to distance learning, differences in teaching strategies, the number of hours of online instruction, or the number of days children attended school after reopenings. If available, these variables would be worth considering in the future. Teachers' assessments may have been biased because of work stress and shifts in their perception of expected behaviors due to the pandemic. The interpretation of results on internalizing behaviors, attitudes towards learning, and especially executive functioning scores should be taken cautiously because the significant effect of the pre-treatment interaction may signal an issue with the assumption of parallel trends. The current study was carried out using large, national cohort data from Uruguay, but may not generalize to other contexts, as characteristics of school closures, mitigation measures, and other COVID-19 related socioeconomic impacts differed across regions. As this is the first study on preschoolers to our knowledge, it would be beneficial to conduct similar research around the world.

Conclusions and implications

Based on our findings, we can draw the general conclusion that the COVID-19 pandemic harmed multiple domains of childhood development, especially within cognitive and motor domains. Inequalities in school readiness already present in the Uruguayan population (Vásquez-Echeverría et al., 2021) appear to have been exacerbated, as children living in the most privileged districts were more protected from the impact of the COVID crisis, and children who were already struggling at age 4 before the onset of the pandemic suffered the most compared to their expected trajectory.

This increased inequity in developmental losses can directly inform public policy to focus on helping children in their transition to primary school. For example, as motor and cognitive development were the most impacted, it would be important for school teachers to focus on these domains in the upcoming school year. Our findings imply that teachers and caregivers need to focus on children from underprivileged backgrounds or who were struggling before the onset of the pandemic. This advice was given to teachers in Uruguay during the latest 2021 INDI teacher training program, in order to help mitigate the negative impact of the pandemic on school readiness. Public policies may also benefit from allocating resources to families in underprivileged school districts (e.g., financial aid for childcare or private tutors to increase instructional time).

Further research on how the pandemic impacts the cognitive and motor development of young children needs to be conducted because our study was the first to our knowledge and generalization may be dubious. Exposure to the COVID-19 pandemic incorporates a myriad of confluent factors that differ considerably across time, cultures, and geographical locations. Therefore, it would be critical to conduct further research in order to ascertain the extent to which findings generalize outside of Uruguay. Considering the epidemiological situation of Uruguay during 2020, the mild restrictions, and other favorable educational policies (e.g., shorter school closures compared to many other countries, social policies, and a strong culture of technology in education), it would not be surprising if the impact of the pandemic on early childhood development in other countries was larger, and even more exacerbated among underprivileged children.

Funding

This work was partially funded by the technical agreement of the National Administration of Education and the Universidad de la República and a post-doctoral fellowship provided by the National Agency for Research and Innovation to Tianna Loose.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting Consequences of the Summer Learning Gap. *American Sociological Review*, *72*(2), 167–180. https://doi.org/10.1177/000312240707200202

Almeida, D. M., & Wong, J. D. (2009). Life transitions and daily stress processes. In G. H. Elder & J. Z. Giele (Eds.), *The craft of life course research* (141–162). Guilford.

Ananat, E., & Gassman-Pines, A. (2020). Snapshot of the COVID crisis impact on working families. Medford, MA: Econofact, The Fletcher School, Tufts University. https://econofact.org/snapshot-of-the-covid-crisis-impact-on-working-families

Anders, Y., Rossbach, H. G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., & Von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly*, *27*, 231-244. https://doi.org/10.1016/j.ecresq.2011.08.003

ANEP (2021). Asistencia, participación virtual y resultados educativos en la escuela pública uruguaya en 2020. https://www.anep.edu.uy/monitorRepo/Presentaciones%20PDF%20y%20documentos/Asistencia%20y%20participacion%20virtual%20y%20resultados%20a%c3%b1o%20covid_v3.pdf

ANEP-CFE (2016). Proyecto de Plan de Estudios: Maestro/a de Primera Infancia - Asistente/a Técnico/a de Primera Infancia. http://www.cfe.edu.uy/images/stories/pdfs/planes_programas/mtro_primer_infancia/plan_estudios_MPI.pdf

ANEP-DGEIP (2021). Estado de Situación 2020 Monitor de Educación Inicial y Primaria. https://www.anep.edu.uy/monitorRepo/Documentos%202020/ESTADO%20DE%20SITUACION%202020%2027_4%20final.pdf

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (1st edition). Princeton University Press.

Araújo, L. A. de, Veloso, C. F., Souza, M. de C., Azevedo, J. M. C. de, & Tarro, G. (2020). The potential impact of the COVID-19 pandemic on child growth and development: A systematic review. *Jornal De Pediatria*. https://doi.org/10.1016/j.jped.2020.08.008

Atteberry, A., & McEachin, A. (2021). School's Out: The Role of Summers in Understanding Achievement Disparities. *American Educational Research Journal*, *58*, 239–282. https://doi.org/10.3102/0002831220937285

Bacher-Hicks, A. & Goodman, J. (2021). The Covid-19 Pandemic Is a Lousy Natural Experiment for Studying the Effects of Online Learning. *Education Next*, Summer 2021, 21. https://www.educationnext.org/covid-19-pandemic-lousy-natural-experiment-for-studying-the-effects-online-learning/

Bailey, D. H., Duncan, G. J., Murnane, R. J., Au-Yeung, N. (2021). Achievement Gaps in the Wake of COVID-19. *Educational Researcher*, *50*, 266-275. https://doi.org/10.3102/0013189X211011237

Bao, X., Qu, H., Zhang, R., & Hogan, T. P. (2020). Modeling Reading Ability Gain in Kindergarten Children during COVID-19 School Closures. *International Journal of Environmental Research and Public Health*, *17*. https://doi.org/10.3390/ijerph17176371

BCU (2021). Informe de Cuentas Nacionales. Cuarto trimestre y año 2020. Retrieved from: https://www.cncs.com.uy/files/Informe%20459%20-%20Cuentas%20Nacionales%20Cuarto%20Trimestre%20y%20a%C3%B1o%20%202020.pdf

Benner, A. D., & Mistry, R. S. (2020). Child Development During the COVID-19 Pandemic Through a Life Course Theory Lens. *Child Development Perspectives*, 14(4), 236–243. https://doi.org/10.1111/cdep.12387

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Cabella, W., Nathan, M., & Tenenbaum, M. (2013). La población afro-uruguaya en el Censo 2011. Retrieved from: https://www.ine.gub.uy/documents/10181/34017/Atlas_fasciculo_2_Afrouruguayos.pdf/ec7ecb3f-ca0a-4071-b05f-28fdc20c94e2

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & Team, T. O. I. (2020). The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data (Working Paper No. 27431; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w27431

Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open*, *2*, 2332858416649593.

Clark, H., Coll-Seck, A. M., Banerjee, A., Peterson, S., Dalglish, S. L., Ameratunga, S., Balabanova, D., Bhan, M. K., Bhutta, Z. A., Borrazzo, J., Claeson, M., Doherty, T., El-Jardali, F., George, A. S., Gichaga, A., Gram, L., Hipgrave, D. B., Kwamie, A., Meng, Q., … Costello, A. (2020). A future for the world's children? A WHO-UNICEF-Lancet Commission. *Lancet*, *395*(10224), 605–658. https://doi.org/10.1016/S0140-6736(19)32540-1

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duncan, R. J., Duncan, G. J., Stanley, L., Aguilar, E., & Halfon, N. (2020). The kindergarten Early Development Instrument predicts third grade academic proficiency. *Early Childhood Research Quarterly*, *53*, 287-300. https://doi.org/10.1016/j.ecresq.2020.05.009

Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538. http://dx.doi.org/10.1037/a0015808

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118,* e2022376118. https://doi.org/10.1073/pnas.2022376118

Gershenson, S. (2013). Do Summer Time-Use Gaps Vary by Socioeconomic Status? *American Educational Research Journal*, *50*, 1219–1248. https://www.jstor.org/stable/24546671

Glynn, L. M., Davis, E. P., Luby, J. L., Baram, T. Z., & Sandman, C. A. (2021). A predictable home environment may protect child mental health during the COVID-19 pandemic. *Neurobiology of Stress*, *14*, 100291. https://doi.org/10.1016/j.ynstr.2020.100291

Gobbi, E., Maltagliati, S., Sarrazin, P., di Fronso, S., Colangelo, A., Cheval, B., Escriva-Boulley, G., Tessier, D., Demirhan, G., Erturan, G., Yüksel, Y., Papaioannou, A., Bertollo, M., & Carraro, A. (2020). Promoting Physical Activity during School Closures Imposed by the First Wave of the COVID-19 Pandemic: Physical Education Teachers' Behaviors in France, Italy and Turkey. *International Journal of Environmental Research and Public Health*, *17*. https://doi.org/10.3390/ijerph17249431

Guerrero, G. (2021). *Midiendo el impacto de la COVID-19 en los niños y niñas menores de seis años en América Latina y el Caribe*. UNICEF. https://www.unicef.org/lac/media/20556/file/Midiendo%20el%20impacto%20de%20la%20COVID-19%20en%20los%20ninos%20y%20ninas%20menores%20de%20seis%20anos%20en%20America%20Latina%20y%20el%20Caribe.pdf

Hiraoka, D., & Tomoda, A. (2020). Relationship between parenting stress and school closures due to the COVID-19 pandemic. *Psychiatry and Clinical Neurosciences*, *74*, 497–498. https://doi.org/10.1111/pcn.13088

Ichino, A., & Winter−Ebmer, R. (2004). The Long−Run Educational Cost of World War II. *Journal of Labor Economics*, *22*, 57–87. https://doi.org/10.1086/380403

INEEd. (2019). Informe Sobre el Estado de la Educación en Uruguay 2017-2018. Retrieved from: https://www.ineed.edu.uy/images/ieeuy/2017-2018/pdf/Informe-sobre-el-estado-de-la-educacion-en-Uruguay-2017-2018.pdf

Kuhfeld, M., Tarasawa, B., Johnson, A., Ruzek, E., & Lewis, K. (2020). Learning during COVID-19: Initial findings on students' reading and math achievement and growth. *NWEA Brief, Portland, OR*.

Kruszewska, A., Nazaruk, S., & Szewczyk, K. (2020). Polish teachers of early education in the face of distance learning during the COVID-19 pandemic – the difficulties experienced and suggestions for the future. *Education 3-13*, https://doi.org/10.1080/03004279.2020.1849346

Liu, Q., Zhou, Y., Xie, X., Xue, Q., Zhu, K., Wan, Z., Wu, H., Zhang, J. & Song, R. (2021). The prevalence of behavioral problems among school-aged children in home quarantine during the COVID-19 pandemic in China. *Journal of Affective Disorders, 279*, 412-416. https://doi.org/10.1016/j.jad.2020.10.008

Loades, M. E., Chatburn, E., Higson-Sweeney, N., Reynolds, S., Shafran, R., Brigden, A., Linney, C., McManus, M. N., Borwick, C., & Crawley, E. (2020). Rapid Systematic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*, 1218-1239.e3. https://doi.org/10.1016/j.jaac.2020.05.009

Maldonado, J. E., & De Witte, K. (2020). The effect of school closures on standardised student test outcomes. *KU Leuven Department of Economics Discussion Paper DPS20*, *17*. http://dx.doi.org/10.1002/berj.3754

Masonbrink, A. R., & Hurley, E. (2020). Advocating for Children During the COVID-19 School Closures. *Pediatrics*, *146*(3). https://doi.org/10.1542/peds.2020-1440

NICHD Early Child Care Research Network (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes?. *Developmental Psychology, 40*, 651-664.

One Laptop Per Child. (n.d.). Retrieved March 21, 2021. https://www.onelaptopperchild.org/

Papay, J., Mantil, A., Murnane, R. , An, L., Donohue, K., & McDonough, A. (2020). Lifting All Boats? Accomplishments and Challenges from 20 Years of Education Reform in Massachusetts. Brown University. Retrieved December 29, 2020 from: https://annenberg.brown.edu/edopportunity/liftingallboats.

Presidencia de la República (2021). *Medidas del Gobierno para atender la emergencia sanitaria por coronavirus (COVID-19) en materia de Educación***.** https://www.gub.uy/presidencia/politicas-y-gestion/medidas-del-gobierno-para-atender-emergencia-sanitaria-coronavirus-covid-19-0

Rothstein, R. (2020). The coronavirus will explode achievement gaps in education. https://www.epi.org/blog/the-coronavirus-will-explode-achievement-gaps-in-education

Sharma, S. V., Haidar, A., Noyola, J., Tien, J., Rushing, M., Naylor, B. M., Chuang, R.-J., & Markham, C. (2020). Using a rapid assessment methodology to identify and address immediate needs among low-income households with children during COVID-19. *PLOS ONE*, *15*(10), e0240009. https://doi.org/10.1371/journal.pone.0240009

Spiteri, J. (2021). Quality early childhood education for all and the Covid-19 crisis: A viewpoint. *PROSPECTS*. https://doi.org/10.1007/s11125-020-09528-4

Stewart, H., Watson, N., & Campbell, M. (2018). The cost of school holidays for children from low-income families. *Childhood*, *25*, 516-529. https://doi.org/10.1177/0907568218779130

Tso, W., Wong, R., Tung, K. T., Rao, N., Fu, K. W., Yam, J. C., Chua, G. T., Chen, E. Y. H., Lee, T. M. C., Chan, S. K. W., Wong, W. H. S., Xiong, X., Chui, C. S., Li, X., Wong, K., Leung, C., Tsang, S. K. M., Chan, G. C. F., Tam, P. K. H., ... Wong, I. C. (2020). Vulnerability and resilience in children during the COVID-19 pandemic. *European Child & Adolescent Psychiatry*, 1-16. https://doi.org/10.1007/s00787-020-01680-8

UNESCO. (2020). *Education: From disruption to recovery*. UNESCO. https://en.unesco.org/covid19/educationresponse

Unicef (2021). Covid-19 and school closures. One year of education disruption. Unicef. https://data.unicef.org/resources/one-year-of-covid-19-and-school-closures/

United Nations (2021). Quality Education - Sustainable Development Goals. Retrieved from: https://www.un.org/sustainabledevelopment/education/

United Nations. (2020). Policy brief: The impact of COVID-19 on children. Retrieved from https://unsdg.un.org/resources/policy-brief-impact-COVID-19-children

Vaillancourt, T., Brittain, H., Krygsman, A., Farrell, A. H., Landon, S., & Pepler, D. (2021). School bullying before and during COVID−19: Results from a population−based randomized design. *Aggressive Behavior, 47*, 557-569. https://doi.org/10.1002/ab.21986

van Verseveld, M. D., Fukkink, R. G., Fekkes, M., & Oostdam, R. J. (2019). Effects of antibullying programs on teachers' interventions in bullying situations. A meta−analysis. *Psychology in the Schools, 56*, 1522–1539. https://doi.org/10.1002/pits.22283

Vásquez-Echeverría, A. (2020). *El Inventario de Desarrollo Infantil y la evaluación sistemática del desarrollo en contextos educativos. Teoría, creación e implementación.* PsyArXiv. https://doi.org/10.31234/osf.io/xg2hj

Vásquez-Echeverría, A., Tomás, C., González, M., Rodríguez, J. I., Alvarez-Nuñez, L., Liz, M., Pérez, M., Rudnitzky, F. Berón, C., Gariboto, G., López-Boo, F. (2021). Developmental Disparities based on socioeconomic status and sex: An Analysis of Two Large, Population-Based Early Childhood Development Assessments in Uruguay. *Early Child Development and Care*. https://doi:10.1080/03004430.2021.1946528

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*, 352-360. https://doi.org/10.3102/0013189X14553660

Wing, C., Simon, K., & Bello-Gomez, R. A. (2018). Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. *Annual Review of Public Health*, *39*, 453–469. https://doi.org/10.1146/annurev-publhealth-040617-013507

World Health Organization (2021). World Health Statistics. https://www.who.int/data/gho/publications/world-health-statistics

Yoshikawa, H., Wuermli, A. J., Britto, P. R., Dreyer, B., Leckman, J. F., Lye, S. J., Ponguta, L. A., Richter, L. M., & Stein, A. (2020). Effects of the global COVID-19 pandemic on early childhood development: Short- and long-term risks and mitigating program and policy actions. *The Journal of Pediatrics*, 223, 188–193. https://doi.org/10.1016/j.jpeds.2020.05.020

Comparaciones sociales en depresión
Social comparision in depression

Autores: **Alejo Acuña y Sebastián Pereira**. Financiación: Proyecto de Apoyo a la Investigación Estudiantil (PAIE): Estudio de las comparaciones sociales en depresión mediante tareas interactivas. Universidad de la República, Uruguay. Este proyecto busca estudiar cómo las comparaciones sociales se ven afectadas en personas con depresión y/o ansiedad.

Authors: **Alejo Acuña** y **Sebastián Pereira**. Funding: Proyecto de Apoyo a la Investigación Estudiantil (PAIE): Estudio de las comparaciones sociales en depresión mediante tareas interactivas. Universidad de la República, Uruguay. This project seeks to study how social comparisons are affected in people with depression and/or anxiety**.**

Numerical Cognition in Uruguay: from clinics and laboratories to the classroom (Cognición numérica en Uruguay: de la clínica y los laboratorios al aula)

Victor Koleszar[1], Dinorah de León[1], Nadir Díaz-Simón[1], Dahiana Fitipalde[1], Ignacio Cervieri[1] & Alejandro Maiche[1]

1 **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República. Uruguay

Victor Koleszar http://orcid.org/0000-0001-6666-6786 , Nadir Díaz-Simón http://orcid.org/0000-0003-4660-438X, Dinorah de León http://orcid.org/0000-0002-4359-1256, Dahiana Fitipalde http://orcid.org/0000-0003-4651-1108, Alejandro Maiche http://orcid.org/0000-0002-5006-1544

Correspondence concerning this manuscript should be addressed to Alejandro Maiche, Facultad de Psicología, Tristán Narvaja 1674, Montevideo 11200, Uruguay. Email: **amaiche@gmail.com**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Koleszar, V., de León, D., Díaz-Simón, N., Fitipalde, D., Cervieri, I., & Maiche, A. (2020). Numerical Cognition in Uruguay: from clinics and laboratories to the classroom (Cognición numérica en Uruguay: de la clínica y los laboratorios al aula). *Studies in Psychology*, 1-25. Available at **https://doi.org/10.1080/02109395.2020.1749000**

Abstract

Research into numerical cognition has yielded results that could have a positive impact on the field of education through the implementation of the models that focus on the development of mathematical skills. In Uruguay, the emergence of studies into numerical cognition has a neuropsychological and experimental tradition. This article describes the main research carried out over the past 30 years using these two approaches, and it describes their results and possible implications. Finally, the importance of research into numerical cognition, the need to progress research ers' formation and the relevance of continuous interaction with educational institutions to enrich this field of knowledge and foster implementation in the classroom are discussed.

Keywords: Numerical cognition; Education; Uruguay

Introduction

Numerical cognition seeks to understand the processes by which individuals perceive or understand mathematical ideas. Some of the main questions in the field are: What factors explain different individual performances in mathematics? How do children acquire the notion of numbers? These questions are key for the design of educational policies and curricular planning. On this basis, it is understood that there are multiple implications to different educational actors' access to this kind of knowledge since they directly impact on the design of the curriculum for the initial years of schooling, and in turn a system that can identify those children who might present difficulties or that can create opportunities to implement specific teaching strategies for these cases.

The field of numerical cognition research is especially relevant in Uruguay for two reasons. Firstly, according to the 2015 PISA tests[1], the percentage of 15-year-old students in Uruguay who do not reach the minimum threshold defined by PISA for mathematics exceeded 50% (this percentage has remained almost unchanged since 2003). Secondly, when the data are broken down by subjects they show that during the first stages of secondary education, mathematics systematically presents the lowest pass rates (ANEP, 2018). If we add to this the importance of mathematics for education in general (Duncan et al., 2007), we can confirm that Uruguay has a significant problem in regard to the learning of mathematics.

The situation in primary education follows similar trends to those described above. International studies such as TERCE[2] found that at the end of primary school, more than half of Uruguay's students (62%) were performing at the lowest levels (I and II) (INEEd, 2017). The national Aristas test[3] shows similar results in terms of distribution by levels (INEEd, 2018). Despite not specifying minimums of expected accomplish ments, Aristas found that 51% of students in year 3 were performing at the lowest levels (levels 1 and 2 of 5). Upon reaching year 6, the results show some improvements: only 34% were performing at levels 1 and 2. However, there is profound inequality of the results according to Uruguayan students' socio-economic and cultural conditions (INEEd, 2019).

Faced with such challenges, findings arising from the research into numerical cognition carried out in Uruguay for more than two decades should be seriously considered when designing educational policies that introduce significant changes in the teaching of mathematics and especially at initial and primary school levels.

Research into numerical cognition in Uruguay

Studies into numerical cognition in Uruguay are based on two approaches that were developed independently in different Uruguayan academic institutions and that had different objectives.

The first approach includes a variety of studies in neuropsychology led by Prof. Sergio Dansilio and in direct collaboration with the Universidad de la República (UdelaR) Faculty of Medicine, initiatied during the nineties. With a fundamentally clinical perspective, this research team basically focused on the study of dyscalculias in a hospital setting, and followed up with more studies into issues in the educational context at the Universidad Católica del Uruguay (UCU) (Dansilio, 2001, 2008a). Over the years, researchers associated with this research team at the UCU covered a variety of areas in the field of numerical cognition, focusing on learning difficulties linked to calculus.

The second approach emerged recently, during the last decade, from the creation of a specific line of research into numerical cognition at the UdelaR Faculty of Psychology[4]. This research team focused on studying the cognitive bases of mathe matics learning from an experimental perspective, with interventions supported by the use of ICTs in the classroom. This group's study is closely associated with developing interventions based on the stimulation of the Approximate Number System (ANS) as well as the possibility of linking non-symbolic and symbolic content through different individual and group games. The group is led by Prof. Alejandro Maiche and works with international collaborators who are references in this research area, such as Prof. Justin Halberda from Johns Hopkins University or Prof. Elizabeth Spelke from Harvard University.

Contributions from neuropsychology The first studies carried out in Uruguay of what we now recognize to be numerical cognition were developed in the early 1990s by the Institute of Neurology at the UdelaR School of Medicine Clinical Hospital.

Techniques and practices related to difficulties and disorders in the acquisition of mathematical skills were developed based on diagnostic studies and therapeutic treat ment for patients with brain damage or degenerative conditions. The research and dissemination studies carried out by this team addressed the clinical aspects of dyscal culias and explored different possibilities for their diagnosis and treatment. This research team also worked hard to differentiate the specific difficulties of calculation from those specific to language disorders, with the purpose of creating guidelines for teachers and actors in the educational context.

Dansilio et al. published a series of papers that present clinical evidence and case studies (children, adolescents and adults) on the psychopedagogical diagnosis of devel opmental discalculias and Developmental Gerstmann's Syndrome, as well as on the interaction with the various dimensions and representations associated with the acqui sition of numbers. They also raise discussions based on theories of the different historical perspectives contributed by Piaget, Gallistel and Gelman, Butterworth or Dehaene (Dalmás & Dansilio, 2000). By focusing on the neuropsychology of learning, and more specifically on the clinical study of disorders, this group was able to imple ment work experiences with specialized teachers and psychopedagogues, sometimes even in coordination with primary and secondary education authorities. As an example, in 1994 this group carried out a project to standardize a protocol for the study of the Numerical Calculation Processing in Adults from a once-off funding by the CSIC[5] (Commission of Scientific Research) which was not consolidated into a definitive protocol. Currently, this group is driving — from the Institute of Neurology — the design of an evaluation protocol to measure number processing and mathematical abilities in children, with the aim of facilitating access to instruments that screen for numerical difficulties.

In 2004, Prof. Dansilio was made director of the Neuropsychology department at the Universidad Católica del Uruguay (UCU). He promoted the formation of a working group whose purpose was to identify the main signs of specific mathematical learning difficulties during the initial years of schooling, in order to treat and guide these issues early (Balbi & Dansilio, 2010).

This initial study resulted in the creation of a research group that currently revolves around three areas associated with the acquisition of mathematical calculation skills (Rodríguez, Cuadro, & Ruiz, 2019): the need to identify difficulties in arithmetic calculation (Balbi, Ruiz, & García, 2017); evaluation and intervention techniques for calculation (Singer, Cuadro, Costa, & Von Hagen, 2014; Singer & Cuadro, 2014a; Ruiz & Balbi, 2019); and studies into the influence of sociodemographic variables on mathematical skills.

The need to identify difficulties in calculation

Balbi et al. (2017) compare the differences that exist in the number of diagnoses of Difficulties in Calculation (DAC) with the amount of diagnoses of Reading Difficulties (DAR). The central hypothesis of this study suggests that teachers detect reading issues much more frequently than calculation problems. To test this hypothesis, the study analysed three groups of children (168 children in years 3 to 6): the first two groups comprised children at risk of DAC and DAR respectively, and a third group that did not present issues in either of those areas. For each group, the number of teacher reports that referred to learning difficulties during those years prior to the group analysis was analysed. The authors found significant differences between the DAR group reports and the control group, i.e., the teachers generally identified children with reading difficulties. However, there were no significant differences between DAC group and the control group reports. In addition, there were also no significant differences between the DAR and DAC group reports. Several explanations have been suggested to explain these somewhat contradictory results: the authors point out that teachers possess scarce knowledge about calculation, that there is lack of evaluation tools and that calculation difficulties tend to coexists with reading difficulties (Balbi et al., 2017).

Evaluation and intervention techniques in calculation

The development of tools to measure numerical skills and for the diagnosis of specific learning difficulties has been another research focus for this group, whose origins stem from the Universidad Católica in Uruguay. The Efficacy Test in Arithmetic Calculation (TECA) (Singer, Cuadro, Costa, & von Hagen, 2014) evalu ates arithmetic efficiency, in terms of accuracy and speed, through basic numerical combinations (combinations of natural numbers from 1 to 20). The test consists of three scales: 'addition and subtraction' scales, comprising 72 items that can be administered to students in years 1 to 6; and 'multiplication' and 'division' scales, for students in years 3 to 6, comprising 36 items each. Content validity analyses (which included item, dimensionality and reliability analysis) were carried out with a sample of 331 students in years 1 to 6 at primary school and showed evidence of the validity of a one-dimensional structure, consistent with the theoretical approaches (Singer & Cuadro, 2014a). The authors noted that the TECA provides information on the process of acquiring calculation skills throughout the school year in the four basic operations, while also facilitating the detection of possible learning difficulties in calculation (DAC). This test is still sometimes used in Uruguayan educational institutions, but it is likely to be a fairly common reference in the private clinical setting.

In another intervention study carried out by this team (Singer, Ruiz, & Cuadro, 2018), the authors analysed the independent variables of linguistic skills and ANS in the performance of calculation efficiency. 679 students in years 3 to 6 from different schools in Montevideo were evaluated in fluency of arithmetic calculation, phonological pro cessing, vocabulary, working memory and ANS. Using multilevel analysis, they found that all the variables studied contribute partially to the fluency in calculation, although linguistic skills explained a greater proportion of the variance.

In another study recently conducted by the authors of this group, Ruiz and Balbi (2018) aimed to influence the cognitive skills of children aged seven by stimulating mental calculation with a process that was based on a quasi-experimental design; they used an experimental group (N = 25) and an active control group (traditional teaching) (N = 25). The experimental group attended a programme that comprised 15 sessions over the course of five weeks, structured from a series of activities similar to those carried out in the classroom but specifically designed to stimulate the learning of numerical facts[6], numerical system rules and calculation strategies. Both the experi mental group and the active control group were evaluated with various tests (TECA, 2-digit Calculation, Number Line). The results found that there were no significant differences between the groups; however, differences were found between the pre- and post-evaluations that measured mathematical performance. The authors note that there was an improvement in the mathematical performance skills of both groups and the difficulty of comparing with a control group that receives traditional mathematics teaching (Ruiz & Balbi, 2018).

Other studies on sociodemographic variables

Research has also focused on the impact variables — such as socioeconomic status, sex and year — have on mathematical skills. Cuadro, Barg, Navarrete, and Suero (2008) investigated the influence of sociodemographic variables on the development of cogni tive and social skills, specifically in children (27 children aged between seven and 13) who have been homeless (from a very low socioeconomic level). The sociodemographic and neuropsychological development variables of the participants were evaluated, such as oral language, their relationship with the written code, logical-mathematical reason ing skills (measured through the Wisc III Arithmetic and Digit Span Subtest) (Wechsler, 1997) and general cognitive functioning. The main objective of this study was to establish a development profile that would outline the main features of possible development trajectories, considering the current state of development and the poten tials on which socio-educational interventions can be based. As a general result, cognitive profiles and logical-mathematical skills were significantly lower than those expected for the children's age and years of schooling. The greatest potentials, although they also had low performance levels, were detected in arithmetic operations and numerical series (Wisc-III Arithmetic), probably the result of activities carried out on the streets in contexts that involve money management.

The research carried out by this team regarding sociodemographic variables high lighted the need to develop strategies that reduced the negative effect of these variables on mathematical performance. In general, the authors noted that these strategies should involve families as well as educational centres (von Hagen, Cuadro, & Giloca, 2017).

Contributions from an experimental perspective

In 2010, the Universidad de la República (UdelaR) Faculty of Psychology created the Centre for Basic Research into Psychology (CIBPsi) within which the numerical cogni tion research group[7] operates. This group studies the cognitive processes involved in the learning of numbers and mathematics during the initial years of schooling.

Large-scale ANS stimulation

In 2013, the UdelaR numerical cognition team launched a large-scale intervention project in 10 public schools from different socio-cultural levels around Uruguay that included approximately 1,000 children in year 1. This project, which was carried out with the collaboration of the Ceibal Plan[8], was based on a pre-post test design and included a stimulation of numerical skills programme which was carried out using Ceibal Plan games for tablets. The programme involved the evaluation of cognitive skills linked to numerical and mathematical concepts in children during their first year of primary school with the objective of determining whether a short intervention could improve the formal learning of mathematics.

The programme sought to strengthen children's ability to estimate quantities (asso ciated with ANS) through different games that were applied in different sessions over five weeks. General and numerical cognitive skills such as comparison of non-symbolic magnitudes, temporal discrimination, numerical recognition, comparison of symbolic quantities, estimation of quantities and magnitudes, visual search, number line, simple arithmetic, working memory, vocabulary and mathematical anxiety were evaluated.

The working hypothesis was based on the idea that refinement in the accuracy of ANS, through the discrimination of non-symbolic quantities, can generate the appro priate conditions to facilitate the understanding of numbers and this, in turn, can improve the ability to perform arithmetic operations. Valle Lisboa et al. (2017a, 2017b) analysed the results of this study and concluded that it is possible to stimulate the ANS and increase its accuracy through short interventions mediated by games installed on tablets and that this positively influences the mathematical skills of children in year 1. However, this project did not include a control group, so it was not possible to establish the presence of a causal relationship. The authors also indicated that the intervention had differential results according to the sociocultural level of the school. Results were more effective in those children who attended lower sociocultural level schools, although they may have been more effective because these children carried out more activities or because their starting point (on average) was lower.

In another article published by this group (Odic et al., 2016), the authors analysed the relationships between time estimation, quantity estimation and mathematical skills within the framework of Walsh's A Theory of Magnitude (ATOM) (Walsh, 2003). Correlation analysis showed that the approximate representations of quantities and time were different, despite the fact that both correlate with mathematical performance. The authors concluded that ANS skills correlate more strongly than time discrimination with formal mathematical performance; for ANS, this association is maintained when controlled for accuracy in time estimation and working memory. The results allow us to confirm that representations of time and number are different and provide evidence to better understand the relationship of these magnitudes (time and quantity) associated with mathematical performance, at least in children in year 1.

The project executed by this research group in collaboration with the Ceibal Programme served to shed light on the progress made in Uruguay in experimental research in the field of numerical cognition and was the first time that findings in experimental cognitive psychology were implemented into school classrooms. On this basis, the project functioned as a base for various initiatives. Firstly, the Ceibal plan created the Ceibal study centre[9], where it could focus its research and which it could use as a jumping board to develop its true potential and influence the country's education. Secondly, it initiated a process of exchanges between teachers and researchers and direct contact between the academy and educational centres, facilitating teachers' access to the world of empirical research and, specifically, to knowledge about cognitive psychology. A fact that validates this greater access is the growing number of teachers and educators who have participated in training activities and educational workshops, which members of this group have organized since 2017 through CICEA[10] — such as the courses Aportes de las Ciencias Cognitivas a la Educación [Contributions of Cognitive Sciences to Education] (200 participants) or the Simposio de Educación, Cognición y Neurociencias [Symposium on Education, Cognition and Neurosciences] (300 participants) and the Reuniones Académicas de Trabajo [Academic Work Meetings][11] (200 participants).

Similarly, as part of this research project, the team also made progress in the design of a tablet-based Uruguayan Mathematics Test (PUMA) that takes a game format and is aimed at assessing mathematical skills in children who are starting school (González et al., 2016; Valle Lisboa et al., 2017a). It is an evaluation that has the peculiarity of not needing to be read and can be self-administered because the software regulates the interaction with the child through audio instructions. This test evaluates different numerical and mathematical skills that include tasks such as numerical recognition, transcoding, composition and numerical decomposition, ordering of Arabic numerals, positioning on the number line and solving addition problems. A more detailed description of its possibilities and its use can be found in the group's publications (Odic et al., 2016; Valle Lisboa et al., 2017a, 2017b) as well as on their website[12], which details all the phases and tasks used in this research study.

Refining a perspective: experimentation in small groups

In the years following the large-scale ANS stimulation project, the UdelaR numerical cognition team continued its research work exploring the importance of refining the accuracy of ANS for mathematical performance through different smaller interventions in the classroom.

González et al. (2016) carried out an intervention study in two separate year 1 classes (N = 44) based on the use of a specially designed card game. A quasi-experimental cross-over design with two groups and three evaluation periods was used. The evalua tion was composed of different mathematics tests that measured the performance in symbolic and non-symbolic numerical skills (approximate numerical system, calcula tion and mathematical reasoning problems). The non-intervention group received normal classes. The game used for the intervention was coordinated by the teacher of each class and was based on a set of cards with different stimuli on both sides. One side of the card shows a typical ANS stimulus with two sets of points (each a different colour), and the other side shows their symbolic representation (in Arabic numerals). The game is played in groups of four or five children who must place each card in the location that corresponds to the colour that has the most points. Children can take whichever card they like, which means they can make their decision based on the ANS (by looking at the side with the dots on them) or based on symbolic information (looking at the side with Arabic numerals). The game seeks to stimulate the connection between non-symbolic percepts (set of points) and the symbolic content that represents them (Arabic numerals). It was expected that for ratios close to one, where non symbolic information (the dots) is ambiguous, the children would perceive the useful ness of the numerical symbols to make their decision and, in this way, the necessary mapping between the non-symbolic and the required symbolic would be made at this stage of mathematical learning.

The authors reported that the intervention had positive effects on the children's symbolic mathematical performance (results from the Symbolic Arithmetic Test and PUMA) and recommend using these types of games in pedagogical strategies for the formal teaching of mathematics to foster the mathematical performance of children in their initial years of schooling (aged six). Although the results are promising and shed light on the potentials of these types of resources, we must also remember that, as the researchers indicate, these results should not be considered conclusive because of the small sample size.

During 2017 and 2018, another classroom intervention was carried out based on educational materials and educational games developed by this same study group (Langfus et al., 2019). The project sought to study the relationships between the notions of space, time and number with symbolic arithmetic. The objectives of the project also included the need to develop teaching materials[13] that could be used directly by teachers in the classroom. The project included an application called Matemáticas Monstruosas [Monstrous Mathematics] for tablets created by the Ceibal Plan that consisted of four mini stimulation games that worked a variety of skills, and comparisons and estimates of quantities, time and space exercises. The project also comprised an educational activities class book that had the same design and characters as the application. The book's activities were chronologically ordered in increasing difficulty and were designed for years 2 and 3 (approximately ages eight to 10). The activities were designed by a group of 10 teachers who voluntarily participated in the study which was part of the process that sought to bring teachers and researchers closer together. The team of teachers worked with researchers for approximately one year, and had a central role in the process as their feedback for the definition of the games, children's daily school life, and the design of the exercises with increasing difficult levels for the activity book was essential.

The research sample comprised 386 children from Montevideo and Canelones from years 3 and 4. The project included an intervention with the materials developed (experimental group) and a control group, which received classes in traditional format during the intervention period. The results showed that both groups improved in ANS and pre-post arithmetic calculation scores, but no major difference was seen between the two groups. The authors did note an improvement for the intervention group when divided into socioeconomic levels (Quintile Very low (1) and Medium (3)). Likewise, an improvement was observed in those students who had repeated classes (Average Quintile (3))[14], although in these cases the group sizes are too small to confirm a conclusion in this regard.

Ongoing projects: promoting the initial learning of mathematics

In 2018, the research team initiated a line of research into parental involvement in children's learning through daily numeracy activities related to mathematics. Under the hypothesis that these activities enhance the learning of early maths concepts (LeFevre et al., 2009), the research study sought to replicate the relationship between the frequency of numeracy activities at home and mathematical performance reported in the literature from a correlational study (N = 37 children) where this association was verified (De León, Sánchez, Koleszar, Cervieri, & Maiche, in press). Subsequently, an experimental study was carried out with 140 dyads (parent–child) from two public primary education institutions which comprised an intervention based on workshops that involved games and use of numbers using everyday household activities. The analysis of the data and the publication of the results are still pending.

With the aim of fostering the use of Ceibal Plan tablets for experiments, another part of the team has been working on improving mathematical abilities during early childhood and for this developed the CETA (Ceibal Tangible) project, which led to the creation of a tangible interaction device for learning mathematical concepts (Marichal et al., 2017). To test the usefulness of this device, the BRUNO video game for year 1 primary school children was developed, and its impact was evaluated through an experimental design with three conditions: two experimental conditions (tangible interaction, virtual interaction) and a control group. Although the sample was small (24 children), the comparative results of the mathematical performance, evaluated through the TEMA-3 instrument, showed significant improvements for the experimen tal condition of tangible interaction (Pires et al., 2019).

From numerical cognition to the teaching of mathematics

The scientific discoveries described above aim to understand how children incorporate numerical knowledge. Many of them emerge from recent research into cognitive neuroscience by authors such as Stanislas Dehaene or Elizabeth Spelke and construct new ways for school-age children to learn mathematics. Dehaene (1997) details how cognitive functions that support the development of mathematical concepts in children emerge from an early age. Some years later, Spelke (2011) shed light on the existence of an innate knowledge that could be the basis of future learning.

However, the ability to understand and use mathematical concepts to solve problems is not based on a single cognitive ability; thus, we cannot conclude that it is either completely innate or only acquired. Mathematics, a complex knowledge with culturally acquired components, is supported by different general processes such as language, visuospatial processing, memory and attention, which have an ontogenetic basis that can be modified by learning (Carey, 2009; Dehaene, 1997; Gelman & Gallistel, 1978). The study of numerical cognition seeks to understand the cognitive and neural bases of mental representations of quantities and their relationship with mathematical concepts in order to refine the didactic and pedagogical strategies teachers use when teaching mathematics.

Most of the studies mentioned in this article fall within that framework. That is to say, the purpose of many of these studies is to explore the basic mechanisms of numerical cognition with the aim of improving mathematical education and the teaching-learning mathematical processes in school-aged children. On this basis, different programmes for the stimulation of nuclear numerical capacities have been designed and evaluated (Spelke & Kinzler, 2007) with the intention of implementing them into pedagogical strategies and didactic activities that can facilitate, for example, the acquisition of the concept of numbers. Progress is in the early stages in terms of the impact and transfer of these investigations to the classroom and its expansion, but the greater involvement of teachers in these areas suggests that these fields are becoming more developed.

From the teaching of mathematics to numerical cognition

In the field of numerical cognition there is still an ongoing discussion regarding the different models, theories and concepts needed for the acquisition of mathematical knowledge. However, the progress that has been seen in recent years in the cognitive bases of learning and, in particular, in the cognitive bases of learning mathematics allows us to envision future changes in teaching.

Over the last 100 years we have made tremendous progress in our understanding of how humans learn and process information but, as is often the case in the early stages of creating a field of knowledge, we are still debating models and learning theories that, in many cases, are opposed. Constructivist ideas strongly influenced the Uruguayan education system (CEIP, 2013) and, consequently, the mathematics curricula. However, unlike what Piaget (1978) and other authors emphasized, in the field of numerical cognition babies seem to respond intuitively to the numerical properties of their visual world, even before they can use language, abstract reasoning or the possibility of manipulating the environment (Butterworth, 2005).

There are several cognitive scientists and mathematical educators who argue that we are born pre-wired with an innate knowledge that can facilitate our interaction with key elements in the world such as faces, intentions or quantities (Dehaene, 2019). Based on empirical evidence of this basic knowledge (core knowledge, Spelke & Kinzler, 2007), they support teaching strategies that rely on certain intuitions of quantities (shared with other species), which could form the basis of the acquisition of mathematics (Dehaene, 1997). This is a good example of how recent discoveries in cognitive science can impact the design of mathematical curricular programmes. And in this sense, designing a curriculum for primary education that is based on the idea that children start school with no notions of numbers is very different to designing one based on the assumption that there is universal knowledge of numbers from birth (Lipton & Spelke, 2003).

In any case, it is also understandable that the incorporation of recent discoveries emerging from research into mathematical curricular programmes is not immediate.

Shared influence between research and school education methods is dependent on local actors and, for that same reason, the objective should not only focus on the influence of one on the other, but rather on continuous collaboration and the formation of mixed working groups, where those responsible for teaching in the classroom are an inherent part of the research processes that are developed at universities and research centres. In our opinion, this is the best way to ensure the relatively rapid incorporation of scientific discoveries into teaching programmes and the design of public policy in education (Simms, McKeaveney, Sloan, & Gilmore, 2019).

In the particular case of Uruguay, we believe that educational actors' acquisition of knowledge of numerical cognition is still in its initial stages, although it is true that the development of the numerical cognition research field has matured over the last few years of production and showing its intent to influence teaching practices. The results of the research analysis carried out by this article have clearly shown that the development of the different lines of research have had, for the moment, limited influence on the design of public policy or on the construction of curricula. Although we do not have a specific study that shows the causes of this limited influence, we can assume that it is related to the fact that changes in teaching practices, like any human process, require an appropriation process and that cannot be carried out from behind the desks the policies are designed at. Changes in teaching must be made from the conceptual changes that educational actors carry out themselves in their daily practice, and cannot be decreed from the policy design. On this basis, although we believe that it is still too early to see changes in the curriculum, there are signs, such as teachers and educational actors' growing interest and involvement in including the cognitive sciences in postgraduate training, that make us believe that a path towards better results is beginning to be forged.

Notes

1. Programme for International Student Assessment carried out by OECD. 2. UNESCO Third Regional Comparative and Explanatory Study.
3. National Assessment System of Educational Achievements carried out by INEEd https:// www.ineed.edu.uy/nuestro-trabajo/aristas.html.
4. www.cognicionnumerica.psico.edu.uy.
5. https://www.csic.edu.uy/.
6. Numerical facts refer to the internalization and recovery of operations between numbers without the need to calculate (e.g., 6 + 4 = 10) (Temple & Sherwood, 2002). 7. http://www.cognicionnumerica.psico.edu.uy/.
8. Uruguayan One Laptop Per Child initiative.
9. https://fundacionceibal.edu.uy/acerca-de/.
10. https://www.cicea.ei.udelar.edu.uy/actividades3/.
11. Interdisciplinary Centre in Cognition for Teaching and Learning.
12. http://www.cognicionnumerica.psico.edu.uy/proyectos/2013/intervencion/. 13. Available at http://www.cognicionnumerica.psico.edu.uy/#recursos.
14. In Uruguay, the repetition rate for initial years of schooling is very high: in the 2013– 18 period, approximately 12% of children repeated a year.

References

ANEP. (2018). Monitor Educativo Liceal. Dirección de Planeamiento y Evaluación Educativa. CES. Retrieved from https://www.ces.edu.uy/files/2019/Liceos/Presentacin_Monitor_ Educativo_Liceal_2018_A4_.pdf
Balbi, A., & Dansilio, S. (2010). Dificultades de aprendizaje del cálculo: Contribuciones al diagnóstico psicopedagógico. *Ciencias Psicológicas*, *4*, 7–15.
Balbi, A., Ruiz, C., & García, P. (2017). ¿Hay diferencias en la habilidad del docente para identificar dificultades en cálculo y en lectura? R*evista Neuropsicología Latinoamericana, 9*(1), 47–55.
Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*, 3–18.
Carey, S. (2009). Where our number concepts come from. T*he Journal of Philosophy*, *106*, 220.
CEIP. (2013). Programa de Educación Inicial y Primaria 2008. Retrieved from http://www.ceip. edu.uy/documentos/normativa/programaescolar/ProgramaEscolar_14-6.pdf
Cuadro, A., Barg, G., Navarrete, I., & Suero, M. (2008). Evaluación de las competencias cognitivas y sociales de niños que han estado en situación de calle. *Ciencias Psicológicas*, *2*, 143–152.
Dalmás, J. F., & Dansilio, S. (2000). Visuographemic alexia: A new form of a peripheral acquired dyslexia. *Brain and Language*, *75*, 1–16.
Dansilio, S. (2001). *Trastornos de las Facultades Matemáticas: Las acalculias y las discalculias*. Montevideo, Uruguay: Publicación del Departamento de Historia y Filosofía de la Ciencia. Instituto de Filosofía.
Dansilio, S. (2008a). *Los trastornos del cálculo y el procesamiento numérico*. Montevideo: Prensa Médica Latinoamerica.
Dansilio, S. (2008b). Síndrome de Gerstmann del Desarrollo y trastornos en la adquisición de las matemáticas. *Ciencias Psicológicas*, *II*, 55–64.
De León, Sánchez, Koleszar, Cervieri, & Maiche. (in press). Actividades numéricas en el hogar y desempeño matemático: Posibles influencias del sistema numérico aproximado y el nivel socioeconómico. R*evista Argentina Ciencias del Comportamiento.*
Dehaene, S. (1997). *The number sense*. Oxford: Oxford University Press.
Dehaene, S. (2019). *¿Cómo aprendemos?: Los cuatro pilares con los que la educación puede potenciar los talentos de nuestro cerebro*. Buenos Aires: Siglo XXI Editores.
Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.
Gelman, R., & Gallistel, C. (1978). *Young children's understanding of numbers. A window on early cognitive development*. Cambridge, MA: Harvard University Press.
González, M., Kittredge, A., Sánchez, I., Fleischer, B., Spelke, E., & Maiche, A. (2016). CARD GAMES: A way to improve math skills through stimulating ANS. *Neuro Educação*, *8*, 34–36.
INEEd. (2017). Informe sobre el estado de la educación en Uruguay 2015-2016. Retrieved from https://www.ineed.edu.uy/images/pdf/Informe-sobre-el-estado-de-la-educacion-en-Uruguay -2015-2016.pdf
INEEd. (2018). *Aristas 2017. Informe de resultados de tercero y sexto de educación primaria*. Montevideo: Author.
INEEd. (2019). *Informe sobre el estado de la educación en Uruguay 2017-2018*. Montevideo: Author.
Langfus, J., Maiche, A., De Leon, D., Fitipalde, D., Mailhos, A., & Halberda, J. (2019). The effects of SES, grade-repeating, and IQ in a game-based approximate math intervention. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Mathematical cognition and learning: Vol. 5. Cognitive foundations for improving mathematical learning.* Amsterdam: Elsevier Academic Press.
LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/ Revue Canadienne des Sciences du Comportement*, *41*, 55–66.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, *14*, 396–401.
Marichal, S., Rosales, A., Perilli, F. G., Pires, A. C., Bakala, E., Sansone, G., & Blat, J. (2017). CETA. *Proceedings of the 19th international conference on human-computer xInteraction with mobile devices and services - MobileHCI '17*. https://doi:10.1145/3098279.3098536
Odic, D., Valle-Lisboa, J. V., Eisinger, R., Olivera, M. G., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school maath abilities in young children. *Acta Psychologica, 163*, 17–26.
Piaget, J. (1987). *Génesis del número en el niño. Buenos Aires: Guadalupe [V.O.: The child's conception of number.* New York, NY: Norton, 1987].
Pires, A. C., González Perilli, F., Bakała, E., Fleischer, B., Sansone, G., & Marichal, S. (2019). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education*, *4*, 81.
Rodríguez, C., Cuadro, A., & Ruiz, C. (2019). Mathematics learning and its difficulties: The cases of Chile and Uruguay. In *International Handbook of Mathematical Learning difficulties* (pp. 213–230). Cham: Springer.
Ruiz, C., & Balbi, A. (2019). The effects of teaching mental calculation in the development of mathematical abilities. *The Journal of Educational Research*, *112*, 315–326.
Simms, V., McKeaveney, C., Sloan, S., & Gilmore, C. (2019). *Interventions to improve mathe matical achievement in primary school-aged children*. London: Nuffield Foundation.
Singer, V., & Cuadro, A. (2014a). Propiedades psicométricas de una prueba experimental para la evaluación de la eficacia del cálculo aritmético básico [Psychometric properties of an experi mental test for the assessment of basic arithmetic calculation efficiency]. *Estudios de Psicología*, *35*, 183–192.
Singer, V., Cuadro, A., Costa, D., & von Hagen, A. (2014). *Manual Técnico del Test de Eficacia del Cálculo Aritmético.*

Montevideo: Magro Editores.

Singer, V., Ruiz, C., & Cuadro, A. (2018). Las habilidades lingüísticas y el sistema numérico aproximado en la eficacia del cálculo aritmético. Bordón. *Revista de Pedagogía*, *70*, 185–197.

Spelke, E. S. (2011). Core systems and the growth of human knowledge: Natural geometry. In A. M. Battro, S. Dehaene, & W. J. Singer (Eds.), The *Proceedings of the working group on human neuroplasticity and education: Human neuroplasticity and education* (Vol. 117, pp. 73–99). Vatican City: The Pontificial Academy of Sciences.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Temple, C. M., & Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *The Quarterly Journal of Experimental Psychology: Section A*, *55*, 733–752.

Valle Lisboa, J., Cabana, A., Eisinger, R., Mailhos, A., Luzardo, M., Halberda, J., & Maiche, A. (2017a). Cognitive abilities that mediate the effect of SES on elementary symbolic mathematics learning in the Uruguayan tablet based intervention. Prospects. *Comparative Journal of Curriculum, Learning, and Assessment*, *47*, 1–15.

Valle Lisboa, J., Mailhos, A., Eisinger, R., Halberda, J., González, M., Luzardo, M., & Maiche, A. (2017b). Estimulación cognitiva a escala poblacional utilizando tabletas. Del sistema numérico aproximado (ANS) a la matemática simbólica. In *Pensar las TIC desde la ciencia cognitiva y la neurociencia* (pp. 147–172). Barcelona: Gedisa.

Von Hagen, A., Cuadro, A., & Giloca, V. (2017). La construcción de hechos numéricos básicos: Incidencia del sexo, curso y nivel socioeconómico del alumno. *Ciencias Psicológicas*, *11*, 67–76.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488.

Wechsler, D. (1997). *Wisc-III. Test de inteligencia para niños*. Buenos Aires: Paidós.

International Virtual Congress in Sign Language Processing
Primer Congreso Virtual Internacional en Procesamiento de Lengua de Señas

In 2020, the group devoted to the study of cognitive metaphor and sign language processing at the CIBPsi organized, with other national and international institutions, the **International Virtual Congress in Sign Language Processing**. Previous to COVID-19 pandemic, the **IVCSLP 2020** was conceptualized as a venue for international scholars to share their work virtually, while also raising public awareness in Uruguay on the status of sign languages and the many applications of sign language research. Each poster had two QR codes at the down right corner. Each QR gave access to a video with the content of the poster. One video explained the contents in the Uruguayan Sign Language (LSU) and the other in International Sign Language (ISL). In the page, at the top, there is the right-down corner of a poster with the explanation of the QRs and, at the bottom, a video capture of the explanation of the event in LSU by the Uruguayan deaf scholar Fabricio Etcheverry. **https://ivcslp20.wordpress.com/**

En 2020, el grupo dedicado al estudio de la metáfora cognitiva y el procesamiento de la lengua de señas en el CIBPsi organizó, con otras instituciones nacionales e internacionales, la **International Virtual Congress in Sign Language Processing**. Previo a la pandemia del COVID-19, el IVCSLP 2020 fue conceptualizado como un espacio para que académicos internacionales compartieran sus trabajos de manera virtual, a la vez que se sensibilizaba al público uruguayo sobre la situación de las lenguas de señas y las múltiples aplicaciones de la investigación en lengua de señas. Cada póster tenía dos códigos QR en la esquina inferior derecha. Cada QR daba acceso a un vídeo con el contenido del póster. Un vídeo explicaba el contenido en lengua de signos uruguaya (LSU) y el otro en lengua de signos internacional (ISL). En la página, en la parte superior, aparece en la esquina inferior derecha un cartel con la explicación de los QR y, en la parte inferior, una captura de vídeo de la explicación del evento en LSU por el académico sordo uruguayo Fabricio Etcheverry. **https://ivcslp20.wordpress.com**

The habitual nature of food purchases at the supermarket: Implications for policy making

Leandro Machín[1], María Rosa Curutchet[2], Vanessa Gugliucci[3], Agustina Vitola[3], Tobias Otterbring[4], Marcela de Alcantara[5], Gastón Ares[1, 6]

María Rosa Curutchethttps://orcid.org/0000-0003-1211-7604, Agustina Vitola https://orcid.org/0000-0003-3723-2748, Tobias Otterbringhttps://orcid.org/0000-0002-0283-8777, Marcela de Alcantara https://orcid.org/0000-0003-0770-5823, Gastón Ares https://orcid.org/0000-0002-0565-8835

1 **Centro de Investigación Básica en Psicología (CIBPsi).** Facultad de Psicología. Universidad de la República, Uruguay
2 Observatorio de Seguridad Alimentaria y Nutricional, Instituto Nacional de Alimentación, Ministerio de Desarrollo Social. Uruguay
3 Espacio Interdisciplinario. Universidad de la República. Uruguay
4 MAPP Centre - Research on Value Creation in the Food Sector. Aarhus University. Fuglesangs Allé 4, 8210. Aarhus V, Denmark
5 PDJ-CNPq/Embrapa Agroindústria de Alimentos. Av. das Américas, 29501, CEP 23.020-470. Rio de Janeiro/RJ, Brazil
6 Instituto Polo Tecnológico de Pando. Facultad de Química. Universidad de la República. Uruguay

Correspondence concerning this manuscript should be addressed to Gastón Ares, Facultad de Química, Instituto Polo Tecnológico de Pando. E-mail: **gares@fq.edu.uy**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Machín L., Curutchet M.R., Gugliucci V., Vitola A., Otterbring T., de Alcantara M., Ares G. (2020). The habitual nature of food purchases at the supermarket: Implications for policy making. *Appetite*, *155*, 104844. Available at **https://doi.org/10.1016/j.appet.2020.104844**

Abstract

Supermarkets have become the most important provider of food products worldwide. However, empirical evidence about how consumers make their food purchase decisions in this environment is still scarce. Therefore, the present field study aimed to: i) explore how people make their in-store food purchases, and ii) identify the information they search for when making those purchases. Consumers (n = 144) were intercepted when entering the facilities of three supermarkets in two Uruguayan cities. They were asked to wear a mobile eye-tracker while they made their purchases as they normally do. The great majority of the consumers bought at least one food product or beverage (92%) and, on average, examined products from 2.8 sections. In total, they investigated 37 categories within 13 self-service sections, corresponding to 26 categories of ultra-processed products. For 67% of the products, the consumers went straight to the product they seemed to be looking for, grabbed it and put it in their shopping basket or cart, without making any comparison among products. A limited information search was observed. On average, consumers spent 22 s examining products within self-service sections and only 6.9 s were elapsed from the moment they grabbed a product until they put it in the shopping cart. These results provide empirical evidence of the habitual nature of supermarket food purchases in a context characterized by wide availability of ultra-processed products. Implications for policy making are discussed.

Keywords: habits; habitual decision-making; repeat purchase products; low-involvement products; eye-tracking; field study; nutrition information; ultra-processed

Introduction

The food system has been recognized as one of the most important drivers of malnutrition in all its forms, the major cause of health loss worldwide (Swinburn et al., 2019). Since the 1980s, the food system has been oriented towards highly processed products, primarily composed of refined starch, free sugars and low-quality fat, which are inexpensive and extremely profitable (Ludwig & Nestle, 2008). In particular, ultra-processed products, defined as "*formulations of ingredients, mostly of exclusive industrial use, that result from a series of industrial processes,*" dominate the food supply in high-income countries and their sales are rapidly growing in middle and low-income countries (Monteiro, Moubarac, Cannon, Ng, & Popkin, 2013; Monteiro et al., 2019; Pan American Health Organization, 2019). A growing body of evidence suggests that consumption of these products is associated with several adverse health outcomes, including hypertension, cardiovascular diseases, increased cancer risk and all-cause mortality (Fiolet et al., 2018; Kim, Hu, & Rebholz, 2019; Mendonça et al., 2016; Rico-Campà et al., 2019; Srour et al., 2019).

One of the key factors that has contributed to the dominance of ultra-processed products in the food system worldwide is the retail sector, which has transformed from small local shops and open markets to supermarkets[1] (Stanton, 2015; Popkin & Reardon, 2018). The emergence of supermarkets in the second half of the twentieth century markedly changed food purchase by having long opening hours and offering access to a vast variety of items within a large number of product categories (Reardon & Berdegué, 2006).

The main profit for supermarkets is related to the provision of processed products, as they are more affordable than fresh produce (Gómez & Ricketts, 2013). Several studies have shown that a large proportion of the products available at supermarkets are ultra-processed with an excessive content of sugar, saturated fat and sodium (Cauchi, Pliakas, & Knai, 2017; Mackay et al., 2019; Vandevijvere, Mackenzie, & Ni Mhurchu, 2007).In this sense, the growth of supermarkets in low and middle-income countries has been identified as one of the underlying factors for the increased consumption of ultra-processed products (Asfaw, 2008; Demmler, Klasen, Nzuma, & Qaim, 2017; Swinburn et al., 2019).

Although supermarkets have become the most important provider of food products in most countries (Drewnowski & Rehm, 2013; Popkin & Reardon, 2018; Stanton, 2015), research on how consumers make their food purchase decisions in this environment is scarce. However, such research is important, given that an in-depth understanding of consumers' shopping behavior in supermarkets can contribute to the design of effective public policies and in-store interventions designed to discourage choices of ultra-processed products.

Food purchase is expected to be highly repetitive, as consumers tend to buy the same products from the same brands across repeated visits to the same supermarket (Wood & Neal, 2009). In fact, research has shown that food purchases tend to be habitual and repeated over time (Carrasco, Labeaga, & Lopez-Salido, 2005; Gardner, de Bruijn, & Lally, 2011). Habits can be regarded as "*learned sequences of acts that have become automatic responses to specific cues"* (Verplanken & Aarts, 1999) and usually arise when a certain behavior is repeatedly used to achieve a specific goal in a particular context (Wood & Neal, 2007). This is typically the case when choosing a food product within a specific category with a certain number of alternatives in a supermarket. A similar behavior has been observed for purchase decisions of laundry detergents (Hoyer, 1984).

One of the main characteristics of habitual behavior is that little information is necessary for decision-making (van't Riet, Sijtsema, Dagevos, & De Bruijn, 2011). When habits involve making choices between a series of alternatives, individuals are expected to minimize their cognitive burden and information acquisition (Verplanken & Aarts, 1999). Thus, when making their purchase decisions in a supermarket, consumers are expected to

[1] The term supermarket is used to describe self-service stores, with more 3-4 cash registers and a surface larger than 350 m² (Reardon & Berdegué, 2002).

not engage in an in-depth information search. Several characteristics of food choice reinforce the idea that: i) they make a large number of choices each time they visit the supermarket; ii) they make similar decisions numerous times; iii) most grocery items are low-involvement products that do not involve relevant risks; and iv) there are a large number of options available within each product category (Hoyer, 1984). Evidence of the limited information search and the use of limited simplified strategies for decision making have already been reported in the food domain. For example, an eye-tracking field study found that although in-store signage exposure to one particular food product significantly increased consumers' visual attention towards the featured product, past product usage was a far more influential variable in predicting consumer choice (Otterbring, Wästlund, Gustafsson, & Shams, 2014). Similarly, Schulte-Mecklenbeck, Sohn, de Bellis, Martin, and Hertwig (2013) found that most consumers did not engage in an in-depth processing of the available information for selecting between two lunch dishes and tended to rely on only one characteristic for making their choices. Supporting this notion, parents of school-aged children reported to use simple cues, such as visuals of fruit and natural foods and health claims, to conclude on the healthfulness of foods targeted at children (Adams, Evans, & Duff, 2015). In addition, the relevance of heuristic processing in food perception and choice has also been shown in other studies under experimental settings (e.g. Skuvisz, 2017; Sütterlin & Siegirst, 2015).

However, apart from a few notable exceptions, information seeking behavior in the context of food purchasing has not received much attention (van't Riet et al., 2011). In particular, field studies that present descriptive details on consumers' food purchases and the information they search for when making their buying decisions are extremely scarce. Therefore, the current work had two main objectives: i) explore how people make their in-store food purchases, and ii) identify the information they search for when making those purchases across a wide variety of product categories. Based on the literature on the relevance of habits in the context of eating behavior (e.g. Cohen & Farley, 2008; van't Riet et al., 2011), the following two key hypotheses were postulated:

H1: Supermarket purchases tend to be habitual and, consequently, consumers quickly select their preferred product within a category.
H2: Consumers do not engage in an in-depth information search when making their food purchase decisions.

The study was conducted in Uruguay, one of the Latin American countries with the highest prevalence of overweight and obesity among all age groups (ANEP, 2019; Ministerio de Salud Pública, 2015). Uruguay has recently approved the compulsory implementation of nutritional warnings to identify packaged products with excessive content of sugar, fat, saturated fat and sodium (Ministerio de Salud Pública, 2018). The data were collected during the 18-months adaptation period granted to the food industry to include the warnings on the packages.

Materials and Methods

Participants The study was conducted in three supermarkets in two cities of Uruguay, Maldonado and Minas, which correspond to the fourth and tenth most populated cities in the country, respectively. The three supermarkets were part of the same chain. Consumers were intercepted by one of the researchers when entering the facilities of the supermarket and were invited to participate. Specifically, they were informed that the study consisted of wearing an eye-tracker while making their purchases as they normally do. Participants were told that the study aimed at understanding how people make their purchases at the supermarket. All participants reported normal or corrected-to-normal vision and full color vision. They all signed an informed consent form before the beginning of the study. The study was conducted during three weekends between October and November 2019. The protocol was approved by the Ethics Committee of the School of Chemistry of Universidad de la República (Protocol No. 101900-001280-19).

Procedure Participants were asked to wear a mobile eye-tracking headset of the brand Pupil Pro (Pupil Labs GmbH, Berlin, Germany), consisting of three built-in cameras. Two small and lightweight eye cameras recorded the left and right eye at 200Hz. A scene camera with a 90-degree horizontal field of view, mounted above participants' eyes, recorded the forward road scene at 120Hz and 480 pixels. All cameras were connected to a cellphone, which was transported by the participant while making his or her purchases at the supermarket. The data were recorded using Pupil Lab application (Pupil Labs GmbH, Berlin, Germany).

First, the eye-tracker was manually calibrated using a bullseye objective. Participants were asked to follow the center of the bullseye with his or her gaze, as it moved through the four corners that formed an imaginary square, situated approximately 1 m from the participant (a distance estimated to be common for standing in front of a shelf). After the calibration process, participants were instructed to make their purchases in the supermarket as they usually do. Figure 1 shows an example of how participants completed their purchases. They were instructed to contact a researcher once they had made all their purchases and before paying, in order not to record the economic transaction (e.g. through debit card codes). After paying, participants answered a series of questions related to the use of nutritional information while making their food purchases and knowledge about the Uruguayan front-of-pack nutrition labelling regulation. Finally, they answered a series of socio-economic questions. After completing the questionnaire, participants were debriefed about the detailed purpose of the study.

Figure 1. Explicative picture of how the study was implemented in the supermarket.

Data analysis

The data from the eye movements and the scene cameras were combined using Pupil Software (Pupil Labs GmbH, Berlin, Germany) into one video file showing the scene camera footage with the gaze data superimposed on the scene camera footage. Figure 2 shows examples of two screenshots of the videos.

Pupil Player v1.19 (Pupil Labs GmbH, Berlin, Germany) was used to analyze the video recordings. For each participant, all the scenes of the video where they examined or bought food products were identified. For each of the identified scenes, the following information was recorded: section, product category and time elapsed examining the products. For the sections requiring assistance from an employee, the time elapsed within the section was registered.

Within each supermarket section, comparison of products was registered and the type of information they read. When a product was grabbed, the following information was registered: product, whether the product was flipped or not, type of information read (if any), and whether the product was bought or not. Descriptive statistics were used to summarize the data.

Food product categories were classified according to the NOVA food classification system into unprocessed products, minimally processed products, processed culinary ingredients, processed foods, ultra-processed foods and culinary preparations (Monteiro et al., 2017).

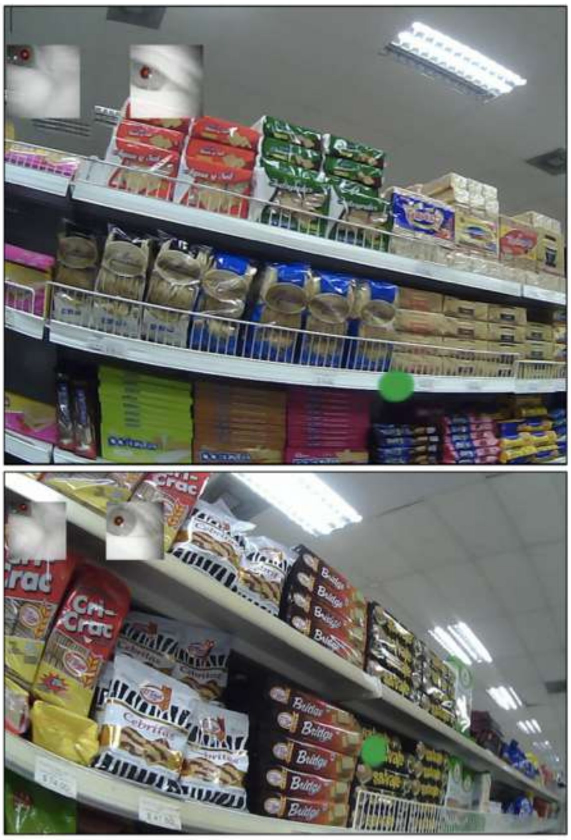


Figure 2. Examples of screenshots of the videos captured using the eye-tracker.

Results

A total of 144 participants provided valid eye-tracking data, whereas recording problems occurred for 19 participants. As shown in Table 1, most participants were frequently in charge of making food purchases for their household. They were mostly female, not highly educated and from medium socio-economic status (Table 1). In terms of age, participants were widely dispersed (18-87 years old).

The great majority of the participants bought at least one food product or beverage (92%, n=134), whereas nine participants did not buy food products (7%) and only one did not buy anything at all (1%). In the following, only participants who bought foods or beverages are considered.

Table 2 shows the percentage of participants who examined products within different sections and product categories, as well as the time they spent doing so. On average, participants examined products from 2.8 sections. Across all participants, a total of 17 sections were visited. *Dairy products*, *Fruits and vegetables*, *Pantry*, *Beverages*, and *Delicatessen,* were the most visited sections. As expected, participants tended to spend more time in the sections that required interaction with an employee to ask for help (e.g. *Bakery*, *Butcher's*, *Delicatessen*) or weight the products (e.g. *Fruits and vegetables*), compared to the self-service sections. For example, participants spent an average of 180s at the *Delicatessen* for purchasing cold cut meats, and 132s at the *Fruits and vegetables* section. For the self-service sections, participants spent between 4s to 51s examining products, with an average of 22 s. The exception was one participant that spent 123s at the *Instant foods* section. In the following, only data for the self-service sections are considered.

Table 1. Characteristics of the participants of the study (n = 144).

Characteristic	Percentage of participants (%)
Gender	
Female	64
Male	36
Age	
18-25	19
25-35	20
36-45	19
46-60	24
Older than 60	18
Educational level	
Primary school	
Secondary school	40
Technical education	32
University	19
Postgraduate studies	7
	1
Socio-economic status	
Low	
Medium	22
High	68
	10
Number of people in the household	
1	
2	15
3	25
4	28
5 or more	19
	13
Responsibility of making food purchases for the household	
Once a week or more	88
1-3 times a week	7
Occasionally	5

Participants examined a total of 37 categories within the 13 self-service sections, corresponding to 26 categories of ultra-processed products, 8 categories of processed foods, 6 categories of culinary ingredients, 4 categories of minimally-processed foods, 2 categories of natural foods, and 1 category including culinary preparations (Table 2). The most frequently visited product category was *Sweetened beverages*: 19% of the participants examined products within this category.

Table 2. Percentage of participants who examined products within different sections and product categories when making their food purchases at the supermarket and average time spent examining products.

Section and Product Category	NOVA classification	Percentage of participants (%)	Average time (s)
Pantry		33	29
Oil	Culinary ingredient	5	17
Rice	Minimally processed	4	46
Sugar	Culinary ingredient	1	4
Instant foods	Ultra-processed foods	1	125
Canned foods	Processed foods	8	34
Flour	Culinary ingredient	5	18
Eggs	Natural foods	1	5
Dried pasta	Processed foods	6	18
Vinegar	Culinary ingredient	1	19
Beverages		33	22
Mineral water	Natural foods	1	15
Flavored waters	Ultra-processed foods	4	29
Alcoholic beverages	N/A	3	32
Juices and nectars	Ultra-processed foods	1	30
Soy milk	Ultra-processed foods	1	19
Powdered drinks	Ultra-processed foods	3	23
Sweetened beverages	Ultra-processed foods	19	18
Coffee, cocoa, tea and mate		7	15
Instant coffee	Processed foods	1	25
Cocoa	Processed foods	1	12
Tea	Minimally processed	1	34
Mate	Processed foods	4	9
*Butcher's**	Natural foods	13	99
Chocolates and sweets		5	35
Alfajor[1]	Ultra-processed foods	3	33
Chocolate	Ultra-processed foods	1	41
Candy	Ultra-processed foods	1	37
Frozen foods		10	29
Frozen foods (e.g. burgers, nuggets)	Ultra-processed foods	7	22
Ice-cream	Ultra-processed foods	2	51
Frozen vegetables	Minimally processed	1	31
Jams and sweet spreads		8	22
Jams	Ultra-processed foods	8	22
*Delicatessen **	Ultra-processed foods	25	180
Refrigerated products		4	24
Pre-made pie crusts	Ultra-processed foods	4	24
Crackers and cookies		15	47
Crackers	Ultra-processed foods	9	51
Cookies	Ultra-processed foods	6	41
Dairy products		43	14
Cream	Culinary ingredient	1	10
Milk	Minimally processed	19	10
Chocolate-flavored milk	Ultra-processed foods	1	3
Butter	Culinary ingredient	3	10

Table 2 (cont). Percentage of participants who examined products within different sections and product categories when making their food purchases at the supermarket and average time spent examining products.

Section and Product Category	NOVA classification	Percentage of participants (%)	Average time (s)
Dairy products		43	14
Cream	Culinary ingredient	1	10
Milk	Minimally processed	19	10
Chocolate-flavored milk	Ultra-processed foods	1	3
Butter	Culinary ingredient	3	10
Milk desserts	Ultra-processed foods	3	14
Cheese	Processed foods	4	16
Cream cheese	Ultra-processed foods	4	26
Yogurt	Ultra-processed foods	9	18
Offers and promotions		4	7
Canned foods	Processed foods	2	4
Crackers	Ultra-processed foods	1	15
Dried pasta	Ultra-processed foods	1	6
*Bakery**	Processed foods	10	53
Packaged breads		9	14
Packaged cakes	Ultra-processed foods	1	16
Toast bread	Ultra-processed foods	7	14
Prepared foods	Culinary preparations	2	16
Sauces and condiments		13	15
Dressings	Ultra-processed foods	5	15
Sauces	Ultra-processed foods	8	14
Savoury snacks		7	17
Olives	Processed foods	1	16
Savoury snacks (e.g. chips)	Ultra-processed foods	5	17
*Fruits and vegetables**	Natural foods	37	132

In all the sections where participants examined products, they bought at least one of them. For 67% of the products, participants went straight to the product they seemed to be looking for, grabbed it and put it in their shopping basket or cart. Conversely, for only 33% of their purchases, participants compared different options within the category or read specific information about a product. As shown in Figure 3, brand and price was the most relevant information for decision making. Meanwhile, for only 3% of their purchases, participants looked at the expiration date before putting a product in the shopping basket or cart.

When participants grabbed a product, they put it straight in their cart. For only 2% of the purchases, participants flipped the products to read the information included on the back of the pack, which corresponded to the expiration date. The average time elapsed since consumers grabbed a product in their hands until they put it in the tray was 6.9 seconds. As shown in Table 3, the time tended to be smaller for culinary ingredients (4.3 s) compared to processed or ultra-processed products (7.5 s and 7.2 s, respectively).

Discussion

The wide availability of ultra-processed products has been identified as one of the main food system drivers of malnutrition (Swinburn et al., 2019). Results from the present work showed that the majority of consumers bought ultra-processed products. These products have been reported to have a large contribution to people's diet worldwide, and particularly in Latin America (Monteiro et al., 2013; Pan American Health Organization, 2019; Popkin & Reardon, 2018). Between 2000 and 20013, sales of ultra-processed foods increased by 68.4% in Uruguay, the highest growth among Latin American countries (Pan American Health Organization, 2015). In particular, sweetened beverages were the most frequently purchased ultra-processed product. In 2019, consumption of sweetened beverages in Uruguay was 98 l per capita (Da Silva, 2020). Given the increasing body of evidence relating consumption of ultra-processed foods with adverse health outcomes and all-cause mortality (e.g. Fiolet et al., 2018; Kim et al., 2019; Srour et al., 2019;), results from the present work reinforce the urgent need to implement regulatory actions to reduce such consumption.

Eye-tracking data confirmed that most consumers do not compare the products available within a specific category when making their purchase decisions. Instead, they seem to go straight to their usual product without much in-store deliberation, supporting the first hypothesis (**H1**). Thus, food purchases in a supermarket setting seem to be a habitual behavior, as previously reported for several eating behaviors (Cohen & Farley, 2008). In this sense, previous studies using panel data have shown that food purchases tend to be repeated over time (Carrasco, Labeaga, & Lopez-Salido, 2005; Gardner, de Bruijn, & Lally, 2011). In addition, similar results have been reported for other low-involvement products. For example, Hoyer (1984) examined laundry detergent choices in a supermarket and reported that most consumers looked at only one product and did not make any comparison between products before a purchase decision was made. However, a key contribution of the present research is that such former non-food findings were here applied to food products and examined across a multitude of different product categories, thereby generalizing prior results to a much wider variety of repeat-purchase products.

The habitual nature of food purchases led to limited information search, consistent with the second hypothesis (**H2**). Pre-purchase evaluation was almost inexistent for the majority of consumers, as they tended to limit themselves to searching for their usual product to confirm their recurrent purchases. This is in agreement with results reported by Betsch, Haberstroh, Glockner, Haar, and Fiedler (2001). None of the consumers in the present research flipped products to read to the information displayed on the back of pack, including nutrition information. This result fits previous experimental research showing that consumers do not perform an in-depth processing of all the information included on food packages before making their choices (Machín, Curutchet, Giménez, Achemann-Witzel, & Ares, 2019). According to Grunert, Fernández-Celemín, Wills, Storcksdieck genannt Bonsmann & Nureeva (2010), 17% of shoppers across six European countries looked for nutrition information on the labels.

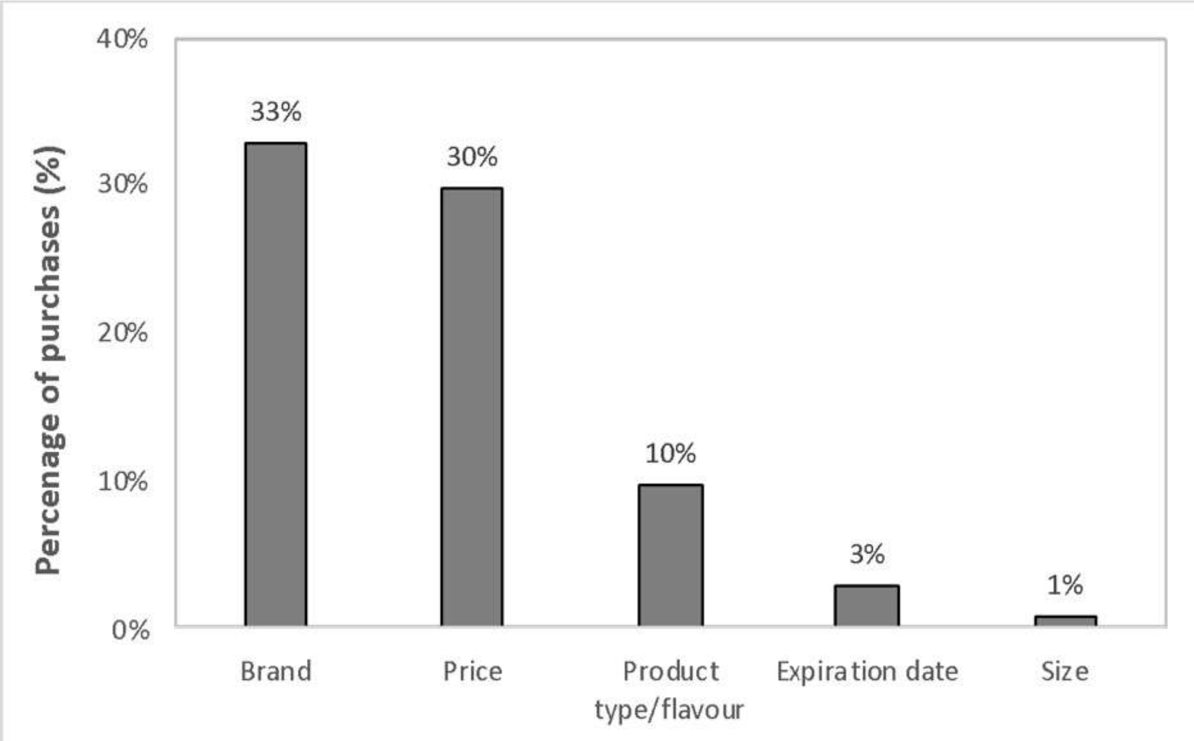


Table 2. Percentage of participants who examined products within different sections and product categories when making their food purchases at the supermarket and average time spent examining products.

However, in the present research none of the participants looked for this type of information.

Participants spent an average of 22 s examining products within self-service sections, confirming limited information search and the use of simplified decision-making strategies. Similar results have been reported for both food and non-food items. Clement (2007) reported that consumers spent an average of 45 s within a category (pasta or jam) to make their choice. Meanwhile, Hoyer (1984) reported that consumers needed an average of 13 s from the time they entered an aisle to the point they had completed their purchase decision of laundry detergents. In addition, the time elapsed since consumers grabbed a product and put it in the cart was only 6.9 s in the present study. This time is in the lower range of the time needed to classify nutrient content in Medium/Low in an experimental task involving food labels presented on a computer screen, which has been shown to vary between approximately 7 to 13 s (Ares et al., 2012).

Among the limited number of consumers who compared different products within a product category and looked for specific information, brand and price had the highest relevance. This is in agreement with results from Machín, Giménez, Vidal, and Ares (2014), who reported that price, quality and brand were the most relevant self-reported factors underlying the choices of Uruguayan consumers in the context of purchasing food. The importance of price fits with well-established economic theory, which postulates that consumer demand is negatively correlated with price for almost all products and services (Mankiw, 2017). This is expected to be particularly relevant in the context of Latin American countries, where food still comprise a relevant percentage of the household budget (Marinov, 2005). Brand has been extensively reported to be a sign of quality and to strongly influence consumers' perception of products (Keller, 1998; Ambler et al., 2002; Varela, Ares, Giménez, & Gámbaro, 2010). In addition, according to Hoyer and Brown (1990), brand awareness has been shown to be one of the most important determinants of product choice.

Conclusions and Implications for policy making

Results from the present work provide empirical evidence of the habitual nature of supermarket food purchases in a context characterized by wide availability of ultra-processed products. This suggests that purchase decisions are expected to be resistant to single interventions, as discussed by Stuben, Chan and Dube (2014). In addition, information provisioning in and of itself cannot be expected to encourage relevant changes in purchase behavior (Verplanken & Wood, 2006). Indeed, the lack of success of information campaigns at achieving enduring behavior change has been reported for several health-related behaviors (Albarracín et al., 2005; Derzon & Lipsey, 2002). In the context of eating habits, campaigns promoting consumption of fruits and vegetables have been shown to have a weak effect (Rekhy & McConchie, 2014). In a similar vein, information provided at the point of purchase requiring large cognitive processing cannot be expected to trigger changes in purchase decisions. This has been the case of conventional nutrition information (i.e. nutrient facts or nutrient declarations) (Balasubramanian & Cole, 2002; Moorman, 1996). Results from this study showed that the average time consumers have a product in their hands is lower than the time needed to classify the content of a single nutrient in medium or high (Ares et al., 2012).

Based on the findings from the current research, policies and interventions aimed at reducing purchase of ultra-products should disrupt habitual decisions at the point of purchase. Habits can be disrupted by changing cues in the context where the decision is made (Verplanken & Wood, 2006; Wood & Neal, 2009). In this sense, policies targeted at introducing salient changes on food packages hold potential to disrupt food purchases and encourage consumers to establish new and more healthful food purchase habits. Two policies could achieve this objective: front-of-pack (FOP) nutrition labelling schemes and restrictions on label design.

Salient FOP nutrition labelling schemes that quickly catch consumers' attention (even if they are not purposely looking for them) and that do not require deep cognitive processing can encourage changes in purchase decisions (Bialkova & van Trijp, 2010, 2011; Grunert & Wills, 2007). Nutritional warnings highlighting ultra-processed products with excessive content of nutrients associated with non-communicable diseases seem to be particularly relevant in the current food environment (Khandpur, Swinburn, & Monteiro, 2018). The inclusion of nutritional warnings may trigger changes in how frequently consumers purchase ultra-processed products. Experimental research has shown that these warnings can change consumers' healthfulness perceptions, particularly for products that are incorrectly assumed to be healthful (Ares et al., 2018; Arrúa et al., 2017).

Changes in package design and particularly the removal of textual and visual elements conveying associations with children, healthfulness and naturalness may also hold potential to disrupt habitual purchases. It may be difficult for consumers to recognize their usual products during their first visit to the supermarket after the implementation of the policy, which may encourage them to re-think their purchases. Food marketing standards have long been recognized as one of the cost-effective policies to promote and tackle the burden of malnutrition worldwide (Mozaffarian, Angell, Lang, & Rivera, 2018). However, restrictions on package design have not been widely adopted worldwide yet. Recent research has shown that the Chilean food law, which introduced nutritional warnings and marketing restrictions, is being efficient at reducing the sales of sugar-sweetened beverages (Taillie, Reyes, Colchero, Popkin & Corvalán, 2020).

Another area that deserves further exploration is the impact of in-store interventions to disrupt habits and nudge consumers towards healthier food products. Recent research suggests that several alternatives hold potential. Promising alternatives include increasing the availability of healthy options in salient places inside the supermarket (Bucher et al., 2016; Gittelsohn et al., 2010), offering healthy samples as a prime for promoting healthier purchases (Tal & Wansink, 2015), and using social influence messages on shopping carts, in-store signs, or communicated by store employees (Kristensson, Wästlund, & Söderlund, 2017; Payne, Niculescu, Just, & Kelly, 2015).

Notably, most consumers in the current study did not perform any in-store information search. However, for those who did compare among products before making their decision, price was one of the most relevant attributes. This suggests that fiscal policies aimed at increasing the price of ultra-processed products can be effective. Evidence of the success of such policies has been reported worldwide, particularly for sweetened beverages (Bridge, Lomazzi, & Bedi, 2020; Fernandez & Raine, 2019).

Product reformulation aimed at reducing the content of nutrients associated with non-communicable diseases may also contribute to improving the healthfulness of diets, even if consumers do not modify their food purchase decisions (Lehmann, Charles, Vlassopoulos, Masset, & Spieldenner, 2017). However, it is important to stress that the negative health effects of ultra-processed products seem to be not only dependent on their content of sodium, sugar and low-quality fat (e.g. Fiolet et al., 2018; Hall et al., 2019).

Finally, the relevance of ultra-processed products on participants' food purchases highlights that enduring changes in the consumption of such products cannot be expected unless major transformations in the food system occur. Comprehensive actions should be taken to create sustainable and health-promoting business models and tackle malnutrition and achieve sustainable development goals worldwide (Swinburn et al., 2019).

Acknowledgements

Financial support was received from Instituto Nacional de Alimentación (Ministerio de Desarrollo Social, Uruguay), Espacio Interdisciplinario (Universidad de la República, Uruguay) and Comisión Sectorial de Investigación Científica (Universidad de la República, Uruguay). The authors are grateful to Adriana Rebon for her collaboration for the implementation of the study.

References

ANEP. (2019). *Evaluación del Programa de Alimentación Escolar y monitoreo del estado nutricional de los niños de escuelas públicas y privadas en Uruguay.* Montevideo: ANEP.

Adams, K.M., Evans, C., Duff, B.R. (2015). Ignorance is bliss. How parents of preschool children make sense of front-of-package visuals and claims on food. *Appetite,* 87, 20-29.

Albarracin, D., Gillette, J.C., Earl, A.N., Glasman, L.R., Durantini, M.R., Ho, M.-H. (2005). A Test of Major Assumptions About Behavior Change: A Comprehensive Look at the Effects of Passive and Active HIV-Prevention Interventions Since the Beginning of the Epidemic. *Psychological Bulletin*, 131, 856–897.

Ambler, T., Bhattacharya, C.B., Edell, J., Keller, K.L., Lemon, K., Mittal, V. (2002). Relating brand and customer perspectives on marketing management. *Journal of Service Research*, 5, 13–25.

Ares, G., Giménez, A., Bruzzone, F., Antúnez, L., Sapolinksi, A., Vidal, L., Maiche, A. (2012). Attentional capture and understanding of nutrition labelling: a study based on response times. *International Journal of Food Sciences and Nutrition*, 63, 679-688.

Ares, G., Varela, F., Machín, L., Antúnez, L., Giménez, A., Curutchet, M.R., et al. (2018). Comparative performance of three interpretative front-of-pack nutrition labelling schemes: Insights for policy making. *Food Quality and Preference*, 68, 215-225.

Arrúa, A., Machín, L., Curutchet, M. R., Martínez, J., Antúnez, A., Alcaire, F., et al. (2017). Warnings as a directive front-of-pack nutrition labelling scheme: Comparison with the Guideline Daily Amount and traffic-light systems. *Public Health Nutrition,* 20, 2308–2317.

Asfaw, A. (2008). Does supermarket purchase affect the dietary practices of households? Some empirical evidence from Guatemala. *Development Policy Review*, 26, 227-243.

Balasubramanian, S.K., Cole, C. (2002). Consumers' search and use of nutrition information: The challenge and promise of the Nutrition Labelling and Education Act. *Journal of Marketing,* 66, 112-127.

Betsch, T., Haberstroh, S., Glockner, A., Haar, T., & Fiedler, K. (2001). The effects of routine strength on adaptation and information search in recurrent decision making. *Organizational Behavior and Human Decision Processes,* 84, 23–53.

Bialkova, S., & van Trijp, H. (2010). What determines consumer attention to nutrition labels? *Food Quality and Preference*, 21, 1042–1051.

Bialkova, S., & van Trijp, H. C. M. (2011). An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Quality and Preference,* 22, 592–601.

Bridge, G., Lomazzi, M., Bedi, R. (2020). Implementation of a sugar-sweetened beverage tax in low- and middle-income countries: recommendations for policymakers. *Journal of Public Health Policy*, 41, 84-97.

Bucher, T., Collins, C., Rollo, M.E., McCaffrey, T.A., De Vlieger, N., Van der Bend, D., et al. (2016). Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. *British Journal of Nutrition*, 115(12), 2252–2263.

Cauchi, D., Pliakas, T., Knai, C. (2017). Food environments in Malta: associations with store size and area-level deprivation. *Food Policy*, 71, 39-47.

Cohen, D.A., Farley, T.A. (2008). Eating as an automatic behavior. *Preventing Chronic Disease*, 5, http://www.cdc.gov/pcd/issues/2008/jan/07_0046.htm.

Carrasco, R., Labeaga, J. M., & Lopez-Salido, J. D. (2005). Consumption and habits. Evidence from panel data. *Economic Journal,* 115, 144–165.

Clement, J. (2007). Visual influence on in-store buying decisions: an eye-track experiment on the visual influence of packaging design. *Journal of Marketing Management*, 23, 917-928.

Da Silva, M. (2020). Cuántos consumieron los uruguayos de cerveza, whisky, agua y refrescos en 2019? Montevideo: El País. Available at: https://negocios.elpais.com.uy/noticias/cuantos-consumieron-uruguayos-cerveza-whisky-agua-refrescos.html

Demmler, K.M., Klasen, S., Nzuma, J.M., Qaim, M. (2017). Supermarket purchase contributes to nutrition-related non-communicable diseases in urban Kenya. *PLoS One,* 12, e0185148.

Derzon, J.H. Lipsey, M.W. (2002). A Meta-Analysis of the Effectiveness of Mass-Communication for Changing Substance-Use Knowledge, Attitudes, and Behavior. In: in W.D. Crano and M. Burgoon (Eds), *Mass Media and Drug Prevention: Classic and Contemporary Theories and Research*, pp. 231-258. Mahwah, NJ: Lawrence Erlbaum Associates.

Drewnowski, A., Rehm, C.D. (2013). Energy intakes of US children and adults by food purchase location and by specific food source. *Nutrition Journal,* 12 (1), 1.

Fernandez, M.A., Raine, K.D. (2019). Insights on the Influence of Sugar Taxes on Obesity Prevention Efforts. *Current Nutrition Reports*, 8, 333-339.

Fiolet, T., Srour, B., Sellem, L., Kesse-Guyot, E., Allès, B., Méjean, C., et al. (2018). Consumption of ultra-processed foods and cancer risk: results from NutriNet-Santé prospective cohort. *BMJ*, 360, k322.

Gardner, B., de Bruijn, G.J., Lally, P. (2011). A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. *Annals of Behavioral Medicine,* 42, 174-187.

Gittelsohn, J., Suratkar, S., Song, H.J., Sacher, S., Rajan, R., Rasooly, I.R., et al. (2010). Process evaluation of Baltimore Healthy Stores: a pilot health intervention program with supermarkets and corner stores in Baltimore City. *Health Promotion Practice,* 11 (5), 723–732.

Gómez, M.I., Ricketts, K.D. (2013). Food value chain transformations in developing countries: selected hypotheses on nutritional implications. *Food Policy,* 42, 139–150.

Grunert, K. G., & Wills, J. M. (2007). A review of European Research on consumer response to nutrition information on food labels. *Journal of Public Health,* 15, 385–399.

Grunert, K.G., Fernández-Celemín, L., Wills, J.M., Storcksdieck genannt Bonsmann, S., Nureeva, L. (2010). Use and understanding of nutrition information on food labels in six European countries. *Journal of Public Health*, 18, 261-277.

Hall, K.D., Ayuketah, A., Brychta, R., Cai, H., Cassimatis, T., Chen, K.Y., et al. (2019). Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake. *Cell Metabolism*, 30, 67-77.

Keller, K. L. (1998). *Strategic brand management: Building, measuring and managing brand equity.* Englewood Cliffs, NJ: Prentice-Hall.

Khandpur, N., Swinburn, B., Monteiro, C.A. (2018). Nutrient based warning labels may help in the pursuit of healthy diets. *Obesity*, 26, 1670-1671.

Kim, H., Hu, E.A., & Rebholz, C.M. (2019). Ultra-processed food intake and mortality in the USA: results from the Third National Health and Nutrition Examination Survey (NHANES III, 1988-1994). *Public Health Nutrition,* 22, 1777-1785.

Kristensson, P., Wästlund, E., & Söderlund, M. (2017). Influencing consumers to choose environment friendly offerings: Evidence from field experiments. *Journal of Business Research*, *76*, 89-97.

Hoyer, W.D. (1984). An examination of consumer decision making for a common repeat purchase product. *Journal of Consumer Research*, 11, 822-829.

Hoyer, W. D., & Brown, S. P. (1990). Effects of brand awareness on choice for a common, repeat-purchase product. *Journal of Consumer Research,* 17(2), 141-148.

Lehmann, U., Charles, V. R., Vlassopoulos, A., Masset, G., & Spieldenner, J. (2017). Nutrient profiling for product reformulation: Public health impact and benefits for the consumer. *Proceedings of the Nutrition Society*, 76, 255–264.

Ludwig, D. S., & Nestle, M. (2008). Can the food industry play a constructive role in the obesity epidemic? *JAMA*, 300, 1808-1811.

Machín, L., Curutchet, M.R., Giménez, A., Aschemann-Witzel, J. & Ares, G. (2019). Do nutritional warnings do their work? Results from a choice experiment involving snack products. *Food Quality and Preference,* 77, 159-165.

Machín, L., Giménez, A., Vidal, L., Ares, G. (2015). Influence of context on motives underlying food choice. *Journal of Sensory Studies,* 29, 313-324.

Mankiw, N.G. (2017). *Principles of microeconomics*. Boston: Cengage Learning.

Marinov, M. (2005). *Marketing in the emerging markets of Latin America.* Hampshire: Palgrave MacMillan.

Mendonça, R.D., Lopes, A.C.S., Pimenta, A.M., Gea, A., Martinez-Gonzalez, M.A., Bes-Rastrollo, M. (2016). Ultra-processed food consumption and the incidence of hypertension in a Mediterranean cohort: the Seguimiento Universidad de Navarra project. *American Journal of Hypertension,* 30(4), 358-366.

Ministerio de Salud Pública. (2015). 2ª Encuesta nacional de factores de riesgo de enfermedades no transmisibles. Montevideo, Uruguay: Ministerio de Salud Pública.

Ministerio de Salud Pública. (2018). *Decreto N°272/18.* Montevideo: Ministerio de Salud Pública.

Monteiro, C.A., Cannon, G., Moubarac, J.C., Levy, R.B., Louzada, M.L.C., Jaime, P.C. (2017). The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutrition*, 21, 5-17.

Monteiro, C. A., Moubarac, J.-C., Cannon, G., Ng, S. W., & Popkin, B. (2013). Ultraprocessed products are becoming dominant in the global food system. *Obesity Reviews,* 14, 21-28.

Mackay, S., Ni Mhurchu, C., Swinburn, B., Eyles, H., Young, L., & Gontijo de Castro, T.(2019). *State of the Food Supply New Zealand 2019.* Auckland, New Zealand: The University of Auckland.

Moorman, C. (1996). A quasi experiment to assess the consumer and informational determinants of nutrition information processing activities: The case of the Nutrition Labelling and Education Act. *Journal of Public Policy & Marketing*, 15, 28-44.

Mozaffarian, D., Angell, S.Y., Lang, T., Rivera, J.A. (2018). Role of government policy in nutrition—barriers to and opportunities for healthier eating. *BMJ*, 361, k2426.

Otterbring, T., Wästlund, E., Gustafsson, A., & Shams, P. (2014). Vision (im) possible? The effects of in-store signage on customers' visual attention. *Journal of Retailing and Consumer Services*, *21*(5), 676-684.

Pan American Health Organization. (2015). *Ultra-processed food and drink products in Latin America: Trends, impact on obesity, policy implications.* Washington, DC: Pan American Health Organization.

Pan American Health Organization. (2019). *Ultra-processed food and drink products in Latin America: Sales, sources, nutrient profiles and policy recommendations*. Washington, DC: Pan American Health Organization.

Payne, C.R., Niculescu, M., Just, D.R., Kelly, M.P. (2015). Shopper marketing nutrition interventions: Social norms on grocery carts increase produce spending without increasing shopper budgets. *Preventive Medicine Reports,* 2, 287–291

Popkin, B.M., Reardon, T. (2018). Obesity and the food system transformation in Latin America. *Obesity Reviews,* 19, 1028-1064.

Reardon, T., Berdegué, J.A. (2002). The rapid rise of supermarkets in Latin America: challenges and opportunities for development. *Development Policy Review,* 20, 371-388.

Reardon, T., Berdegué, J.A. (2006). *The Retail-Led Transformation of Agrifood Systems and its Implications for Development Policies.* Santiago: RIMISP - Latin American Centre for Rural Development.

Rekhy, R., McConchie, R. (2014). Promoting consumption of fruit and vegetables for better health. Have campaigns delivered on the goals? *Appetite*, 79, 113-123.

Rico-Campà, A., Martínez-González, M.A., Alvarez-Alvarez, I., de Deus Mendonça, R., de la Fuete-Arrillaga, C., Gómez-Donoso, C., et al. (2019) Association between consumption of ultra-processed foods and all cause mortality: SUN prospective cohort study. *BMJ,* 365, l1949.

Kim Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, 71, 242–251.

Skubisz, C. (2017). Naturally good: Front-of-package claims as message cues. *Appetite,* 108, 506-511.

Sütterlin, B. & Siegrist, M. (2015). Simply adding the word "fruit" makes sugar healthier: The misleading effect of symbolic information on the perceived healthiness of food. *Appetite,* 95, 252–261.

Srour, B., Fezeu, L.K., Kesse-Guyot, E., Allès, B., Méjean, C., Andrianasolo, R.M., et al. (2019). Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). *BMJ,* 365, l1451

Stanton, R.A. (2015). Food retailers and Obesity. *Current Obesity Reports*, 4, 54-59.

Stuben, J., Chan, D., Dube, L. (2014). Policy insights from the nutritional food market transformation model: the case of obesity prevention. *Annals of the New York Academy of Sciences*, 1331, 57-75.

Swinburn, B.A., Kraak, V.I., Allender, S., Atkins, V.J., Baker, P.I., Bogard, J.R., et al. (2019). The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report. *The Lancet,* 393, 791-846.

Taillie, L.S., Reyes, M., Colchero, M.A., Popkin, B., Corvalán, C. (2020). An evaluation of Chile's Law of Food Labeling and Advertising on sugar-sweetened beverage purchases from 2015 to 2017: A before-and-after study. *PLOS Medicine*, 17, e1003015.

Tal, A., Wansink, B. (2015). An apple a day brings more apples your way: Healthy samples prime healthier choices. *Psychology & Marketing,* 32 (5), 575–584.

Vandevijvere, S., Mackenzie, T., Ni Mhurchu, C. (2007). Indicators of the relative availability of healthy versus unhealthy foods in supermarkets: A validation study. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 53.

van't Riet, J., Sijtsema, S.J., Dagevos, H., De Bruijn, G.J. (2011). The importance of habits in eating behaviour. An overview and recommendations for future research. *Appetite*, 57, 585-596.

Varela, P., Ares, G., Giménez, A., Gámbaro, A. (2010). Influence of brand information on consumers' expectations and liking of powdered drinks in central location tests. *Food Quality and Preference,* 21, 873-880.

Verplanken, B., & Aarts, H. (1999). Habit, attitude and planned behavior. Is habit an empty construct or an interesting case of goal-directed automaticity? *European Review of Social Psychology,* 10, 101–134.

Verplanken, B. & Wood, W. (2006). Interventions to Break and Create Consumer Habits. *Journal of Public Policy and Marketing*, 25, 90-103.

Wood, W., Neal, D.T. (2007). A new look at habits and the habit-goal interface. *Psychological Review,* 114, 843-863.

Wood, W., Neal, D.T. (2009). The habitual consumer. Journal of Consumer Psychology, 19, 79-592

Taza decorativa de Dearrollo cognitivo infantil
Children Cognitive Development's decorative mug

Autores: **Alejandra Carboni, Verónica Nin, Hernán Delgado, Carina Aldecosea, Ñeranei Menéndez, Richard Rodríguez.** Financiación: CSIC - *Programa de Apoyo a la Investigación Estudiantil* (PAIE). Universidad de la República, Uruguay. Estas tazas fueron utilizadas en el marco de los proyectos del grupo de Desarrollo Cognitivo Infantil. Estos trabajos exploraron la manera en la cual las características de los entornos modulan el desarrollo de habilidades cognitivas referidas como funciones ejecutivas frías y calientes durante la infancia, con énfasis en contextos de vulnerabilidad social.

Authors: **Alejandra Carboni, Verónica Nin, Hernán Delgado, Carina Aldecosea, Ñeranei Menéndez, Richard Rodríguez.** Funding: CSIC - *Programa de Apoyo a la Investigación Estudiantil* (PAIE). Universidad de la República, Uruguay. These cups were used in the framework of the projects of the Child Cognitive Development group. These works explored the way in which the characteristics of environments modulate the development of cognitive skills referred to as cold and warm executive functions during childhood, with emphasis on contexts of social vulnerability.

Partial agreement between task and BRIEF-P-based EF measures depends on school socioeconomic status

Verónica Nin[1], Hernán Delgado[1], Graciela Muniz-Terrera[2], Alejandra Carboni[1]

1 **Centro de Investigación Básica en Psicología (CIBPsi)**, Facultad de Psicología y Centro Interdisciplinario para la Cognición en la Enseñanza y el Aprendizaje, Universidad de la República, Uruguay
2 Centre for Dementia Prevention, University of Edinburgh, Scotland, UK

Verónica Nin https://orcid.org/0000-0003-3162-1582, Hernán Delgado https://orcid.org/0000-0002-5738-1516, Graciela Muniz-Terra https://orcid.org/0000-0002-4516-0337, Alejandra Carboni https://orcid.org/0000-0003-4009-1122

Correspondence concerning this manuscript should be addressed to Verónica Nin, Centro de Investigación Básica en Psicología, Facultad de Psicología, Universidad de la República, Uruguay. Email: **veronica.nin@gmail.com**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Nin, V., Delgado, H., Muniz−Terrera, G., & Carboni, A. (2022). Partial agreement between task and BRIEF−P−based EF measures depends on school socioeconomic status. *Developmental Science*, *25*(5), e13241. Available at **https://doi.org/10.1111/desc.13241**

Abstract

Executive functions (EF), either conceptualized as skills involved in regulation of cognition and emotion in service of goal oriented behavior, or reductively as working memory, flexibility and inhibitory control, are commonly invoked constructs in developmental science. Two main traditions on EFs measurement prevail, one consisting of ratings obtained through questionnaires that inquire on behavior in common situations, the other based on performance in laboratory tasks. Whether both types of assessment actually refer to the same constructs is not consensual. Further, the role of school context in the degree of correspondence between both types of measures remains largely unexplored. Here, we show in a sample of over 220 children (age M= 5.6, SD= 0.4 years), by means of multilevel models, that whether EF tasks can predict BRIEF-P ratings and vice-versa, depends on the process considered and on the school SES. Inhibitory control, planning, and global executive functioning are associated with BRIEF-P ratings in all schools. In contrast, we found no association among measures of flexibility independently of school SES. For working memory, we found that questionnaire rating predicts span only in high SES schools, but span predicts behaviors across schools. Our findings contribute to a growing body of literature that proposes constructs assessed by questionnaires and tasks only partially overlap and suggests that school SES may be a relevant factor to consider when questionnaires are answered by teachers.

Keywords: executive functions, multilevel analysis, SES, self regulation, cognitive assessment, child development.

Introduction

The ability to voluntarily and consciously align thoughts and emotion with behavior in the pursuit of goals is the most usual definition of Executive Functions (EF) (Diamond, 2013; Zelazo, 2015, 2020). In spite of what seems to be a clear definition, conceptual and methodological challenges regarding EF persist in the literature (Doebel, 2020; Inzlicht et al., 2021). The work by Miyake and colleagues (Miyake et al., 2000) constitutes a turning point in terms of EF conceptualization. Their seminal work aimed to identify (part of) the cognitive processes that underlie performance in tasks that are frequently used to assess EF. To do so, they selected a number of tasks based on convenience criteria and analyzed the results by means of latent variable analysis. The results supported that EF can be partially dissociated in three components: memory (WM), inhibitory control (IC), and cognitive flexibility (CF), and a unitary factor. Although later publications do not always find support for a tripartite model in adults (Karr et al., 2018), and a less differentiated model is proposed for younger children (Karr et al., 2018; Lehto et al., 2003), the reduction of EF skills to these 3 specific components has gained traction in the literature. As a result, it has become frequent to encounter definitions that refer to 3 EF: WM, inhibition and flexibility, and to assume these are the building blocks of more complex, or higher, EF as planning or reasoning (Diamond, 2013).

This predominant view has recently been challenged by Doebel who proposes that EF should be instead conceptualized as the integration of knowledge, beliefs and values when applying control over cognitive, motor and perceptual processes for goal pursuit in specific situations (Doebel, 2020). For instance, performance in a classic self regulation task depends on beliefs about group norms (Doebel & Munakata, 2018), whether others would find out about the subject's behavior (Ma et al., 2020), or takes place in isolation or in pairs (Koomen et al., 2020). In addition, this author has shown that exposure to language appropriate to deal with contrasting stimuli characteristics improves performance in inhibitory control and flexibility tasks (Doebel & Zelazo, 2016b). Doebel proposes that children do not simply develop an inhibitory process they then can apply across a wide range of situations, but rather, what children develop is a skill to use control in particular situations.

These two different ways of conceptualizing EF are in turn related to two clear perspectives regarding EF measurement (Malanchini et al., 2019; Toplak et al., 2013). The first one, rooted in clinical and experimental psychology, employs tasks that are administered individually, under optimal experimental conditions, and that may or not be emotionally neutral, to measure specific cognitive processes. Examples of tasks with neutral emotional content used to evaluate WM are the Corsi blocks and the digit span in their forward and backward versions (Oberauer et al., 2018). Inhibition capacity is usually evaluated through semantic incongruence tasks such as Stroop or spatial incompatibility tasks such as Simon (Davidson et al., 2006). For younger children, tasks that do not require well developed reading skills have been developed to replace tasks based on semantic incongruence. Examples of these are the day-and-night stroop (Gerstadt et al., 1994), the fruit stroop (Archibald & Kerns, 1999) and the animal stroop (I. Wright et al., 2003). Flexibility tasks generally involve switching between prompts or rules, for example, the dimensional-change card sorting (Doebel & Zelazo, 2016a), or the mixed block of the Hearts and Flowers task (Davidson et al., 2006).

On the other hand, the second tradition of evaluation proposes the use of questionnaires, which in the case of children are usually completed by parents or teachers, to inquire about everyday situations that neither reflect emotionally neutral nor controlled situations, and require the coordination of multiple underlying processes, contextual knowledge and motivation. One scale that is gaining popularity in the literature is the Behavioral Rating Inventory of Executive Functions (BRIEF) (Gioia et al., 2000) and its preschool version BRIEF-P (Isquith et al., 2004). Both present items that evaluate how common certain behaviors are, grouped in 5 scales and 3 indices.

Taking into consideration the differences described for performance-based and questionnaire-based evaluations, it is not surprising that several studies on the convergent validity of these types of measures suggest that they weakly correlate (Duckworth & Kern, 2011; Enkavi et al., 2019; Mahone & Hoffman, 2007; Malanchini et al., 2019; Saunders et al., 2017; Toplak et al., 2013). Toplak et. al. (2013) reviewed specifically the relation between BRIEF and EF tasks in clinical and non-clinical populations and found significant, although low associations, between both.

Another important issue to consider regarding EF is the influence of context in its acquisition and manifestation (Ellis et al., 2017). According to contemporary theories of development, EF developmental trajectories are the product of a multilevel, culturally and historically situated dynamic system, in which biological and environmental factors modulate each other over time (Blair & Raver, 2012; Lerner, 2006). An improvement of EF during childhood has been consistently reported, together with the maturation of the prefrontal cortex that supports EF skills (Farah, 2017; Zelazo, 2020). In addition, the support for an association between socioeconomic status (SES) and EF development measured by means of laboratory tasks is strong (Farah, 2017; Hackman et al., 2015; Lawson et al., 2018). In Doebel's understanding of EF, socioeconomic differences in the ability to excerpt control in similar situations is expected given that the knowledge, values, norms and types of circumstances encountered during development varies across contexts (Doebel, 2020).

Taking in consideration the concerns regarding whether or not questionnaires and tasks evaluate the same constructs, and whether EFs should be conceptualized mechanistically or holistically, it is not clear what to expect of the relationship between informant-based measures and socioeconomic circumstances, although Sherman and Brooks report that children from lower SES tend to get ratings that indicate difficulties with EF (Sherman & Brooks, 2010). In addition to all of the above, another concern regarding informant-based measures lies in the fact that responses are subject not only to typical measurement errors, but also to expectation effects and biases that affect the raters (Fuhs et al., 2015; Mashburn et al., 2006). For instance, teachers' evaluations of EF may be subject to what is known as the 'reference group effect', the fact that each teacher will judge a child's abilities relative to their peers' (Fuhs et al., 2015). The perception that teachers have in relation to their students' abilities is in turn associated with the classroom context: teachers in disadvantaged schools tend to underestimate their students' literacy abilities (Ready & Wright, 2011). Whether this finding holds true for basic skills evaluations made by teachers of younger children is, at the best of our knowledge, unknown.

In light of these conflicting views, understanding the relationship between EF measured by tasks and questionnaires may shed light on the relation between specific cognitive processes and the resolution of real life challenges for children from diverse socioeconomic backgrounds. The present work seeks to provide evidence in relation to two key issues. First, on whether individual, test-based measures of EF, and measures provided by teachers through the BRIEF-P questionnaire tap into the same underlying processes, that is, whether one can predict the other. Based on the previous findings reported in the literature and recapitulated in this introduction, we hypothesized that a low to moderate association for the performance-based and rating measures of inhibitory control, flexibility and planning would emerge. The second issue is whether school's socioeconomic characteristics are associated with, and moderate, EF measures obtained by both methodologies. The absence of prior studies on this matter, and the multiple factors that may influence the evaluation teachers make about their students, makes it difficult to advance hypotheses in this regard: we can not anticipate whether a modulation of the associations for each specific process will emerge. Given the nested structure of the data, we used multilevel analysis, a statistical methodology that allows to disentangle within and between classes variance, and the inclusion of predictors at both individual and school levels.

Methods

Participants A total of 308 children (150 girls) were recruited for the study from 14 kindergarten classes in 10 public schools in Montevideo, Uruguay, classified either as the lowest or highest Sociocultural Level (SCL) quintile by the National Administration of Public Education. Data from 4 children were removed from the analytical sample because they were outliers according to their age (age above /below 3 SD samples' average age). Teachers were asked to complete the BRIEF-P for each child in the classroom (see below), but failed to return questionnaires for 7 children, and 1 questionnaire was incomplete. We excluded children whose BRIEF-P questionnaires had unacceptable inconsistency scores (for further detail on BRIEF-P treatment refer to the corresponding paragraph in this section). The final sample consisted therefore of 261 children (129 girls) in kindergarten aged M= 5.6, SD= 0.4 years. Not all children attended all testing sessions; the number of participants for each analysis is reported in the results section. Informed consent was obtained from parents/caregivers in accordance with the World Medical Association (WMA) Declaration of Helsinki. The study was approved by the Research Ethics Committee of -blinded for review-.

Design and procedures Children were administered a battery of EF tasks at their schools, during 2 sessions (two to three tasks per session) at times reported by teachers not to interfere with regular meals and activities. The detailed description of the procedures is available in the Supplemental Material. The tasks used in this article have been reported before (Goldin et al., 2014), and will be briefly described in the corresponding section. In addition, teachers filled the Behavior Rating Inventory of Executive Function-Preschool version (BRIEF-P) (Isquith et al., 2004) for every child in their classroom.

School Sociocultural Level and Home Socioeconomic Status Schools in Uruguay are classified by the National Administration of Public Education according to a continuous variable (the School Sociocultural Level, SCL) a composite of three sets of measures: socioeconomic status of the school's overall population, a measure of maternal education for every child in the school, and and indicator of social integration that combines the proportion of homes with school-aged children that do not attend formal education and the proportion of children that live in urban slums (ANEP, 2016). The specific score of each school is not open access, however, the distribution in quintiles is public: those schools classified as the lowest quintile serve the poorest population and are usually located in vulnerable quarters. On the other hand, schools in the highest quintile serve children from homes located in urban neighborhoods characterized by middle and middle high-class homes. Taking schools in the extremes of the distribution ensures that the differences attributable to socioeconomic circumstances are captured. This approach has also been followed in the US (Torff & Sessions, 2009).

Cognitive Assessment

Inhibitory Control & Cognitive flexibility Task We used an adaptation of the Hearts and Flowers task (A. Wright & Diamond, 2014). Two types of stimuli were presented either at the left or right of the screen on each trial. The congruent stimuli required children to press the button on the same side of the screen and the incongruent stimuli required a button press on the opposite side of the screen. An initial block of 12 congruent trials (all responses on the same side) was followed by a block of 12 incongruent trials (all responses on the opposite side), and then by a mixed block of 24 trials where congruent and incongruent trials were randomly intermixed. Instructions and a practice block of 8 trials were given before each block. We used the Proportion of Correct Answers (PCA) in the incongruent block as the measure of inhibitory control, like other studies in the region with similar samples (Giovannetti et al., 2020; Hermida et al., 2015). We used the PCA in the mixed block as a measure of flexibility (Camerota et al., 2019; Davidson et al., 2006).

Visuo-spatial Working Memory Task Corsi Blocks task (Corsi, 1972). Children were presented with a 3x3 matrix of light bulbs. First, the number of lights that would turn-on was shown. Then, the sequence of lamps lit up in a random order. Children had to reproduce the sequence by touching the corresponding light bulbs. Each block included five trials ranging from 1 lamp to the maximum amount possible before reaching the cutoff criteria of three consecutive errors. We used the longest remembered sequence as a measure of WM span.

Planning Task Tower of London task (Shallice, 1982). Children were shown a goal configuration with 3 colored balls in poles of three different heights, an initial configuration, and the number of moves required to transition from the initial to the goal configuration. As children progress through trials the amount of moves required to solve the problem increases. Cutoff criteria was three consecutive errors. The measure for planning was the number of moves of the problem with the highest amount of required moves successfully resolved.

Behavioral Rating Inventory of Executive Function-Preschool (BRIEF-P) Questionnaires were completed by teachers. We did not include in the analysis 35 questionnaires with unacceptable inconsistency scores. We did not find differences in the frequency of unacceptable inconsistency scales among low and high socioeconomic status according to a chi squared test (χ-squared = 2.0909, df = 1, p-value = 0.1). Fifteen questionnaires had higher than cutoff values in the negativity scale. We carefully analyzed these and found that all had T-scores greater than 65 in at least one index, values considered clinically significant (Isquith et al., 2004). Given that clinical populations may have high negativity scores, and according to Sherman and Brooks recommendations, we decided to include questionnaires with a high negativity score that also presented T-scores higher than 65 in any scale (Sherman & Brooks, 2010).

Statistical analyses

Before running the analytical plan to check our hypothesis, we ran several models to investigate whether the quality of the data acquired during the evaluation sessions was adequate. To do so, we first run logistic regressions with the complete dataset (that is, all trials for all children included in the final sample) with children nested into their classes. The results of these analysis corroborated that the data follows the behavior expected for each task. The methodology and results are presented in detail in the Supplemental Material. Next, we performed multilevel analysis to determine whether BRIEF measures can predict classic EF measures, whether performance in tasks can predict BRIEF-P ratings, and how socioeconomic background influences these associations. All models were fitted using R (R Core Team, 2020) using packages lme4 (Bates et al., 2015), emmeans (Lenth, R, 2020), and sjstats (Lüdecke D, 2021). We built the models by sequentially including child level variables, then school level variables, and finally an interaction term. The child level variables of interest were BRIEF scales T-scores and performance in the selected tasks, child level covariates were age, gender and home SES, and the higher level variable was School Sociocultural Level (SCL). Because of the high correlation between Home SES and School SCL, the former was included only if it significantly improved the model's fit. Children's measurements were nested within their classes by means of a random intercept term. We fitted several models: a null model with no predictors to calculate the intraclass correlation, a model with child level variables, and a model which added the school variable. We also fitted a model with an interaction term between school SCL (a variable entered as Low or High SCL) and the variables of interest. Age was centered by calculating the difference in relation to the youngest child in the sample, that is, the youngest child entered with an age of 0. EF task performance measures and home SES were transformed to z-values and BRIEF raw scores to T-scores. We used a Type III Analysis of Variance with Satterthwaite's method to calculate the estimated p-values of the fixed effects. The procedure described was conducted to optimize fitting while taking in consideration data structure in a hypothesis driven approach to data analysis (Meteyard & Davies, 2020).

Results

Demographic variables The basic demographic for both groups of schools are shown in Table 1. There were no differences in age or gender frequency among low and high SCL schools. We found that, as expected, the SES score of homes in low SCL schools was on average 1.4 SD below homes in high SES schools (t = -15.48, df = 254.93, p-value < 0.001). However, it is important to notice that children who attend high SCL public schools come from homes actually classified as medium to medium-high SES and do not represent the highest SES in absolute terms. This is in fact expected because high-class children in Uruguay tend to attend private schools. Descriptive statistics for all variables and zero-order correlations among all variables are shown in Supplementary Tables 1 and 2.

Table 1 Demographic variables

	School SCL		
Variable	*Low, N = 141*†	*High, N = 120*†	*p*
Gender			0.5
F	67(48%)	62(52%)	
M	74(52%)	58(48%)	
Age, years	5.61 (0.45)	5.65 (0.36)	0.6
Home SES	27 (10)	46 (10)	<0.001

†n (%); Mean (SD)

Inhibitory Control A total of 230 children had data on the second block of the Hearts and Flowers-like test and BRIEF-P questionnaires. Logistic regressions of the dataset showed the quality of the dataset allowed for further processing (Supplementary Tables 3 to 6). A chi squared test to compare the proportion of children from low and high SCL schools that undertook the test showed no difference in attrition (χ-squared = 0.89, df = 1, p-value = 0.3). We started by analyzing whether BRIEF-P scores predict the Proportion of Correct Answers (PCA) in the incongruent block of the task. We ran a null model to determine the PCA attributable to class belonging, and found that it is only 10% (ICC=0.1). Inclusion of age and gender improved the model fit in every model we ran, and were always included. Addition of Home SES did not improve fit and therefore was not included. The following step, addition of the Inhibit scale T-scores resulted in improved fit (p<0.01), its estimate shows the appropriate sign and it significantly explains variance in the incongruent block performance. Then, we included the variable of interest at the second level, school sociocultural level (SCL), which also resulted in a better fit (p<0.05), and also explained part of the performance in the task. Model results are presented in Table 2. As can be seen, both variables of interest, BRIEF-P assessment and school SCL are predictors of inhibitory control. The last step was the introduction of the interaction term between school SCL and the perception of inhibitory control abilities measured with the questionnaire. The interaction term did not result in an improvement of fit (p=0.2), which suggests that BRIEF-P has similar predictive value for both low and high SCL schools. Indeed the parameters from this model shown in Supplementary Table 7 indicate that the interaction term is not statistically significant, and its inclusion undermines the predictive value of the isolated terms.

Inhibitory control could also be associated with the ability of children to control their emotions (Carlson & Wang, 2007). In order to check this possibility, we ran similar models as the ones described in the previous paragraph, this time with the Emotion Control (EC) scale of the BRIEF-P questionnaire. Inclusion of the Emotion Control scale improved the fit (p<0.05), and inclusion of School SCL did too (p<0.05). The results are summarized in Table 2. In the following step, the interaction term did not result in an improvement of fit (p=0.4), which suggests that BRIEF-P EC scale has similar predictive value of task performance in both low and high SCL schools. Again, the parameters of this model, shown in Supplementary Table 7, indicate that the interaction term is not statistically significant, and its inclusion undermines the predictive value of the isolated terms.

We then assessed the association in the opposite direction, that is, whether PCA can predict the broader skills reflected in BRIEF-P Inhibit and Emotional Control scales. We followed the same steps described in the previous paragraph. The ICC in both cases is 0.2, therefore, the effect of class belonging is twice as big when we consider ratings by teachers. Again, adding home SES did not improve fit (p=0.9 for both the Inhibit and EC scales). However, inclusion of PCA in the incongruent block further improved the fit (p<0.01 and p<0.05 for IC and EC scales) and significantly explained part of the variance in the Inhibition scale ratings. In contrast to what we found in the models to predict performance in the incongruent block of the Hearts and Flowers-like task, adding school SCL did not explain further variance in neither model (p=0.3 and p=0.7 for IC and EC scales) in the BRIEF-P ratings. All model estimates are shown in Table 3. Given this lack of a main effect of school SCL we did not run the models with the interaction term.

Flexibility The third, mixed block, of the Hearts and Flowers-like task requires children to alternate responses between congruent and incongruent stimuli presented in a pseudo aleatory order. It has been argued that therefore this block requires a flexibility component (Camerota et al., 2019; Davidson et al., 2006). We analyzed whether the Shift Scale of the BRIEF-P is a valid predictor of cognitive flexibility as assessed by the PCA in the mixed block. Two hundred twenty-eight children completed the task, for whom we followed the same procedure as before. Similar to the previous EF measures, a small, although greater proportion of variance of PCA in the third block of the task is explained by group belonging (ICC=0.16). Unlike the previous results, we found that Home SES significantly improved the fit and was therefore also incorporated (p<0.001). We then incorporated the Shift Scale T-values and, different from the results described up to this point, we did not find an improvement of fit (p=0.3). To check if SCL explains part of the performance, we proceeded to incorporate the school SCL term, which improved the fit (p<0.05). In summary, results shown in Table 4 indicate that the score obtained in the Shift Scale is not associated with performance in the mixed block of the Hearts and Flowers task but it is to school SCL. Due to the lack of main effect of the BRIEF-P we did not include an interaction term.

Table 2. Association of the Proportion of Correct Answers (PCA) in the incongruente block of the Hearts and Flowers-like task and the Inhibition or Emotion Control scales of the BRIEF-P.

	B2 PCA_z			B2 PCA_z		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.19	-0.67 – 1.05	0.658	-0.11	-0.95 – 0.74	0.804
age	0.21	-0.12 – 0.54	0.212	0.22	-0.11 – 0.55	0.197
Gender [M]	0.47	0.22 – 0.73	**<0.001**	0.43	0.18 – 0.69	**0.001**
Inh_T	-0.02	-0.03 – -0.01	**0.005**			
SchoolSCL	0.11	0.01 – 0.20	**0.027**	0.12	0.03 – 0.21	**0.011**
EC_T				-0.01	-0.03 – -0.00	**0.043**
Random Effects						
σ^2	0.87			0.89		
τ_{00}	0.06 $_{Class}$			0.06 $_{Class}$		
ICC	0.07			0.06		
N	14 $_{Class}$			14 $_{Class}$		
Observations	230			230		
Marginal R^2 / Conditional R^2	0.132 / 0.192			0.115 / 0.169		

Next, we ran the models to assess whether performance in the mixed block holds predictive value of the BRIEF-P Shift scale. Again the ICC for the null model was higher in this case (ICC=0.24). Contrary to the previous models, home SES did not improve fit (p=0.6) and therefore was not included. Inclusion of the z-scores in the mixed block did not improve fit (p=0.3), and neither did adding School SCL, suggesting that performance in the mixed block of the task does not explain the ratings in the Shift scale of BRIEF-P, and nor does school type. The estimates of this latter model are shown in Table 5.

Working memory Two hundred twenty-three children had Corsi blocks span data and BRIEF-P questionnaires. A chi squared test to compare the proportion of children from low and high SCL schools that undertook the test showed no difference in attrition (χ-squared = 0.19, df = 1, p-value = 0.7). The quality of the dataset allowed for further analysis (see Supplemental Material Tables 8 and 9). We first fitted the null model (ICC=0.11). Inclusion of home SES did not

Table 3. Association of the the Inhibition scale rating and the Emotional Control scale with the Proportion of Correct Answers (PCA) in the incongruent block of the Hearts and Flowers-like task.

	EC_T			Inh_T		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	45.92	39.91 – 51.93	**<0.001**	46.03	40.02 – 52.05	**<0.001**
age	1.84	-1.42 – 5.11	0.268	2.15	-1.07 – 5.36	0.190
Gender [M]	3.82	1.54 – 6.11	**0.001**	5.17	2.93 – 7.41	**<0.001**
Block 2 PCA_z	-1.19	-2.36 – -0.02	**0.047**	-1.71	-2.86 – -0.56	**0.004**
SchoolSCL	-0.24	-1.53 – 1.05	0.718	-0.75	-2.06 – 0.57	0.265
Random Effects						
σ^2	70.42			67.84		
τ_{00}	18.36 $_{Class}$			19.28 $_{Class}$		
ICC	0.21			0.22		
N	14 $_{Class}$			14 $_{Class}$		
Observations	230			230		
Marginal R^2 / Conditional R^2	0.057 / 0.252			0.124 / 0.318		

Table 4. Association between BRIEF-P Shift Scale and the Proportion of Correct Answers (PCA) in the Mixed block of the Hearts and Flowers-like task.

	B3 PCA_z		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.79	-1.57 – -0.01	**0.048**
Age	0.33	0.03 – 0.63	**0.031**
Gender [M]	0.33	0.10 – 0.55	**0.004**
SES_z	0.18	0.01 – 0.34	**0.034**
Shif_T	-0.01	-0.02 – 0.01	0.294
SchoolSCL	0.10	0.00 – 0.20	**0.047**
Random Effects			
σ^2	0.70		
$\tau_{00\ Class}$	0.05		
ICC	0.06		
N_{Class}	14		
Observations	228		
Marginal R^2 / Conditional R^2	0.184 / 0.236		

Table 5. Association between the Proportion of Correct Answers (PCA) in the Mixed block of the Hearts and Flowers-like task and ratings in the Shift scale.

	Shif_T		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	47.63	40.98 – 54.27	**<0.001**
Age	0.13	-3.18 – 3.45	0.937
Gender [M]	3.52	1.25 – 5.80	**0.002**
Block 3 PCA_z	-0.75	-2.06 – 0.57	0.266
SchoolSCL	-0.02	-1.50 – 1.47	0.982
Random Effects			
σ^2	69.36		
$\tau_{00\ Class}$	25.59		
ICC	0.27		
N_{Class}	14		
Observations	228		
Marginal R^2 / Conditional R^2	0.033 / 0.294		

improve fit (p=0.1) and therefore was not incorporated. We then included the variable of interest at child level, the standardized T score of the Working Memory scale, which improved model fit ($p<0.01$). Next, we included school SCL, which further improved the fit ($p<0.01$). In the last step we included an interaction term between school SCL and BRIEF-P WM, that also improved fit ($p<0.01$). We used this final model to evaluate the predictive capacity of the Working Memory index of the BRIEF-P questionnaire and whether it varies for high and low SCL schools. As shown in Table 6, the relation between both measures shows the appropriate sign, although it is not statistically significant. However, as can also be seen in Figure 1, the relation of both variables is modulated by school SCL.To further understand the modulation by school SCL, we computed the slopes for both types of schools and found that the slope for low SCL schools was -0.00819 (CI 95% [-0.0262 to 0.00981]) and the slope for high SCL schools was -0.05401 (CI 95% [-0.0830 to -0.02503]), which means that the slope in high SCL schools is approximately 7 times steeper. Importantly, the 95% confidence interval for High SCL schools does not contain zero, but for Low SCL schools it does. These results indicate that the WM scale of the BRIEF-P holds predictive value of performance in Corsi Blocks only in High SCL schools. Moreover, the pairwise difference between slopes was significant (t ratio= 2.69, $p<0.01$).

We then followed the same steps but for the opposite association. Again, we found that the ICC for teachers ratings of working memory (ICC=0.23) is twice the value for task measures. In this case, adding home SES did not improve fit (p=0.2). Adding the variable of interest, span in Corsi blocks, resulted in better fitting ($p<0.01$), but addition of school SCL did not significantly improve the fit (p=0.1), results are shown in Table 7.

Planning Due to a technical problem in two low and two high SCL schools, some data points for the Tower of London (TOL) task were lost, therefore, we assessed a smaller sample consisting of 191 children. A chi squared test to compare the proportion of children from low and high SCL schools that undertook the test showed no difference in attrition (χ-squared = 0.12, df = 1, p-value = 0.7). We then run a similar set of analysis as with Corsi blocks in order to check the quality of our data set, the results are shown in Supplementary Tables 10 and 11. Finally, we proceeded to run the sequence of models to test our hypothesis. The ICC for TOL in the null model was 0.25, the higher variance attributable to class clustering in our study. Introduction of gender and age improved the fit of the model ($p<0.05$), also inclusion of Home SES ($p<0.01$) and the variable of interest the Plan/Organize scale of BRIEF-P ($p<0.01$). In contrast with the previous tasks, inclusion of the school level variable did not improve model fit (p=0.8). We tested whether this result was explained by the high correlation among home SES and school SCL. To do so, we ran a model removing home SES and keeping school SCL, however, in this model the main effect of school SCL did not achieve significance (p=0.4), nor did its inclusion improve the fit (p=0.3). All the estimates and the corresponding p values in each of these models are presented in Table 8.

Next, we ran the models to evaluate the predictive value of TOL performance of planning behaviors. The ICC of the PO scale null model was 0.25. Inclusion of home SES did not result in better fitting (p=0.8), but adding the performance in TOL did ($p<0.01$). Again, inclusion of school SCL did not (p=0.2), so we did not proceed to include the interaction term; estimates are shown in Table 9.

Global Executive Functioning The successful resolution of the mixed block of the Hearts and Flowers-like task requires cognitive flexibility to alternate between rules, inhibitory control to refrain the automatic, prepotent response and working memory to hold both rules in consciousness (Davidson et al., 2006); therefore it could be considered an indicator of integrated executive functioning. The BRIEF-P questionnaire combines all its scales in a Global Executive Composite (GEC). We evaluated the relation among these following the same analytical procedure. The first steps of model building were described in the section referring to flexibility, so we continued with the inclusion of GEC-T values into the model, which resulted in better fit ($p<0.01$). The incorporation of school SCL did not result in an

Table 6 Association between BRIEF-P WM scale and Corsi Blocks span.

	MaxSpan_z			MaxSpan_z		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.07	-0.92 – 1.06	0.888	-1.15	-2.46 – 0.17	0.088
Age	0.47	0.14 – 0.80	**0.005**	0.48	0.15 – 0.80	**0.004**
Gender [M]	-0.02	-0.27 – 0.22	0.851	0.03	-0.22 – 0.27	0.826
SchoolSCL	0.12	0.03 – 0.20	**0.005**	0.66	0.26 – 1.07	**0.001**
WM_T	-0.02	-0.04 – -0.01	**0.008**	0.00	-0.02 – 0.03	0.780
WM_T * SchoolSCL				-0.01	-0.02 – -0.00	**0.007**
Random Effects						
σ^2	0.83			0.81		
τ_{00}	0.04 Class			0.03 Class		
ICC	0.04			0.04		
N	14 Class			14 Class		
Observations	223			223		
Marginal R^2 / Conditional R^2	0.161 / 0.197			0.185 / 0.217		

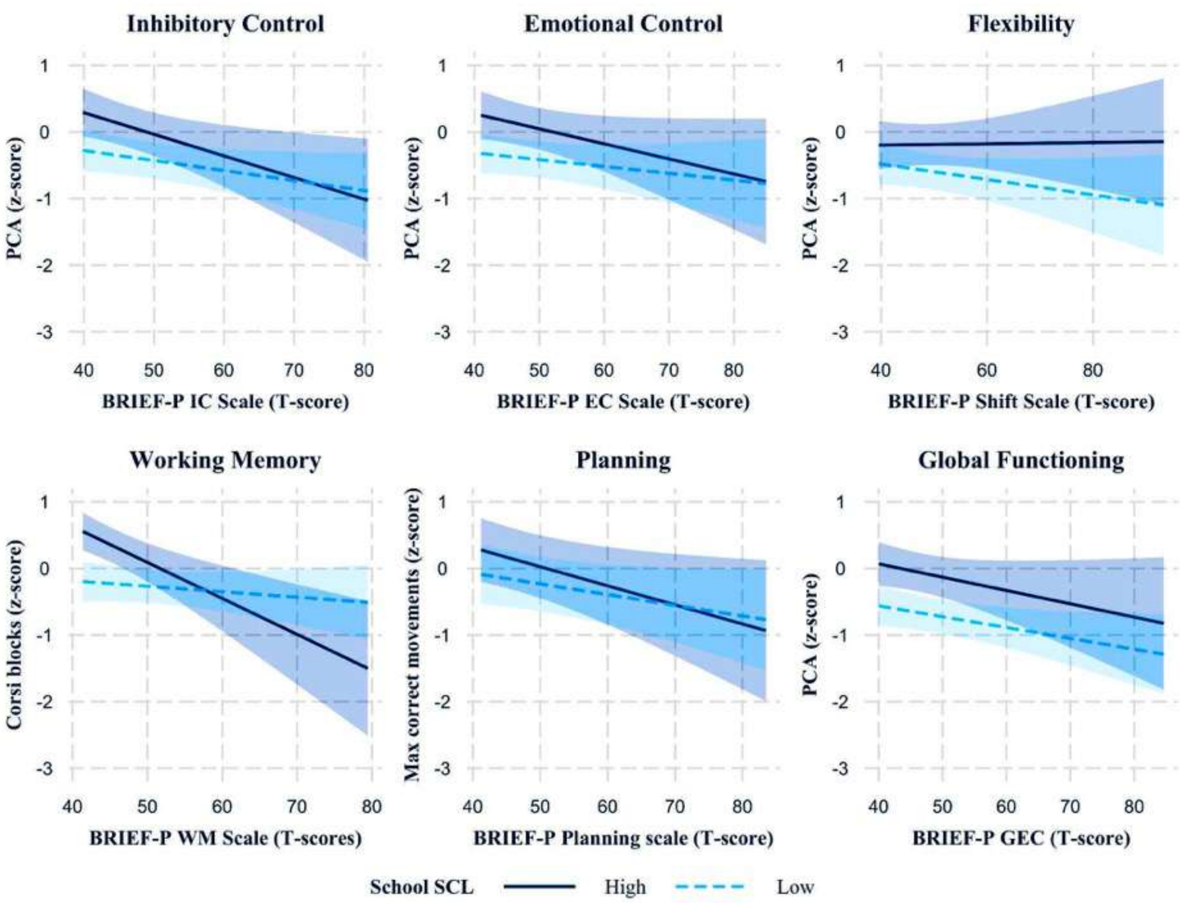


Figure 1. Interaction plots showing the modulation of the association between BRIEF-P and performance in laboratory tasks by School SCL

Table 7 Association between Corsi Blocks span and BRIEF-P WM scale ratings.

	WM_T		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	52.88	47.47 – 58.28	**<0.001**
Age	-1.89	-4.94 – 1.17	0.226
Gender [M]	2.27	0.22 – 4.32	**0.030**
MaxSpan_z	-1.59	-2.68 – -0.50	**0.004**
SchoolSCL	-0.90	-1.99 – 0.19	0.104
Random Effects			
σ^2	56.02		
τ_{00} Class	12.68		
ICC	0.18		
N Class	14		
Observations	223		
Marginal R^2 / Conditional R^2	0.128 / 0.289		

improvement of fit (p=0.1), not surprisingly in view of the high correlation between home SES and school SCL. Indeed, a model that includes School SCL but not home SES results in a significant proportion of variance explained by this variable. The results of these models are shown in Table 10. Noteworthy, the GEC is a statistically significant predictor of the mixed block PCA in the three models. Finally, in accordance with the theory-driven approach to analysis, we also ran a model with the interaction term between school SCL and the GEC. In this model the interaction term did not reach significance (p=0.8), therefore, the relationship between both approaches to global functioning is not modulated by school SCL (Supplementary Table 12).

The ICC for BRIEF-P GEC ratings was 0.3, almost twice as large as the ICC for the performance in the mixed block of the Hearts and Flowers-like task. Adding the covariates resulted in better fitting (p<0.001), home SES did not (p=0.3), and performance in the mixed block improved fit (p<0.01). Akin to all models where we used to explain BRIEF-P variance, inclusion of school SCL did not result in improvement of the model (p=0.3). This latter model estimates are shown in table 11.

The plots illustrating the modulation of school SCL of the predictive capacity of the BRIEF-P of performance in tasks are presented in Figure 1.

Conclusions

The present work sought to provide empirical evidence to contribute to the ongoing discussion of methodological approaches used to measure EF. In the first place, we assessed the extent of the predictive capacity of the teacher version of the BRIEF-P regarding the performance in classic neurocognitive tasks, and vice-versa, using multilevel analysis. Then, we conducted analyses to understand whether school socioeconomic characteristics (here referred to as school sociocultural level, SCL) modulate the association among task-based and questionnaire-based assessment of EF. To the best of our knowledge, this is the first work that considers school characteristics when analyzing the convergence of task and informant-based EF measures. We found an heterogeneous set of results that we proceed to discuss.

For the constructs of inhibitory control and planning, our results are consistent with studies reporting low but significant associations among performance-based and informant based measures (Garon et al., 2016; Tamm & Peugh, 2019; Toplak et al., 2013). For these

Table 8. Predictive value of the BRIEF-P Plan/Organize Scale of the maximum amount of movements in correct answers (Max_z) in the Tower of London Task

	Max_z			Max_z			Max_z		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.46	-0.48 – 1.39	0.338	0.51	-0.53 – 1.54	0.336	0.22	-0.86 – 1.29	0.694
Age	0.39	0.03 – 0.76	**0.035**	0.40	0.03 – 0.76	**0.034**	0.39	0.01 – 0.77	**0.047**
Gender [M]	-0.11	-0.37 – 0.14	0.380	-0.11	-0.37 – 0.14	0.379	-0.08	-0.34 – 0.18	0.534
SES_z	0.25	0.08 – 0.41	**0.004**	0.25	0.07 – 0.44	**0.007**			
PO_T	-0.02	-0.04 – -0.01	**0.005**	-0.02	-0.04 – -0.01	**0.005**	-0.02	-0.04 – -0.01	**0.009**
SchoolSCL				-0.02	-0.16 – 0.13	0.826	0.07	-0.08 – 0.22	0.353
Random Effects									
σ^2	0.68			0.68			0.72		
τ_{00}	0.17 Class			0.17 Class			0.24 Class		
ICC	0.20			0.20			0.25		
N	14 Class			14 Class			14 Class		
Observations	191			191			191		
Marginal R^2 / Conditional R^2	0.133 / 0.305			0.126 / 0.299			0.097 / 0.324		

constructs it could be interpreted that the items in the corresponding BRIEF-P Inhibitory control, Emotional Control and Planning/Organizing scales capture situations that do indeed require at least partially the ability to contain preponderant actions and self-organize behavior. The opposite situation is also true, from our results, it seems that the basic cognitive process captured by the tasks are building blocks for behaviors in the complex situations encountered at school that require self-regulation and organization.

With regards to flexibility, while the BRIEF-P inquires mainly about the ability of children to adapt to new social situations and to changes in the routine, the mixed block of the Flower and Hearts task involves flexibility in terms of a rule change. We found that ratings in the Shift scale do not predict performance in the task, nor performance in the mixed block of the Hearts and Flowers-like test relate to flexible behavior in social situations. Consistent findings were reported by Tamm and Peugh (2018) using a broader set of tasks. These results suggest that both measures do not reflect the same underlying construct, possibly because the type of situations depicted in the BRIEF-P items require a wide range of socioemotional skills not involved in the Flower and Hearts task, while also suggesting that flexibility as measured in the mixed block of the task does not play a role in complex, social situations that require children to adapt to novel situations.

A different situation emerged for the working memory construct. Working memory is, broadly defined, a system, or set of processes, that hold mental representations for use in thought and action (Oberauer et al., 2018). The differences among Corsi blocks and the BRIEF-P WM scale are perhaps the most evident. Whereas the forward version of the Corsi Blocks is a serial recall type of task (a.k.a simple span), the BRIEF-P items address performance in a wide range of broadly presented situations, generally pertaining to distractibility and the ability to remain on task. Even though the large distance between both assessments, BRIEF-P WM scale does predict the simple memory span, strikingly, only for children that attend high SCL schools. Why? The analysis of items contained in the WM scale of BRIEF-P suggests that it reflects not only WM capacity, but also self regulating abilities

Table 9. Association between the Maximum amount of movements in correct answers (Max_z) in the Tower of London Task and the BRIEF-P Plan/Organize Scale.

	PO_T		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	51.47	45.08 – 57.86	<0.001
Age	-1.51	-5.04 – 2.02	0.402
Gender [M]	3.13	0.78 – 5.47	0.009
Max_z	-1.75	-3.05 – -0.46	0.008
SchoolSCL	-0.77	-2.06 – 0.51	0.237
Random Effects			
σ^2	61.19		
$\tau_{00\ Class}$	17.99		
ICC	0.23		
N_{Class}	14		
Observations	191		
Marginal R^2 / Conditional R^2	0.113 / 0.315		

Table 10. Predictive power of global executive functioning (GEC) scale of performance in the mixed block of the Hearts and Flowers-like task.

	Block 3 PCA_z			Block 3 PCA_z			Block 3 PCA_z		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.03	-0.74 – 0.79	0.947	-0.38	-1.20 – 0.44	0.364	-0.26	-1.07 – 0.56	0.534
Age	0.33	0.04 – 0.62	**0.026***	0.27	-0.02 – 0.56	0.065	0.31	0.02 – 0.60	**0.035***
Gender [M]	0.37	0.15 – 0.60	**0.001 ****	0.39	0.16 – 0.61	**0.001 ****	0.37	0.15 – 0.60	**0.001****
GEC_T	-0.02	-0.03 – -0.01	**0.005***	-0.02	-0.03 – -0.00	**0.011 ***	-0.02	-0.03 – -0.00	**0.011***
HomeSES_z	0.25	0.12 – 0.38	**<0.001*****				0.18	0.02 – 0.33	**0.028***
SchoolSCL				0.15	0.07 – 0.23	**<0.001*****	0.09	-0.01 – 0.18	0.083
Random Effects									
σ^2	0.67			0.68			0.67		
τ_{00}	0.05_{Class}			0.04_{Class}			0.04_{Class}		
ICC	0.07			0.06			0.06		
N	14_{Class}			14_{Class}			14_{Class}		
Observations	228			228			228		
Marginal R^2 / Conditional R^2	0.171 / 0.226			0.194 / 0.242			0.209 / 0.253		

that are, in turn, strongly associated with socioeconomic factors (Blair, 2010; Evans & Kim, 2013; McEwen & McEwen, 2017). So, in kids from more privileged schools, better abilities to deal with everyday situations that require enough span to hold relevant information and simultaneously self monitoring abilities is also evident in a higher capacity to remember longer sequences in WM. Therefore the association between both measures is present, as has been shown for adults (McCabe et al., 2010). However, for kids that attend low SCL schools, the self monitoring component of WM could be compromised and given that performance in everyday situations not only requires good memory span, but also the ability to control attention and behavior, the association between span and actual behavior would be lost.

Given the heterogeneous capacity of the BRIEF-P scales to predict performance in tasks, we were surprised to find a significant association of the GEC with performance in the mixed block of the Flowers and Hearts-like task. This may imply that the cognitive processes needed for successful execution in the mixed block also take part in the successful navigation of the complex behaviors required in classroom situations. On the other hand, BRIEF-P GEC indeed captures situations that require the integration of working memory, flexibility and inhibitory control.

Finally, we found that for all tasks, school SCL partially explains performance; however, none of the BRIEF-P scales ratings show an association with School SCL. Therefore, it seems that tasks are sensitive to the school context characteristics but the questionnaire is not. One possible explanation for this discrepancy is that teachers judge children in relation to their peers, what is known as the reference group effect (Fuhs et al., 2015). The fact that all BRIEF-P models show higher ICCs than the tasks counterparts suggests that, indeed, teachers' measures lack the granularity obtained with the tasks. It has also been shown that teachers underestimate the literacy skills of children in disadvantaged schools (Ready & Wright, 2011). However, expanding the concept of "teacher bias" into the construct of EF is complicated due to the convergent validity issues reported in previous studies and confirmed here. The only process for which we found that teachers ratings in low SCL schools tended to be below the task counterpart was WM. However, as discussed before, we do not have elements to interpret this result as a teacher bias, instead, the analysis of the BRIEF-P items suggest that it evaluates what is known as the "central executive" in Baddeley and Hitch's model (Baddeley & Hitch, 1994) rather than WM span.

Table 11. Prediction of the GEC Scale by performance in the mixed block of the hearts and flowers-like task.

	GEC		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	81.06	64.61 – 97.51	**<0.001**
Age	1.19	-6.70 – 9.07	0.768
Gender [M]	11.99	6.63 – 17.36	**<0.001**
Block 3 PCA_z	-4.42	-7.52 – -1.31	**0.005**
SchoolSCL	-1.71	-5.50 – 2.09	0.378
Random Effects			
σ^2	385.54		
$\tau_{00\ Class}$	172.06		
ICC	0.31		
N_{Class}	14		
Observations	228		
Marginal R^2 / Conditional R^2	0.109 / 0.384		

Limitations and future directions The generalization of the findings presented here to other performance-based measures is challenging.The battery of tasks we used is limited to one per domain, and its associations (or lack of) with ratings are therefore circumscribed to that particular task. Despite the fact that the ones we selected are widely reported in the literature, we did not include a typical measure of cognitive flexibility, as the Dimensional Change Card Sort.

Another limitation of the present study is that we have not provided evidence of a potential age effect of the associations reported. The age of our sample is in the upper limit for the BRIEF-P, a situation that limits the possibility of following this cohort with the same instrument. In addition, we can not provide any evidence about whether one type of measurement holds better predictive value for scholastic achievement later on.

Finally, due to space limitations, we did not include Confirmatory Factor Analysis with our data to assess if performance-based and ratings measurements load into two separate factors, as would be expected if both types of measures do tap into different underlying constructs.

In sum, our work presents evidence consistent with a party agreement between the BRIEF-P and individual task performance, dependent both on the process considered, the socioeconomic background of the school, and the direction of the prediction. In this sense, we provide evidence of the importance of considering school level variables when evaluating EF, especially if these are based on teachers' judgments. Furthermore, our results are relevant because the possibility of evaluating EF through questionnaires is becoming more attractive as it does not require specifically trained personnel and it is less time consuming than applying test batteries individually, therefore reducing overall testing costs. Last but not least, by no means do we state that one measure is better than the other, rather, that one can not be substituted by the oher.

Disclosures

The authors declare no conflict of interests.

Funding

This work was supported by ANII grants FSPI_X_2015_1_108417, ININ_1_2017_1_137164 and CSIC grant 191120-000676-15

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

ANEP. (2016). *Relevamiento de características socioculturales de las escuelas públicas del Consejo de Educación Inicial y Primaria 2015*.

Archibald, S. J., & Kerns, K. a. (1999). Identification and Description of New Tests of Executive Functioning in Children. *Child Neuropsychology*, *5*(2), 115-129. https://doi.org/10.1076/chin.5.2.115.3167

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485-493. https://doi.org/10.1037/0894-4105.8.4.485

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist*, *67*(4), 309-318. https://doi.org/10.1037/a0027493

Camerota, M., Willoughby, M. T., & Blair, C. B. (2019). Speed and accuracy on the Hearts and Flowers task interact to predict child outcomes. *Psychological Assessment*, *31*(8), 995-1005. https://doi.org/10.1037/pas0000725

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development*, *22*(4), 489-510. https://doi.org/10.1016/j.cogdev.2007.08.002

Corsi, PM. (1972). *Human memory and the medial temporal region of the brain.*

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006).

Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(4), 942-956. https://doi.org/10.1177/1745691620904771

Doebel, S., & Munakata, Y. (2018). Group Influences on Engaging Self-Control: Children Delay Gratification and Value It More When Their In-Group Delays and Their Out-Group Doesn't. *Psychological Science*, *29*(5), 738-748. https://doi.org/10.1177/0956797617747367

Doebel, S., & Zelazo, P. D. (2016a). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children Sabine. *Developmental Review*, *8*(5), 583-592. https://doi.org/10.1002/aur.1474.Replication

Doebel, S., & Zelazo, P. D. (2016b). Seeing conflict and engaging control: Experience with contrastive language benefits executive function in preschoolers. *Cognition*, *157*, 219-226. https://doi.org/10.1016/j.cognition.2016.09.010

Duckworth, A. L., & Kern, M. L. (2011). A Meta-Analysis of the Convergent Validity of Self-Control Measures. *Journal of research in personality*, *45*(3), 259-268. https://doi.org/10.1016/j.jrp.2011.02.004

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond Risk and Protective Factors: An Adaptation-Based Approach to Resilience. *Perspectives on Psychological Science*, *12*(4), 561-587. https://doi.org/10.1177/1745691617693054

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472-5477.

Farah, M. J. (2017). The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences. *Neuron*, *96*(1), 56-71. https://doi.org/10.1016/j.neuron.2017.08.034

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *Journal of Educational Psychology*, *107*(1), 207-221. https://doi.org/10.1037/a0037366

Garon, N. M., Piccinin, C., & Smith, I. M. (2016). Does the BRIEF-P Predict Specific Executive Function Components in Preschoolers? *Applied Neuropsychology. Child*, *5*(2), 110-118. https://doi.org/10.1080/21622965.2014.1002923

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129-153. https://doi.org/10.1016/0010-0277(94)90068-X

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Behavior Rating Inventory of Executive Function. *Child Neuropsychology*, *6*(3), 235-238. https://doi.org/10.1076/chin.6.3.235.3152

Giovannetti, F., Pietto, M. L., Segretín, M. S., & Lipina, S. J. (2020). Impact of an Individualized Cognitive Training Intervention in Preschoolers from Poor Homes. *International Journal of Environmental Research and Public Health*, *17*(8), 2912. https://doi.org/10.3390/ijerph17082912

Goldin, A. P., Hermida, M. J., Shalom, D. E., Elias Costa, M., Lopez-Rosenfeld, M., Segretin, M. S., Fernandez-Slezak, D., Lipina, S. J., & Sigman, M. (2014). Far transfer to language and math of a short software-based gaming intervention-Supplementary material. *Proceedings of the National Academy of Sciences*, *111*(17), 6443-6448. https://doi.org/10.1073/pnas.1320217111

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). *Socioeconomic status and executive function: Developmental trajectories and mediation*. 1-17. https://doi.org/10.1111/desc.12246

Hermida, M. J., Segretin, M. S., Prats, L. M., Fracchia, C. S., Colombo, J. A., & Lipina, S. J. (2015). Cognitive neuroscience, developmental psychology, and education: Interdisciplinary development of an intervention for low socioeconomic status kindergarten children. *Trends in Neuroscience and Education*, *4*(1), 15-25. https://doi.org/10.1016/j.tine.2015.03.003

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology*, *72*(1), 319-345. https://doi.org/10.1146/annurev-psych-061020-105721

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behaviour. *Developmental Neuropsychology*, *26*(1), 403-422.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147-1185. https://doi.org/10.1037/bul0000160

Koomen, R., Grueneisen, S., & Herrmann, E. (2020). Children Delay Gratification for Cooperative Ends. *Psychological Science*, *31*(2), 139-148. https://doi.org/10.1177/0956797619894205

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), e12529. https://doi.org/10.1111/desc.12529

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59-80.https://doi.org/10.1348/026151003321164627

Lenth, R. (2020). *Mmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0.* https://CRAN.R-project.org/package=emmeans

Lerner, R. M. (2006). Developmental Science, Developmental Systems, and Contemporary Theories of Human Development. En *Handbook of child psychology: Theoretical models of human development, Vol. 1, 6th ed* (pp. 1-17). John Wiley & Sons Inc.

Lüdecke D. (2021). *sjstats: Statistical Functions for Regression Models (Version 0.18.1)*. https://CRAN.R-project.org/package=sjstats

Ma, F., Zeng, D., Xu, F., Compton, B. J., & Heyman, G. D. (2020). Delay of Gratification as Reputation Management. *Psychological Science*, *31*(9), 1174-1182. https://doi.org/10.1177/0956797620939940

Mahone, E. M., & Hoffman, J. (2007). Behavior Ratings of Executive Function among Preschoolers with ADHD. *The Clinical Neuropsychologist*, *21*(4), 569-586. https://doi.org/10.1080/13854040600762724

Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2019). «Same but different»: Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and Social Psychology*, *117*(6), 1164-1188. https://doi.org/10.1037/pspp0000224

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and Classroom Characteristics Associated With Teachers' Ratings of Prekindergartners' Relationships and Behaviors. *Journal of Psychoeducational Assessment*, *24*(4), 367-380. https://doi.org/10.1177/0734282906290594

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). *The Unity and Diversity of Executive Functions and Their Contributions to Complex " Frontal Lobe " Tasks: A Latent Variable Analysis and*. *100*, 49-100. https://doi.org/10.1006/cogp.1999.0734

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885-958. https://doi.org/10.1037/bul0000153

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* URL https://www.R-project.org/

Ready, D., & Wright, D. (2011). Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities: The Role of Child Background and Classroom Context. *American Educational Research Journal - AMER EDUC RES J*, *48*, 335-360. https://doi.org/10.3102/0002831210374874

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2017). *Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian analysis*. PsyArXiv. https://doi.org/10.31234/osf.io/bxfsu

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions:Biological Sciences*, *298*, 199-209.

Sherman, E. M. S., & Brooks, B. L. (2010). Behavior Rating Inventory of Executive Function – Preschool Version (BRIEF-P): Test Review and Clinical Guidelines for Use. *Child Neuropsychology*, *16*(5), 503-519. https://doi.org/10.1080/09297041003679344

Tamm, L., & Peugh, J. (2019). Concordance of teacher-rated and performance-based measures of executive functioning in preschoolers. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *25*(3), 410-424. https://doi.org/10.1080/09297049.2018.1484085

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 131-143. https://doi.org/10.1111/jcpp.12001

Torff, B., & Sessions, D. (2009). Teachers' attitudes about professional development in high-SES and low-SES communities. *Learning Inquiry*, *3*(2), 67-77. https://doi.org/10.1007/s11519-009-0040-1

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, *5*(MAR), 1-9. https://doi.org/10.3389/fpsyg.2014.00213

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: Typical developmental trends. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *44*(4), 561-575. https://doi.org/10.1111/1469-7610.00145

Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review*, 612-624. https://doi.org/10.1016/j.dr.2015.07.001Zelazo, P. D. (2020). Executive Function and Psychopathology: A Neurodevelopmental Perspective. *Annual Review of Clinical Psychology*, *16*(1), 431-454. https://doi.org/10.1146/annurev-clinpsy-072319-024242

Taza decorativa de Alfabetización inicial
Early literacy's decorative mug

Autores: **Soledad Assis, Camila Zugarramurdi** y **Juan Valle-Lisboa.** Financiación: CSIC- Iniciación: *Adaptación y evaluación de una aplicación de juegos para la alfabetización inicial.* Universidad de la República, Uruguay. Este proyecto busca optimizar el desarrollo de una herramienta educativa basada en evidencia científica y centrada en la intervención escolar para facilitar la enseñanza de la lectura.

Authors: **Soledad Assis, Camila Zugarramurdi** and **Juan Valle-Lisboa**. Funding: CSIC- Iniciación: Adaptación y evaluación de una aplicación de juegos para la alfabetización inicial. Universidad de la República, Uruguay. This project seeks to optimize the development of an educational tool based on scientific evidence and focused on school intervention to facilitate the teaching of reading.

Approximate number and approximate time discrimination each correlate with school math abilities in young children

Darko Odic[1], Juan Valle Lisboa[2], Robert Eisinger[3], Magdalena Gonzalez Olivera[1], Alejandro Maiche[1], & Justin Halberda[3]

1 University of British Columbia. Department of Psychology, Vancouver, Canada
2 **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República. Uruguay
3 Johns Hopkins University. Department of Psychology and Brain Sciences. Baltimore, USA

Darko Odic https://orcid.org/0000-0001-5783-2789, Juan Valle Lisboa https://orcid.org/0000-0002-5662-5937, Robert Eisinger https://orcid.org/0000-0003-3191-2355, Magdalena Gonzalez Olivera https://orcid.org/0000-0001-5718-8217, Alejandro Maiche https://orcid.org/0000-0002-5006-1544, Justin Halberda https://orcid.org/0000-0002-5619-3249

Correspondence concerning this manuscript should be addressed to Darko Odic Department of Psychology. University of British Columbia 2136 West Mall. Vancouver, British Columbia V6T 1Z4. Email :**darko.odic@psych.ubc.ca**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Odick, D., Valle Lisboa, J., Eisinger, R., Gonzalez, M, Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychologica*, vol. 163 pp 17-26. Available at **https://doi.org/10.1016/j.actpsy.2015.10.010**

Abstract

What is the relationship between our intuitive sense of number (e.g., when estimating how many marbles are in a jar), and our intuitive sense of other quantities, including time (e.g., when estimating how long it has been since we last ate breakfast)? Recent work in cognitive, developmental, comparative psychology, and computational neuroscience has suggested that our representations of approximate number, time, and spatial extent are fundamentally linked and constitute a "generalized magnitude system". But, the shared behavioral and neural signatures between number, time, and space may alternatively be due to similar encoding and decision making processes, rather than due to shared domain-general representations. In this study, we investigate the relationship between approximate number and time in a large sample of 6-8 year old children in Uruguay by examining how individual differences in the precision of number and time estimation correlate with school mathematics performance. Over four testing days, each child completed an approximate number discrimination task, an approximate time discrimination task, a digit span task, and a large battery of symbolic math tests. We replicate previous reports showing that symbolic math abilities correlate with approximate number precision and extend those findings by showing that math abilities also correlate with approximate time precision. But, contrary to approximate number and time sharing common representations, we find that each of these dimensions uniquely correlates with formal math: approximate number correlates more strongly with formal math compared to time and continues to correlate with math even when precision in time and individual differences in working memory are controlled for. These results suggest that there are important differences in the mental representations of approximate number and approximate time and further clarify the relationship between quantity representations and mathematics.

Keywords: approximate number system, time discrimination, formal mathematics, individual differences

Introduction

What is the source of our intuitions about number? Recent work in cognitive development has focused on young children's ability to quickly and intuitively represent the number of items in a collection through the Approximate Number System (ANS; Dehaene, 2009; Feigenson, Dehaene, & Spelke, 2004; Halberda & Feigenson, 2008; Halberda, Mazzocco, & Feigenson, 2008; Halberda & Odic, 2014; Odic, Hock, & Halberda, 2014; Odic, Libertus, Feigenson, & Halberda, 2013). The multimodal ANS provides us with a rough and noisy sense of number, such as when guessing how many people are sitting in a lecture hall or how many items are in our shopping basket.

The ANS is characterized by three empirical signatures (Feigenson et al., 2004; Halberda & Odic, 2014). First, discrimination performance in the ANS is ratio-dependent (i.e., obeys Weber's law): discriminating a collection of 10 items from 5 items (a ratio of 2.0) is much easier than discriminating a collection of 10 items from 9 items (a ratio of 1.11). The precision with which an individual can successfully discriminate difficult ratios is often quantified through the Weber fraction (w), and theoretically corresponds to the amount of noise in the underlying ANS representations (Cordes, Gallistel, Gelman, & Latham, 2007; Halberda & Odic, 2014; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Second, there are large individual differences in ANS precision, and children's ANS continues to improve from birth onward, peaking around age 30 (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Odic et al., 2013; Piazza et al., 2010). Finally, the ANS has been localized in both the human brain and in non-human animals to a region of the intraparietal sulcus (IPS; Dehaene, Piazza, Pinel, & Cohen, 2003; Nieder, 2005, 2012; Piazza et al., 2010; Roitman, Brannon, & Platt, 2007); physiological modulations of the IPS can, for example, enhance ANS discrimination (Cappelletti et al., 2013).

Researchers have also focused on the relationship between the ANS and formal mathematical abilities[21]. Individual differences in ANS precision show a small but significant relationship with formal math, including in preschoolers (Feigenson, Libertus, & Halberda, 2013; Libertus, Feigenson, & Halberda, 2011; Starr, Libertus, & Brannon, 2013) and adults (DeWind & Brannon, 2012; Libertus, Odic, & Halberda, 2012; Lyons & Beilock, 2011). Temporary modulations of the ANS can also selectively enhance or impair subsequent math performance (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013; Wang, Odic, Halberda, & Feigenson, under review). Finally, individuals with math learning disabilities also show impaired ANS precision (Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010). This work, though not unchallenged (e.g., De Smedt, Noël, Gilmore, & Ansari, 2013), suggests that our basic intuitions about math may emerge, in part, from a universal and ontologically ancient core cognitive system, and that intervention methods that improve the ANS may also help children in acquiring formal math concepts.

But the ANS is not alone in showing these behavioral and neural signatures. Many other dimensions, including surface area, time, density, weight, brightness, line length, etc., also obey Weber's law (Cantlon, Platt, & Brannon, 2009; Cheng, Srinivasan, & Zhang, 1999a; Feigenson, 2007; Gescheider, 1997; Meck & Church, 1983; Möhring, Libertus, & Bertin, 2012; Stone & osley, 1965), develop with age (Brannon, Lutz, & Cordes, 2006; Droit-Volet, Clément, & Fayol, 2008; Odic, Le Corre, & Halberda, 2015; Odic et al., 2013), and are localized in the IPS (Cantlon et al., 2009; Castelli, Glaser, & Butterworth, 2006; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Tudusciuc & Nieder, 2007). For example, transcranial noise stimulation of the IPS modulates both number and time discrimination (Cappelletti et al., 2013), and 6-month-old infant's Weber fractions for surface area discrimination appear identical to Weber fractions for number and time discrimination (Brannon et al., 2006; Feigenson, 2007). These similarities between distinct dimensions have led many

2 In this paper, we refer to "formal" math in the sense of symbolic, abstract, school-taught mathematics, rather than differentiating between more informal math skills, such as addition and counting, and more formal math skills, such as word problems (e.g., see Libertus, Feigenson, & Halberda, 2013)

researchers to suggest that number, time, and space are all represented by common mechanisms – a domain-general "generalized magnitude system" (Bueti & Walsh, 2009; Cantlon et al., 2009; Lourenco & Longo, 2010; Vicario, 2013; Walsh, 2003). Additional evidence for the generalized magnitude system comes from correlations of Weber fractions across dimensions (e.g., time and number; Meck & Church, 1983, but see Droit Volet et al., 2008), and from persistent congruency and interference effects between quantities, whereby manipulation of one dimension affects discrimination performance of another (Barth, 2008; Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Gebuis & Reynvoet, 2012; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Leibovich & Henik, 2013; Lourenco & Longo, 2010; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013; Wood, Willmes, Nuerk, & Fischer, 2008).

Although researchers have frequently invoked the generalized magnitude system as an explanation for the behavioural and neural commonalities amongst quantity representations, it remains unclear what the shared mechanism between number and other dimensions might be. There are at least three (non-mutually exclusive) possibilities. First, quantity representations could share low-level sensory encoding processes. Dakin and colleagues (2011), for example, suggest that number and density are both encoded through low spatial-frequency filters; hence, modulations of density (and thus of low spatial-frequency) will simultaneously impact number discrimination. Second, various dimensions might all be represented on an identical domain general quantity scale and by identical sets of neurons that code for "more" or "less" of any and every dimension (Bueti & Walsh, 2009; Lourenco & Longo, 2010; Tudusciuc & Nieder, 2007); in this case, representations for, e.g., time and number, will show identical Weber fractions, identical individual and developmental differences, and will be equivalently impacted by any modulation of the IPS. Finally, different dimensions may share common decision making or comparison computations, such as determining a threshold before the response is initiated; as a result, quantity representations may compete for behavioural responses and interfere with one another (DeWind & Brannon, 2012; Hurewitz et al., 2006; Van Opstal, Gevers, De Moor, & Verguts, 2008a) and bottlenecks on attentional, memory, or decision making processes may result in similar Weber fractions across dimensions. Droit-Volet and colleagues (2008), for example, find that time and number Weber fractions only correlate when both dimensions are presented sequentially, suggesting that attentional and memory processes may be responsible for their correlation.

The existing evidence has not determined the best explanation for the common behavioral and neural signatures between number, time, and space (though most researchers seem to prefer the shared representations account). Recently, an increasing number of studies have attempted to dissociate quantity representations by examining how they relate to other cognitive abilities, such as affect (Droit-Volet, 2013; Young & Cordes, 2013) or formal mathematics (DeWind & Brannon, 2012; Lourenco, Bonny, Fernandez, & Rao, 2012). If, for example, the ANS correlates with formal math independently from non-numeric dimensions such as surface area, then we would have evidence for an important degree of independence between these dimensions. DeWind and Brannon (2012) recently found that while number and line-length discrimination correlate in precision in adults, only number correlates with formal math (as assessed by SAT scores). Similarly, Lourenco and colleagues (2012) found that while number and cumulative surface area correlate in precision amongst adults, individual differences in the ANS uniquely correlate with arithmetic math problems, while individual differences in cumulative area precision uniquely correlate with geometric math problems. Combined, this work suggests important distinctions in the representations of number and spatial extent and their relationship to formal math, and further implies that the commonality between these dimensions is unlikely to be due to both number and spatial extent being represented on an identical scale.

A similar kind of approach has been used to differentiate the ANS from approximate time perception. Time perception provides a useful case-study because its relationship to the ANS is still very actively debated. Meck and Church (1983) famously proposed that both time and number are encoded by an accumulating pacemaker mechanism, and found that amphetamine administration equally affects time and number perception in rats. Furthermore, affective stimuli, such as sad or happy faces, appear to impact both time and number equally (Droit-Volet, 2013), there are known shared neural substrates for time and number perception (Dormal, Dormal, Joassin, & Pesenti, 2012), and these dimensions show mapping and interference effects (Bueti & Walsh, 2009; Müller & Schwarz, 2008; Oliveri et al., 2008). Focusing on children and adults with math learning disabilities, previous work has shown mixed results in dissociating time from number perception. For example, Cappelletti, Freeman, and Butterworth (2011) find that time perception is not affected in adults diagnosed with dyscalculia. On the other hand, both Hurks and Loosbroek (2012) and Vicario and colleagues (2012) find that children with math learning disabilities show abnormal time estimation and production. Combined, the existing work does not conclusively show evidence for or against time and number being part of a single generalized magnitude system.

The existing work on time, number, and their relationship to formal mathematics leaves open the possibility that quantity representations diverge and differentiate with development, especially as children acquire formal math concepts from preschool onward. Additionally, previous work has only tested children with math learning disabilities and used small sample sizes. Here, we examine the relationship between the ANS, time perception, and a series of formal math tests in a large sample of children tested at schools in Uruguay. By examining the relationship between ANS, time discrimination, and math performance, we ask whether time and number share a common relationship with formal mathematics performance or two distinct relationships. If time and number pattern together, this would be strong evidence for them sharing a common representational resource. If they pattern separately (e.g., if number correlates more strongly with formal math and continues to correlate even after time performance is controlled for), this would be strong evidence for time and number relying on distinct representations. In short, by investigating the relationships between approximate number, approximate time and formal mathematics we can gain insight into the representational resources that support each of these abilities.

Methods

Participants The present study was performed as part of Plan Ceibal, a Uruguayan initiative whereby each student in the public education system receives a technological device (e.g., computer, Tablet, etc.) for use in the classroom. The plan promotes digital inclusion and creativity in learning. We had access to 10 different schools and 31 different classrooms in Montevideo, the Uruguayan capital and its largest city. Within each classroom, each child received a tablet (described below) and, on the days that they were in the classroom and willing, participated in up to ten days of games and/or assessments. For the purposes of this paper, we only report data from the first four days (data from other days will appear in a series of other publications that address questions beyond number and time).

In total, we tested 503 unique first graders (Mean Age = 7.25; SD = 0.46; Age Range = 6.42 – 8.76). However, due to the opportunistic sampling, not all children completed all of the games. We include in our current sample only children who completed both the ANS and Time Discrimination games (N = 244; Mean Age = 7.26; SD = 0.47; Age Range = 6.42 – 8.71; 135 girls and 109 boys); analyses reported in the Results section show no significant differences between children who completed and did not complete both of these two games. Table 1 reports the Ns across the different formal math games that were used for all our statistical analyses. All children spoke Spanish as their first language and all tasks were administered in Spanish.

General Procedure and Apparatus All testing was done in the child's classroom by trained researchers who followed a written protocol. In all, we tested children on seven games across four days: the Prueba Uruguaya de Matemática (PUMA) (days 1 and 2), Time Discrimination (day 1), Digit Span (day 2), Timed Arithmetic (day 3), Symbolic Magnitude Judgement (day 3), Symbolic Ordinal Judgement (day 3), and the ANS Discrimination game (day 4). These tasks are described in more detail below.

The games were played online in Spanish on a 7.6" x 4.65" x 0.39" XO Tablet running the Android 4.2 (Jelly Bean) with a 7" screen size. XO Tablets are hand-held, touch-responsive Tablets specifically designed for children aged 4-14 (see Figure 1). Each participant had their own XO Tablet that they used throughout the duration of testing. All games were completely computerized and were developed using JavaScript, PHP, SQL, and JQuery. For games that involved sound, we provided each child with headphones to wear.

The Tablet was locked with game-specific passwords that children could not begin playing without the trained researcher informing them of the password. On the day of testing, the researcher would read a pre-written script explaining to children the games for that day and encouraging them to do their best. Games were explained on a blackboard to the whole class. Then, for each game, children were instructed to type in a specific password that would trigger the game. In this way, each game stayed paused until the researcher fully read the instructions and explained the game on the blackboard, assuring that children would not being playing the games early. While children played, the researcher walked around the classroom and provided any assistance.

Figure 1: An illustration of the XO Tablet used for all the seven tasks, the ANS Discrimination task, the Time Discrimination task, and the Digit Span task. In the Time Discrimination task, the two monsters would open their mouth and make a signing sound. In the Digit Span task, the child would have to re-create the set-size (here, set-size two is shown) after the numbers disappear.

ANS Discrimination (Panamath).

Children's ANS precision was assessed via the standardized Panamath task (Halberda & Ly, in prep) on the fourth day. As illustrated in Figure 1, children were shown two empty rectangles (a yellow one on the left, and a blue one on the right); subsequently, yellow dots appeared in the yellow rectangle and blue dots in the blue rectangle. The dots disappeared after 1600 ms. Children had to decide which side had more dots and indicated their response by tapping on the appropriate rectangle. We presented five ratios: 1.17 (e.g., 7 blue vs.6 yellow dots), 1.2, 1.5, 2.0, and 3.0. The number of dots within each rectangle was always between 4 and 24. To control for surface area, half of the trials had cumulative surface area congruent with the number of dots (i.e., cumulative area and number gave identical answers) and half the trials were incongruent with the number of dots (i.e., cumulative area gave the opposite answer to that of number).

The first three trials acted as practice and were very easy. Children were allowed to play for 6 minutes. During the task, children were given feedback (a positive "ding!" sound for correct or an "err!" sound for incorrect answers). The dependent variable was percent correct across all completed trials.

Time Discrimination Children's time precision was assessed via the Time Discrimination game, which children played on the first day. As illustrated in Figure 1, children were shown two monsters on a screen: a green one on the left and a purple one on the right. On each trial, each monster would take its turn making a singing-like sound for a certain amount of time. To provide additional visual cues, each monster would open its mouth while singing and put its hand over its mouth for a second once it was done. Children had to touch the monster that sang longer.

The first five trials acted as practice and were very easy. Children were presented with eight ratios: 1.20 (e.g., 1200ms sound vs. 1000ms sound), 1.25, 1.50, 1.60, 2.00, 2.40, 2.50, and 3.00. On half the trials, the left monster made sounds first. The singing sound also varied from trial to trial. Children were given feedback (a positive "ding!" sound for correct or an "err!" sound for incorrect answers). The dependent variable was percent correct across all completed trials.

Digit-Span To assess each child's working memory capacity, we administered a child friendly version of the classic digit-span task (Baddeley, 1992). On each trial, children touched the screen to reveal between 1-5 cards (set-size) with a single-digit Arabic digit on each (see Figure 1). The cards were shown simultaneously and stayed face-up on the screen for 1 second per card (i.e., three cards stayed face-up for 3 seconds, five cards for 5 seconds, etc.). After this time period, the cards turned face-down. Subsequently, blank cards appeared underneath the face-down cards and the child had to input the correct identity for each card in the correct order. Trials on which children got the numbers correct but in the wrong order were counted as incorrect. To determine each child's digit-span, we used a staircase method: each child started with set-size of 1, and the set-size increased by one every time the child got two trials correct; if the child got a trial wrong, the set-size would decrease by 1. The span is determined by the maximum set-size that the child successfully completed. The task was limited to 4 minutes.

Formal Math Assessment #1: Prueba Uruguaya de Matemática (PUMA) The PUMA was used to test a broad set of math skills through a series of mini-games, including number symbol knowledge, Arabic number ordering, number composition and decomposition, number line placement, and basic word problems. These were tested within the PUMA using eight different mini-games across two testing days. On each day, children were given up to 10 minutes total to complete as many questions (intermixed from various mini-games) as they could; if they spent more than one minute on any single question, the game automatically advanced to the next one. All children did the trials in the same order. Children received instructions verbally in their headphones via the Tablet after typing in the game-specific password given them by the researcher.

In the first mini-game, children had a set of cards with numbers on the screen (1, 2, 5, 10, 20, 50, and 100). On each trial, the on-screen game character desired a target number (written on the screen), and children had to drag and add several cards to a workspace in order to match the target number (targets were 10, 8, 20, 34, 52, and 100). In the second mini-game children performed the reverse of this challenge – i.e., a game character on-screen showed a collection of cards (e.g., 50, 1, 1, 1) and the child had to select the correct number for the total array from among 3 options (e.g., 53). In the third mini-game ("base ten", Figure 2, top), children had to add cards (each with a value of 10) to match a target value (e.g., 100). Target values were 100, 60, 120, 80, 40, and 30. The fourth mini-game tested number line ordering (Figure 2, middle). Children were shown a train with a number line and some values missing. Children then had to select a card with a number and drag it into the missing spot (the targets were 5, 8, 10, 12, 15, 18, and 20). The fifth mini-game tested verbal counting and cardinality. Children were shown a character and some dots and they had to count the number of dots. They then had to move the character along a line to match the number of the dots (targets were 16, 12, 20, and 9). The sixth mini-game tested number order. Children were shown numbers 10 – 1 in a scrambled order, and had to arrange the cards in decreasing order by dragging them from the bottom of the screen to their correct position on the top. In the seventh mini-game, children saw an empty number line with a single anchor number (either 5, 7, 6, 3, 5, or 8), and had to drag two numbers onto the appropriate position of the line given the anchor (e.g., place 4 and 9 on the line relative to the anchor of 5). Targets were 4-9, 1-9, 4-10, 5-9, 1-7, or 1-6. In the eighth mini-game, testing basic word problems, children saw toys on the screen with labeled prices (Figure 2, bottom). Three math problems indicated the toys to select: (1) "Select two toys totalling $90". (2) "If the girl only has $90, which toy can she not buy?" (3) "Matthew paid $80 and had $2 in chance. What did he buy?"

The dependent measure was percent correct across all completed questions. The PUMA was designed as an overall assessment of math ability, and preliminary analyses revealed insufficient numbers of trials to investigate each game individually. As a result, we combined the percent correct across the eight mini-games; as shown in the Results section, this single score showed a good, normal distribution of scores.

Formal Math Assessment #2: Timed Arithmetic The timed arithmetic game was played the third day and was used to assess children's addition and subtraction abilities. As illustrated in Figure 2, children were shown single Arabic digit subtraction and addition problems (e.g., 4+2 = ?) and had to type their answer on a linear number pad. Problems always involved single-digit operations, but could include decade breaks (i.e., going above 10, 20, etc.). All children saw an identical order of trials. When done, children pressed a green checkmark to register the answer or an eraser to clear the answer and change it. The first two trials were considered practice. Children were given feedback (a positive "ding!" sound for correct or an "err!" sound for incorrect answers). No time limit was enforced for each arithmetic problem. The dependent variable was percent correct across all completed questions.

Formal Math Assessment #3: Symbolic Magnitude Judgement The symbolic magnitude judgement task was also played on the third day and was used to assess children's knowledge of Arabic digits. As illustrated in Figure 2, children were shown two digits (one on each side of the screen) within colored rectangles and had to tap on the one that was numerically larger. The lowest number presented was 1, and the highest number presented was 21. All children saw the same order of trials. Children were allowed to play for 4 minutes, but could spend as much time as they needed on each problem. The first two trials were considered practice. Children were given feedback (a positive "ding!" sound for correct or an "err!" sound for incorrect answers). The dependent variable was percent correct across all completed questions.

Formal Math Assessment #4: Symbolic Ordinal Judgement The symbolic ordinal judgement game was also played on the third day and assessed children's knowledge of number order. As shown in Figure 2, children were shown three single-digit numbers on the screen and had to decide if the numbers were in increasing order (e.g., 3-5-8) or not (e.g., 3-8-5). If the children thought the numbers were increasing, they had to press an upward facing green arrow; if they thought the numbers were not increasing they had to press a red X. The first three trials were considered practice. All children saw the identical order of trials. Children could play for up to 4 minutes, but as long as needed on each trial. Children were given feedback (a positive "ding!" sound for correct or an "err!" sound for incorrect answers). The dependent variable was percent correct across all completed questions.

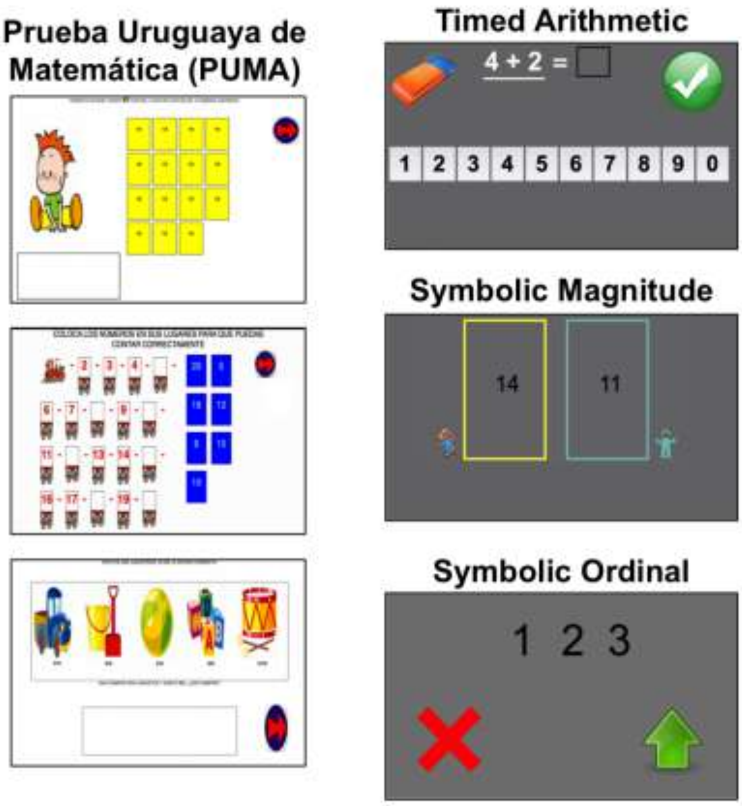


Figure 2: An illustration of the four Formal Math tasks. The PUMA task is comprised of eight mini-games, of which three are shown on the left. Details on each of the eight games are in-text.
Videos outlining each of the eight tasks can be found (in Spanish) online (http://www.ceibal.psico.edu.uy/2013/preparacion/prueba-de-evaluacion-de-matematica/). Examples of three of the mini-games are shown in Figure 2.

Results

Because all testing occurred in the school setting over four days, many children did not complete all seven of the tasks (i.e., ANS Discrimination, Time Discrimination, Digit Span, and the four formal math assessments). Our sample sizes and pairwise correlations for each of the tasks are presented in Table 1.

To make sure that our attrition was random and not because some children were doing poorly on the games and subsequently dropped out, we first examined the scores of children who completed only the ANS Discrimination (on day 4) or only the Time Discrimination game (on day 1) and compared them to children who completed both tasks. We found no significant difference in ANS Discrimination performance between children who completed only the ANS Discrimination game (M = 0.73, SD = 0.11, N = 151) and those who completed both the Time Discrimination and ANS Discrimination games (M = 0.74, SD = 0.09, N = 251; t(400) = -0.54; p = 0.59). Similarly, there was no difference in Time Discrimination performance between children who completed only the Time Discrimination game (M = 0.68, SD = 0.20, N = 162) and those who completed both the ANS Discrimination and Time Discrimination games (M = 0.68, SD = 0.20, N = 251; t(411) = -0.12; p = 0.91).

We next tested whether performance on both ANS and Time Discrimination obeyed Weber's law (i.e., was ratio-dependent). As shown in Figure 3, a one-way Repeated-Measures ANOVA (Ratios: 1.17, 1.25, 1.50, 2.00, 3.00) over ANS Discrimination percent correct showed a significant effect of Ratio (F(4,972) = 347.82, p < .001) and a normal distribution of scores. And, as also shown in Figure 3, a one-way Repeated-Measures ANOVA (Ratios: 1.20, 1.25, 1.50, 1.60, 2.00, 2.40, 2.50, and 3.00) over Time Discrimination percent correct showed a main effect of Ratio (F(7,1050) = 25.93, p < .001), with a slightly left-skewed distribution of scores. Additionally, we found that ANS Discrimination performance (M = 73.7%, SD = 9.37%) was

Table 1. The pairwise correlations between all the tested variables. Stars indicate significance at $p < .05$.

	Discrimination Tasks			Math Tasks			Working Memory
	ANS	*Time*	*PUMA*	*Timed Arithmetic*	*Symbolic Magnitude*	*Symbolic Ordinal*	*Digit Span*
ANS (n = 244)	1.0	.26*	.39*	.37*	.51*	.44*	.21*
Time (n = 244)	--	1.0.	.31*	.25*	.22*	.14	.29*
PUMA (n = 233)	--	--	1.0	.57*	.39*	.49*	.35*
Timed Arithmetic (n = 153)	--	--	--	1.0	.44*	.47*	.35*
Symbolic Magnitude (n = 170)	--	--	--	--	1.0	.46*	.22*
Symbolic Ordinal (n = 119)	--	--	--	--	--	1.0	.37*
Digit Span (n = 183)	--	--	--	--	--	--	1.0

significantly better than Time Discrimination performance (M = 68.83%, SD = 9.37%; t(243) = 3.89; $p < .001$), consistent with previous work (Droit-Volet et al., 2008). We found no significant relationship between age and ANS Discrimination (r(233) = -.07; $p = .29$) or between age and Time Discrimination (r(233) = -.01; $p = .87$), probably due to our truncated age range.

Consistent with previous work, we found a weak but significant correlation between ANS Discrimination and Time Discrimination performance (r(242) = .26; $p < .001$; Figure 3). As discussed in the Introduction, this correlation could be due to shared encoding, identical domain general representations, or decision-making components. To further understand this correlation, we investigated the relationship of both ANS and Time Discrimination to individual differences in Digit Span – the classic measure of working memory performance. Digit Span performance correlated with both ANS Discrimination (r(182) = .21; $p < .01$), and with Time Discrimination Performance (r(182) = .29; $p < .001$), suggesting that working memory contributes to individual differences in each task. But – consistent with the idea that number and time only share decision making components – we failed to find a correlation between ANS and Time Discrimination when Digit Span was controlled for (r(179) = .10; $p = .16$; Figure 3). In other words,the relationship between the ANS and time in our sample appears to be largely accounted for by shared working memory (see also Droit-Volet & Wearden, 2001; Genovesio, Tsujimoto, & Wise, 2012). However, as it is difficult to draw conclusions from null results, we next turn to the relationship between time, number and our formal math assessments.

Each of the individual math tasks ranged in performance from 0 to 100%, but chance performance was different for each task (e.g., for PUMA and Timed Arithmetic, chance was less than 5%, while for Symbolic Magnitude and Symbolic Ordinal was 50%). As shown in Figure 4, most of the individual math tasks were normally distributed, with the exception of the

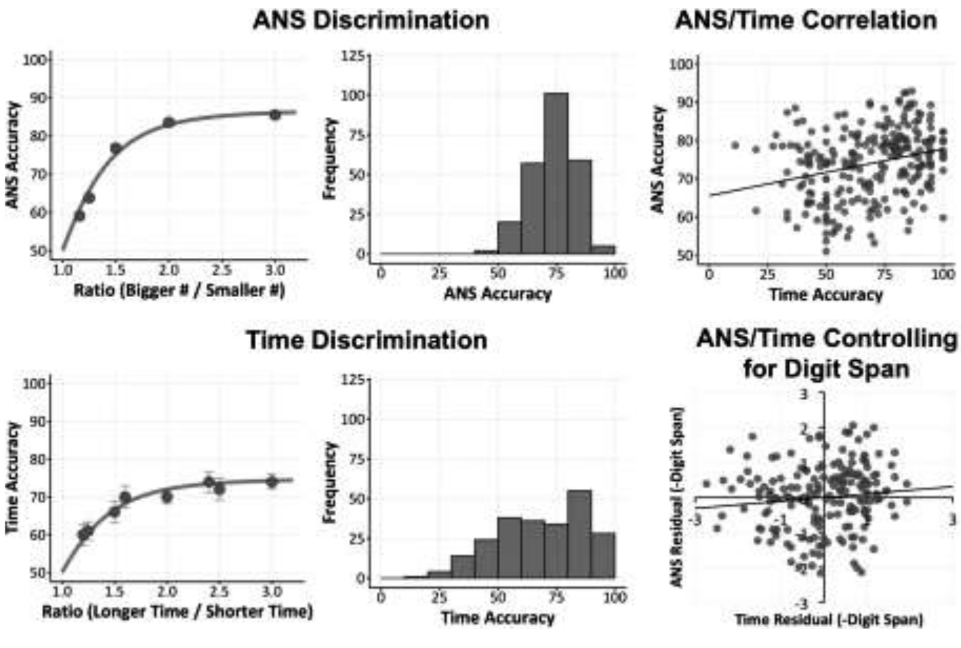


Figure 3: Performance on the ANS Discrimination and Time Discrimination tasks, averaged across all children and as a histogram. Results from both tasks show clear ratio-dependent effects and normal distributions of scores (the curve fits are from a psychophysical model of Halberda and Feigenson, 2008). On the right, the correlation between ANS and Time performance, and the correlation between ANS and Time when controlling for Digit Span.

Symbolic Magnitude task, which showed a strong negative skew and was easier than the other three tasks. The average performance on the PUMA task was 50.5% (SE = 1.4%), on the Timed Arithmetic task 53.1% (SE = 2.0%), on the Symbolic Magnitude task 80.8% (SE = 1.5%), and on the Symbolic Ordinal task 66.2% (SE 1.6%). These distributions show that children understood and engaged with all of the math tasks.

Because of the strong correlations between the four math tasks and in order to maximize our sample size, we calculated a single combined Formal Math Score. To do so, we first normalized performance on each individual math task by re-computing each child's score as a Z Score. Then, to calculate the combined and standardized Formal Math Score, we averaged these individual Z-Scores across all children, ignoring the scores for the tasks they did not complete. This gave us a final sample size of 244 individual children for whom we had a combined Formal Math Score, ANS Discrimination and Time Discrimination. As shown in Figure 5, the Formal Math Score distribution was normal with a mean of -0.05 (SD = 0.88). We found no correlation between age and the Formal Math Score (r(232) = -.03; $p = .69$), probably due to our truncated age range.

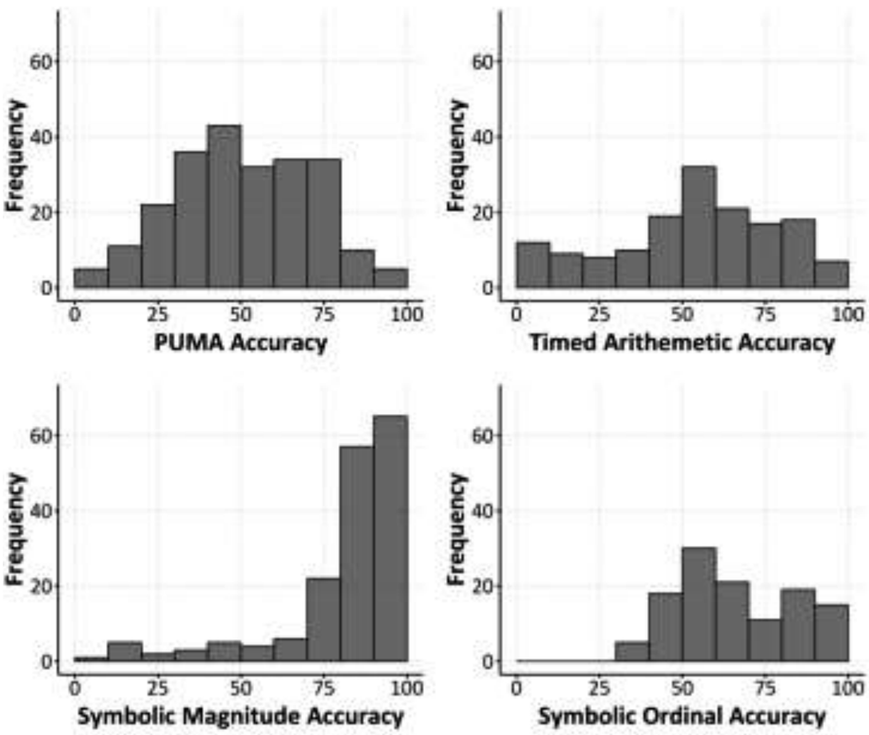


Figure 4: Histograms for each of the four math tasks (PUMA, Timed Arithmetic, Symbolic Magnitude, and Symbolic Ordinal).

Next, we turn to the main question of interest – what is the relationship between the ANS Discrimination and the Formal Math Score and is it in any way different from the relationship between Time Discrimination and the Formal Math Score? We found a moderate correlation between ANS Discrimination performance and the Formal Math Score (r(242) = .51; $p < .001$; Figure 5) and a weak correlation between Time Discrimination performance and the Formal Math Score (r(242) = .29; $p < .001$; Figure 5). This relationship held for each of the individual math tests that comprised the combined Formal Math Score, with the exception of Symbolic Ordinal task, which did not correlate with Time Discrimination (see Table 2). We also found that the correlation between ANS Discrimination and the Formal Math Score was significantly higher than the correlation between Time Discrimination and the Formal Math Score (Z = 3.22; $p < .001$). This difference could, at least in part, be due to the lower reliability of the Time Discrimination scores and does not by itself suggest independence between time and number.

To further understand these results, we performed a set of partial correlations; if time and number are both represented on a common scale by common Gaussian tuning curves, we should find that controlling for one should remove the correlation with the Formal Math Score. However, contrary to this, we found that ANS Discrimination performance still correlated with the Formal Math Score, even when the variability in Time Discrimination was entirely controlled for (r(242) = .47; $p < .001$). We also found the converse Time Discrimination performance still correlated with the Formal Math Score, even when the variability in ANS Discrimination was entirely controlled for (r(242) = .19; $p < .05$). These results held true for all of the individual formal math games, with the exception that Symbolic Magnitude no longer correlated with Time Discrimination when ANS was controlled for (see Table 2). Hence, each dimension has a unique correlation with Formal Math Score.

A potential criticism, however, is that the differences captured by the partial correlations between ANS, time, and formal math tasks are not due to independent representations of time and number, but due to differences between the two discrimination tasks. Specifically, because stimuli in the Time Discrimination task are, unlike the ANS Discrimination task, presented sequentially, performance in this task will highly depend on individual differences in working memory (Droit-Volet & Wearden, 2001). Thus, a correlation between Time Discrimination and the Formal Math Score could remain

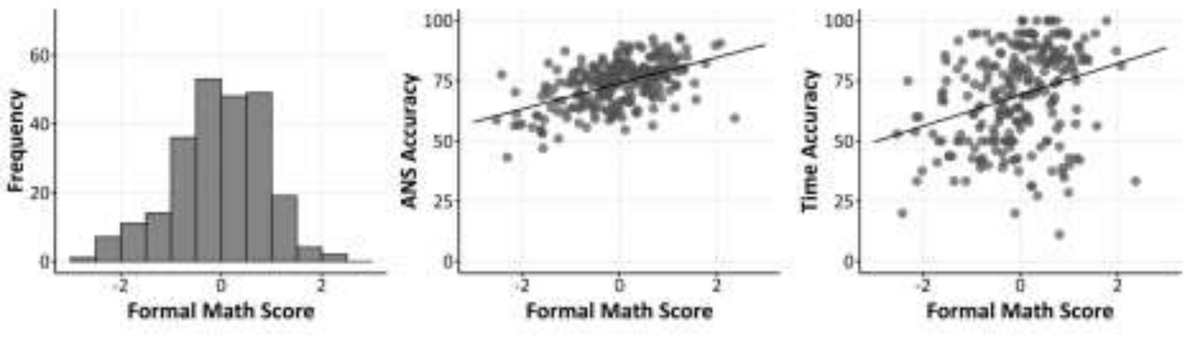


Figure 5: Histogram of the combined Formal Math Score and the correlations between the Formal Math Score and the ANS and Time Discrimination tasks.

Table 2. The data from the different tasks, alongside each task's correlation with ANS Discrimination performance, with the ANS Discrimination performance when controlling for Time Discrimination, correlation with Time Discrimination performance, and correlation with Time Discrimination performance when controlling for ANS discrimination. Stars indicate significance ($p < .05$)

	N	Average (SE)	ANS Discrimination Correlation	ANS Discrimination Controlling for Time	Time Discrimination Correlation	Time Discrimination Controlling for ANS
ANS Discrimination	244	73.4 (0.6)	--	--	.26*	--
Time Discrimination	244	68.8 (1.3)	.26*	--	--	--
Formal Math Score	244	-0.05 (0.06)	.51*	.47*	.29*	.19*
PUMA	233	50.5 (1.4)	.39*	.33*	.31*	.23*
Timed Arithmetic	153	53.1 (2.0)	.37*	.34*	.25*	.18*
Symbolic Magnitude	170	80.8 (1.5)	.51*	.49*	.22*	.14
Symbolic Ordinal	119	66.2 (1.6)	.44*	.42*	.14	.08

significant even if the ANS forms a domain-general magnitude system with time, since the leftover variance could be due to different working memory demands of the sequential vs. simultaneous tasks.

To investigate the possibility that working memory can account for the correlation between Time Discrimination and the Formal Math Score, we further investigated individual differences in the Digit Span task. If ANS and time comprise a generalized magnitude system and the significant partial correlation between Time Discrimination and the Formal Math Score is due to working memory demands, we should find that further controlling for Digit Span performance should eradicate this correlation. Contrary to this prediction, however, we found that Time continued to weakly correlate with Formal Math even when both ANS and Digit Span were partialled out (Figure 6; $r(178) = .18$; $p < .05$). Similarly, the ANS continued to moderately correlate with Formal math even when both Time and Digit Span were partialled out (Figure 6; $r(178) = .43$; $p < .001$). In other words, ANS and Time uniquely and independently correlated with Formal Math performance, even when individual differences in working memory were accounted for.

General Discussion

While previous work has largely focused on the commonalities between time and number discrimination, including shared encoding and similar Weber fractions, it has remained unclear just how similar these representations really are. Under many popular accounts of the generalized magnitude system, time and number are represented on an identical, domain-general scale that codes for "more" or "less" of any quantity; under extreme version of such a view, time and number should be near-perfectly correlated and show an identical relationship to other cognitive abilities, including formal mathematics. Contrary to this account, however, our findings show that the ANS and time perception do not correlate in preschoolers when individual differences in working memory are controlled for, and that ANS and time uniquely and independently correlate with school math abilities. Furthermore, the independence between ANS and time perception is unlikely to be caused by task-related differences (e.g., by time being presented sequentially), as ANS and time continue to correlate with formal math performance when working memory is controlled for.

Our results are most consistent with the idea that the scale that codes approximate number is distinct from the scale that codes for approximate time. In other words, while time and number might share encoding and decision-making components – including working memory demands - the evidence presented

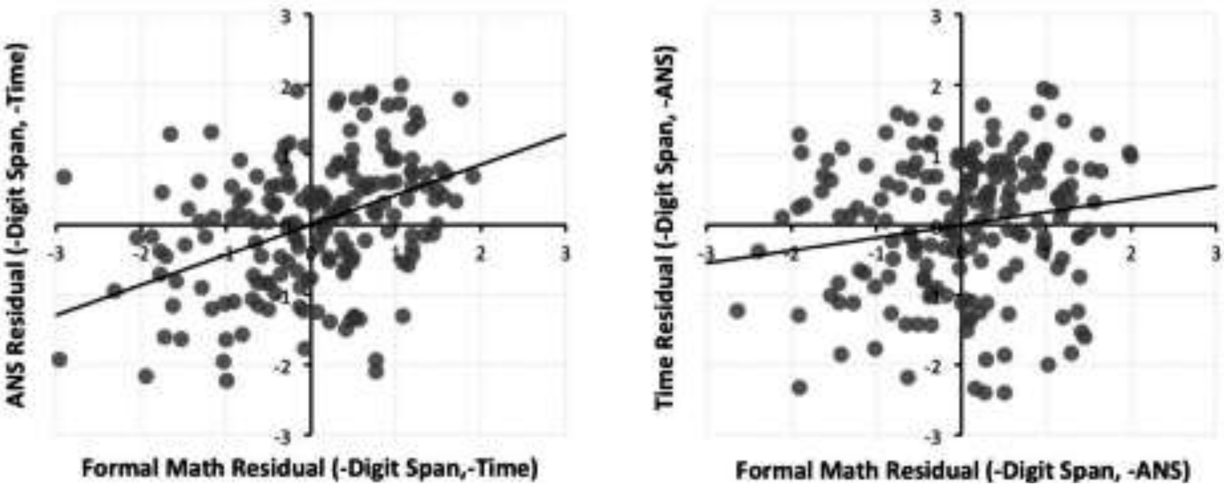


Figure 6: Partial correlations between the ANS and the Formal Math Score (controlling for Digit Span and Time) and between Time and the Formal Math Score (controlling for Digit Span and ANS). The units are standardized residuals.

here suggests that they do not share identical, domain-general representations. Thus, our results suggest that approximate magnitude representations encompass a constellation of abilities rather than a single system. That is, while time and number discrimination are related, and they each relate to school math ability, these relations are varied and textured rather than identical or reducible to a single construct. This is consistent with other findings in the literature that have found differences in the neural localization for time and number (Dormal, Andres, & Pesenti, 2008), and findings that suggest that the interference and mapping effects between time and number are the product of a metaphorical mapping between them, rather than shared resources (Bottini & Casasanto, 2013; Vicario & Martino, 2010). More broadly, they are consistent with recent critiques of the generalized magnitude system, including work dissociating number and surface area (Odic et al., 2013), and work suggesting that many of the commonalities between quantity dimensions may stem from shared decision-making components (Van Opstal et al., 2008a; Van Opstal & Verguts, 2013).

Our findings are broadly consistent with previous work on the ANS, non-numeric dimensions and formal math. For example, much like in the case of spatial extent in adults (DeWind & Brannon, 2012; Lourenco et al., 2012), we find that the ANS uniquely correlates with formal math compared to time. Similarly, our results are consistent with the finding that adults with dyscalculia have unimpaired time perception (Cappelletti et al., 2011). Our data also extend this work by demonstrating a dissociation between number and time in children, removing the possibility that this division occur only after years of experience and formal math education.

At the same time, our results are not entirely consistent with some previous work showing that time estimation and production are impaired in children with math learning disabilities (Hurks & Loosbroek, 2012; Vicario et al., 2012). One possibility is that the specific nature of the tasks may be responsible: in time estimation and production tasks, participants are asked to assign a numeric value to a duration, or are given a numeric value and are asked to produce a duration of that length. In both of these situations, children must use their knowledge of numbers to perform a task about time. Hence, if their math learning disability had impacted their number knowledge broadly, then we might expect them to be impaired in any task that used numbers, and numerical magnitudes, as stimuli (Cappelletti et al., 2011; Hurks & Loosbroek, 2012; Vicario & Martino, 2010). Furthermore, Vicario and colleagues (2012) failed to find a correlation between time performance and tests of formal math in their sample; however, without a positive correlation between formal math and the ANS (which they did not test) it is hard to interpret this null result.

Perhaps surprisingly, we found a robust and independent correlation between formal math and the time discrimination performance, even when ANS discrimination performance was controlled for. Although this correlation was consistently weaker than the correlations with the ANS (across the Formal Math Score and the four individual games), it still remains to be explained: why are formal math and time discrimination performance correlated? One possibility is that this correlation is an artifact of a third factor, such as children's ability to pay attention or their executive functions. This, however, seems unlikely given that time continued to correlate with formal math even when digit span performance was controlled for. Alternatively, one's intuitive sense of time may contribute to only certain math abilities (e.g., perhaps providing an analogy to concepts such as multiplication). Future work could investigate whether temporary modulation of time precision has an impact on formal math performance (temporary modulations of the ANS have been shown to selectively impact subsequent math tasks; Wang et al., under review).

In conclusion, the work presented here demonstrates a unique relationship between formal mathematics, the ANS, and time discrimination performance. This suggests that the commonalities between the ANS and time are not due to shared representations on a domain general "more/less" scale, but more likely due to shared encoding or decision-making processes; and that time and ANS may each independently relate to formal math abilities. This helps to clarify the nature of the theorized generalized-magnitude system and has implications for future work on mathematics interventions.

Acknowledgements

This work was supported by Centro Ceibal para el Apoyo a la Educación de la Niñez y la Adolescencia. We thank Alvaro Mailhos and Mario Luzardo for helpful discussions. Thank you also to Diego Cuevasanta, Cecilia Hontou, Gonzalo Grau, and Leticia Carve for their help in the first 10 days of the intervention. We also thank the Principals, Teachers and Staff of the participant schools for their help during the study.

References

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Barth, H. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition*, *109*(2), 251–266. http://doi.org/10.1016/j.cognition.2008.09.002

Bottini, R., & Casasanto, D. (2013). Space and time in the child's mind: metaphoric or ATOMic? *Cognition*, *4*, 803. http://doi.org/10.3389/fpsyg.2013.00803

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science*, *9*(6), F59–F64. http://doi.org/10.1111/j.1467-7687.2006.00530.x

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1831–1840. http://doi.org/10.1098/rstb.2009.0028

Cantlon, J. F., Platt, M., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. http://doi.org/10.1016/j.tics.2008.11.007

Cappelletti, M., Freeman, E. D., & Butterworth, B. L. (2011). Time Processing in Dyscalculia. *Frontiers in Psychology*, *2*. http://doi.org/10.3389/fpsyg.2011.00364

Cappelletti, M., Gessaroli, E., Hithersay, R., Mitolo, M., Didino, D., Kanai, R., … Walsh, V. (2013). Transfer of Cognitive Training across Magnitude Dimensions Achieved with Concurrent Brain Stimulation of the Parietal Lobe. *The Journal of Neuroscience*, *33*(37), 14899–14907. http://doi.org/10.1523/JNEUROSCI.1692-13.2013

Castelli, F., Glaser, D., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences*, *103*(12), 4693–4698. http://doi.org/10.1073/pnas.0600444103

Cheng, K., Srinivasan, M. V., & Zhang, S. W. (1999a). Error is proportional to distance measured by honeybees: Weber's law in the odometer. *Animal Cognition*, *2*(1), 11–16. http://doi.org/10.1007/s100710050020

Cheng, K., Srinivasan, M. V., & Zhang, S. W. (1999b). Error is proportional to distance measured by honeybees: Weber's law in the odometer. *Animal Cognition*, *2*(1), 11–16.

Cordes, S., Gallistel, C. R., Gelman, R., & Latham, P. (2007). Nonverbal arithmetic in humans: Light from noise. *Attention, Perception, & Psychophysics*, *69*(7), 1185–1203. http://doi.org/10.3758/BF03193955

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A Common Visual Metric for Approximate Number and Density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552–19557. http://doi.org/10.1073/pnas.1113195108

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, *1156*(1), 232–259.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3-6), 487–506.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3329901/

Dormal, V., Andres, M., & Pesenti, M. (2008). Dissociation of numerosity and duration processing in the left intraparietal sulcus: A transcranial magnetic stimulation study. *Cortex*, *44*(4), 462–469. http://doi.org/10.1016/j.cortex.2007.08.011

Dormal, V., Dormal, G., Joassin, F., & Pesenti, M. (2012). A common right fronto-parietal network for numerosity and duration processing: An fMRI study. *Human Brain Mapping*, *33*(6), 1490–1501. http://doi.org/10.1002/hbm.21300

Droit-Volet, S. (2013). Emotion and magnitude perception: number and length bisection. *Frontiers in Neurorobotics*, *7*. http://doi.org/10.3389/fnbot.2013.00024

Droit-Volet, S., Clément, A., & Fayol, M. (2008). Time, number and length: Similarities and differences in discrimination in adults and children. *Quarterly Journal of Experimental Psychology*, *61*(12), 1827–46.

Droit-Volet, S., & Wearden, J. H. (2001). Temporal Bisection in Children. *Journal of Experimental Child Psychology*, *80*(2), 142–159. http://doi.org/10.1006/jecp.2001.2631

Feigenson, L. (2007). The equality of quantity. *Trends in Cognitive Sciences*, *11*(5), 185–187. http://doi.org/10.1016/j.tics.2007.01.006

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. http://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/cdep.12019/full

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642.

Genovesio, A., Tsujimoto, S., & Wise, S. P. (2012). Encoding goals but not abstract magnitude in the primate prefrontal cortex. *Neuron*, *74*(4), 656–662.

Gescheider, G. A. (1997). *Psychophysics: The Fundimentals* (3rd ed.). Mahwah: Lawrence Erblaum Associates.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. http://doi.org/10.1037/a0012682

Halberda, J., & Ly, R. (in prep). PANAmath: The psychophysical assessment of number-sense acuity. *Manuscript in Preparation*.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–11120. http://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. http://doi.org/10.1038/nature07246

Halberda, J., & Odic, D. (2014). The Precision and Internal Confidence of Our Approximate Number Thoughts. In D. C. Geary, D. Berch, & K. Koepke (Eds.), *Evolutionary Origins and Early Development of Number Processing* (p. 305). Academic Press.

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*(2), 77–96.

Hurks, P. P. M., & Loosbroek, E. van. (2012). Time Estimation Deficits in Childhood Mathematics Difficulties. *Journal of Learning Disabilities*, 0022219412468161. http://doi.org/10.1177/0022219412468161

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107. Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Cognition*, *4*, 375. http://doi.org/10.3389/fpsyg.2013.00375

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300.

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829–838.

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive Sense of Number Correlates With Math Scores on College-Entrance Examination. *Acta Psychologica*, *141*(3), 373–379.

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences*, *109*(46), 18737–18742. http://doi.org/10.1073/pnas.1207212109

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873–881.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number sense and arithmetic competence. *Cognition*, *121*(2), 256–261.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia). *Child Development*, *82*(4), 1224–1237.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334.

Möhring, W., Libertus, M. E., & Bertin, E. (2012). Speed discrimination in 6- and 10-month-old infants follows Weber’s law. *Journal of Experimental Child Psychology*, *111*(3), 405–418. http://doi.org/10.1016/j.jecp.2011.11.002

Müller, D., & Schwarz, W. (2008). “1-2-3”: Is There a Temporal Number Line?: Evidence from a Serial Comparison Task. *Experimental Psychology (formerly “Zeitschrift Für Experimentelle Psychologie”)*, *55*(3), 143–150. http://doi.org/10.1027/1618-3169.55.3.143

Nieder, A. (2005). Counting on neurons: the neurobiology of numerical competence. *Nature Reviews. Neuroscience*, *6*(3), 177.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *109*(29), 11860–11865.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis Affects Approximate Number Discrimination in Young Children. *Journal of Experimental Psychology: General*, *43*(1), 255–265. http://doi.org/10.1037/a0030825

Odic, D., Le Corre, M., & Halberda, J. (2015). Children’s mappings between number words and the approximate number system. *Cognition*, (138), 102–121.

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental Change in the Acuity of Approximate Number and Area Representations. *Developmental Psychology*, *49*(6), 1103–1112.

Oliveri, M., Vicario, C. M., Salerno, S., Koch, G., Turriziani, P., Mangano, R., … Caltagirone, C. (2008). Perceiving numbers alters time perception. *Neuroscience Letters*, *438*(3), 308–311. http://doi.org/10.1016/j.neulet.2008.04.051

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, 0956797613482944.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., … Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. http://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, *44*(3), 547–555. http://doi.org/10.1016/j.neuron.2004.10.014

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993.

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biology*, *5*(8), e208.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, *110*(45), 18116–18120. http://doi.org/10.1073/pnas.1302751110

Stone, H., & Bosley, J. J. (1965). Olfactory Discrimination and Weber’s Law. *Perceptual and Motor Skills*, *20*(2), 657–665. http://doi.org/10.2466/pms.1965.20.2.657

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T.

(2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Developmental Psychology*, *4*, 444. http://doi.org/10.3389/fpsyg.2013.00444

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(36), 14513–14518.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008a). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008b). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425.

Van Opstal, F., & Verguts, T. (2013). Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Cognition*, 435. http://doi.org/10.3389/fpsyg.2013.00435

Vicario, C. M. (2013). On a generalized magnitude system in the brain: an integrative perspective. *Cognition*, *4*, 829. http://doi.org/10.3389/fpsyg.2013.00829

Vicario, C. M., & Martino, D. (2010). The neurophysiology of magnitude: One example of extraction analogies. *Cognitive Neuroscience*, *1*(2), 144–145. http://doi.org/10.1080/17588921003763969

Vicario, C. ., Rappo, G., Pepi, A., Pavan, A., & Martino, D. (2012). Temporal Abnormalities in Children With Developmental Dyscalculia. *Developmental Neuropsychology*, *37*(7), 636–652. http://doi.org/10.1080/87565641.2012.702827

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. http://doi.org/10.1038/nn1057

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. http://doi.org/10.1016/j.tics.2003.09.002

Wang, J., Odic, D., Halberda, J., & Feigenson, L. (under review). A five-minute intervention improves preschoolers' intuitive sense of number with significant transfer to symbolic mathematics performance.

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, *50*(4), 489.

Young, L. N., & Cordes, S. (2013). Fewer Things, Lasting Longer The Effects of Emotion on Quantity Judgments. *Psychological Science*, *24*(6), 1057–1059

Taza decorativa de Metáforas en alemán como L2
Decorative mug of Cognitive metaphor in German as L2

Autores: **Emiliano de Armas Cabrera** y **Luis Andrés Ochoa**. Financiación: Programa de Apoyo a la Investigación Estudiantil (PAIE) Universidad de la República, Uruguay. Este estudio evaluó los efectos del enmarcado lingüístico de metáforas conceptuales entre lenguas. En este caso, en nativos de español uruguayo que aprenden alemán como lengua extranjera. Los resultados mostraron que las metáforas compartidas entre L1 y L2 con idéntica realización lingüística muestran contrastes significativos respecto a la referencia literal sin diferencia de grupos de aprendizaje incial o avanzado. También, sugirieron que la exclusividad lingüística de las metáforas afecta su procesamiento. Los resultados fueron limitados ante la escasa muestra pero sugieren vías interesantes para seguir explorando.

Authors: **Emiliano de Armas Cabrera** and **Luis Andrés Ochoa.** Funded by: Programa de Apoyo a la Investigación Estudiantil (PAIE) Universidad de la República, Uruguay. This study evaluated the effects of linguistic framing of conceptual metaphors between languages. In this case, in native speakers of Uruguayan Spanish learning German as a foreign language. The results showed that metaphors shared between L1 and L2 with identical linguistic realization show significant contrasts with respect to literal reference with no difference in initial or advanced learning groups. They also suggested that the linguistic uniqueness of metaphors affects their processing. The results were limited by the small sample size but suggest interesting avenues for further exploration

Building Blocks of Mathematical Learning: Virtual and Tangible Manipulatives Lead to Different Strategies in Number Composition

Ana Cristina Pires [1,2]*, Fernando González Perilli [2,3], Ewelina Bakała[4], Bruno Fleisher [2,3], Gustavo Sansone[5] and Sebastián Marichal [6]

Ana Cristina Pires https://orcid.org/0000-0001-7747-7112, Fernando González Perilli https://orcid.org/0000-0001-5832-716X, Ewelina Bakała https://orcid.org/0000-0003-2507-2118, Gustavo Sansone https://orcid.org/0000-0002-8693-4749, Sebastián Marichal https://orcid.org/0000-0001-5067-049X

1 LASIGE, Faculty of Sciences, Universidade de Lisboa, Lisbon, Portugal
2 **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República. Montevideo, Uruguay
3 Facultad de Información de Comunicación. Universidad de la República. Montevideo, Uruguay
4 Facultad de Ingeniería. Instituto de Computación. Universidad de la República. Montevideo, Uruguay
5 Facultad de Arquitectura. Universidad de la República. Montevideo, Uruguay
6 Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, España

Correspondence concerning this article should be addressed to Ana Cristina Pires. Faculdade de Ciências da Universidade de Lisboa. Departamento de Informática. Edifício C6 Piso 3 - Sala 6.3.30. Lisboa, Lisbon, Portugal. Contact: **acdpires@di.fc.ul.pt**

This is a free version -with the the consent of the authors- of the open access manuscript Pires, A. C., González Perilli, F., Bakała, E., Fleisher, B., Sansone, G., & Marichal, S. (2019, September). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education* (Vol. 4, p. 81) Available at **https://doi.org/10.3389/feduc.2019.00081**

Abstract

Multiple kinds of manipulatives, such as traditional, virtual, or technology-enhanced tangible objects, can be used in primary education to support the acquisition of mathematical concepts. They enable playful experiences and help children understand abstract concepts, but their connection with cognitive development is not totally clear. It is also not clear how virtual and physical materials influence the development of different strategies for solving instructional tasks. To shed light on these issues, we conducted a 13-day intervention with 64 children from first grade, divided into three groups: Virtual Interaction (VI), Tangible Interaction (TI), and Control Group (CO). The VI group played a fully digital version of a mathematics video game and the manipulation of the blocks took place on the tablet screen. The TI group played the same video game with digitally augmented tangible manipulatives. Finally, the CO group continued with their classroom curricular activities while we conducted the training, and only participated in the Pre and Post-Test evaluations. Our results highlighted that the use of tangible manipulatives led to a positive impact in children's mathematical abilities. Of most interest, we recorded children's actions during all the training activities, which allowed us to achieve a refined analysis of participants' operations while solving a number composition task. We explored the differences between the use of virtual and tangible manipulatives and the strategies employed. We observed that the TI group opted for a greater number of blocks in the number composition task, whereas the VI group favored solutions requiring fewer blocks. Interestingly, those children whose improvement in mathematics were greater were the ones employing a greater number of blocks. Our results suggest that tangible interactive material increases action possibilities and may also contribute to a deeper understanding of core mathematical concepts.

Keywords: digital manipulatives, tangible manipulatives, technology-enhanced learning activities, mathematics, additive composition

Introduction

Learning mathematics at an early age is fundamental to ensuring academic success in STEM (science, technology, engineering, and mathematics) disciplines and maximizing future integration into professional life (Wang and Goldschmidt, 2003). Research has been concerned with how to foster this core cognitive ability and enable a deep understanding of mathematical concepts. This research explores how virtual and tangible manipulatives can be used to strengthen math learning at 6 years of age. In the current study, we used the activity of composing and decomposing sets of manipulatives representing numbers, an exercise that has been traditionally practiced with concrete material in order to foster an understanding of numerosity (Geary et al., 1992; Morin and Franks, 2009). We focused on a set of three properties (additive composition, commutativity, and associativity) and the mastery of the basic number combinations. Additive composition is the knowledge that larger sets are made up of smaller sets; the commutative property implies that changing the order of the operands doesn't affect the result; the associative property allows us to add (or multiply) numbers, no matter how the factors are grouped [$(a + b) + c = a + (b + c)$]; while mastering the basic number combinations leads to understanding how numbers can be composed. These properties are crucial for cardinality and number concept acquisition; and lead to the development of key strategies in arithmetical problem solving, such as addition and subtraction (Fuson, 1992; Verschaffel et al., 2007).

In mathematics curricula, teaching is frequently supported by tangible objects (three-dimensional models of geometrical shapes, etc.) that help young students to better understand abstract concepts, for instance in the acquisition of cardinality (Geary et al., 1992; Morin and Franks, 2009). The pioneer in this tradition was Maria Montessori who developed materials for geometry and mathematics specifically aimed at providing children with autonomy during the learning process (Montessori, 1917). Georges Cuisenaire, in turn, created a special set of tiles for arithmetics learning known as Cuisenaire rods (Cuisenaire, 1968). His proposal was based on the relationship between size and number and exploited the possibility of different spatial arrangements to exemplify mathematical principles like number composition. A new version of these materials can be found in Singapore Math's tiles (Wong, 2009; Wong and Lee, 2009); which is considered one of the more influential methods for teaching basic mathematics nowadays (Deng et al., 2013).

Following this vein, the acquisition of the number concept—one of the building blocks of mathematical learning—would benefit from direct interaction with objects (Dienes, 1961; Chao et al., 2000; Anstrom, 2006; McGuire et al., 2012). Interaction with objects may facilitate the passage from a concrete construal (I can see/manipulate three things in front of me) toward an abstract one (3 = * * *). This transformation begins with a process which is strongly based on perceptual, non verbal operations and turns into a symbolic one supported by an abstract association (Feigenson et al., 2004). The first stage has to do with the understanding that a given group of objects has a certain quantity of components (Gelman and Gallistel, 1978); the second with associating this quantity (of objects) to an exact number and its symbolic expression, and then understanding that any time the number is seen or heard it means that an exact quantity is being referred to (Kilpatrick et al., 2001).

The sensitivity to numerosity is improved gradually as the infant develops (Izard et al., 2009). Infants even just a few hours old are already sensitive to numerosity (e.g., Antell and Keating, 1983; Izard et al., 2009). Allegedly, this is possible due to two innate parallel number systems (see Feigenson et al., 2004; for a review see Piazza, 2010): an object file system (Feigenson and Carey, 2003) which accounts for the immediate identification of a discrete quantity of elements—subitizing (Kaufman and lord, 1949)—and is limited by the capability to attend to different objects at the same time; and an approximate number system

(ANS) which accounts for a non-symbolic continuous numerical representation involving large numbers (Gallistel and Gelman, 1992; Dehaene, 2011).

Nevertheless, children are not able to explicitly identify simple quantities involving numbers from 1 to 4 until 4 years old, and up to 5 until 5 years old. To do so, different skills must be developed such as counting and conceptual subitizing; the combination of two “subitizable” numbers, for e.g., recognizing the presence of a 3 (***) and a 4 (****) and implicitly composing a set of 7 (*******) (Steffe and Cobb, 1988; Clements, 1999). Toddlers recognize that sets can be combined in different ways, but this understanding is based on nonverbal, perceptual processes (Sophian and McCorgray, 1994; Canobi et al., 2002). Commutativity is only acquired later between 4 and 5 years old, as also the understanding that commutativity of added groups leads to associativity (Gelman and Gallistel, 1978; Canobi et al., 2002). Thus, associativity reflects conceptual reasoning about how groups can be decomposed and recombined (Sarama and Clements, 2009). Further, as children learn basic number combinations, they can master a broad set of heuristics when faced with addition and subtraction problems.

To foster the conceptualization of unit items children may rely on hand actions such as pointing or grasping (Steffe and Cobb, 1988). For instance, in the case of subtraction, small children often represent the minuend with the fingers (or objects) and fold their fingers (or remove objects) for the value of the subtrahend (Groen and Resnick, 1977; Siegler, 1984). In fact, most children cannot solve complex numerical problems without the support of concrete objects until 5.5 years old (Levine et al., 1992). Later on, children acquire retrieval strategies, accessing results directly from long term memory (Rathmell, 1978; Steinberg, 1985; Kilpatrick et al., 2001). For this to be possible, children need to master basic number combinations (Baroody and Tiilikainen, 2003), but also understand associativity (Sarama and Clements, 2009). Children typically progress throughout three phases to achieve mastery on basic number combinations: (a) Counting strategies—using object counting (e.g., with blocks, fingers) or verbal counting (b) Reasoning strategies—using known information (facts and relationships) to deduce the answer of an unknown combination; (c) Mastery-efficient responses [i.e., fast and accurate (Kilpatrick et al., 2001)].

Children’s addition and subtraction strategies also evolve during childhood. For instance, in order to solve 9 + 8, 4 to 5-year-old children would count from 1 to 9 for the first addend and then from 9 to 17 for the total sum (“counting all strategy”; Fuson, 1992; Verschaffel et al., 2007). Later on between 5 and 6 years old children would develop the more refined strategy of “counting on” in which the count starts from the cardinal of the larger addend (i.e., from 9 to 17; Carpenter and Moser, 1982; Siegler and Jenkins, 2014). More sophisticated part-whole strategies are developed with the achievement of associativity and the knowledge of how numbers from 1 to 10 can be composed (6– 7 years old; Canobi et al., 2002). To solve 9 + 8 children would be able to retrieve that 9 + 1 is one of the forms to compose 10, and then solve the problem by the easier 10 + 7 (also retrieving that 8−1 equals 7; Carpenter and Moser, 1984; Fuson, 1992; Miura and Okamoto, 2003).

Interaction with objects may supports the development of different strategies by diminishing cognitive load and freeing up working memory, given that the perceived entities are cognitively available through the objects that represent them in space (Manches and O’Malley, 2016). Object manipulation gives rise to operations that can work as analogies of abstract operations. For example, joining 2 elements to a group of another 3 forms a new group of 5. This concrete activity would be a metaphor of act of addition: 2 + 3 = 5. These conceptual metaphors work as scaffolding that allows children to grasp abstract ideas such as commutativity or associativity (Manches and O’Malley, 2016).

With the appearance of digital technologies, researchers have been exploring how the manipulation of digital (Yerushalmy, 2005; Moyer-Packenham and Westenskow, 2013) and/or technology-enhanced concrete material (Tangible User Interfaces or TUIs; Manches, 2011) can benefit learning processes, finding promising results (see Sarama and Clements, 2016). Beyond the encouraging results obtained in several technology-based interventions, it has been claimed that the application of digital technology in the classroom posits the risk of replacing rich physical interactions with the environment by much more constrained interactions such as the use of the mouse–keyboard or multi-tactile interfaces (Bennett et al., 2008). In this vein, theories like constructivism, embodied cognition (Wilson, 2002; Anderson, 2003) and physically distributed learning (Martin and Schwartz, 2005) support the idea that physical interaction plays a key role in the learning process (Antle and Wise, 2013; for a review in this matter see Sarama and Clements, 2016).

In this study, we focus on the kinds of actions virtual and physical manipulatives offer and their impact on numerical learning. On one hand, interaction with virtual manipulatives is limited to dragging objects on the screen, but it still allows children to displace, join and isolate objects as traditional manipulatives allow (Moyer-Packenham and Westenskow, 2013). On the other hand, classic manipulatives offer interactive advantages (to grasp the object, for instance) that could have relevant consequences for educational activity (Martin and Schwartz, 2005; Manches and O’Malley, 2016). Several studies have been dedicated to this comparison, providing results which are slightly favorable to physical manipulatives (Martin and Schwartz, 2005; Schwartz et al., 2005; Klahr et al., 2008). Technology-enhanced tangible manipulatives offer several advantages when compared with traditional or virtual manipulatives (Moyer-Packenham and Westenskow, 2013). They allow autonomous and active learning by using physical material and enable us to record a child’s performance. In addition, they enable us to explore which kind of actions are relevant in specific learning activities. Importantly for the present research, our system permits analyzing and comparing the use of physical and virtual manipulatives to solve a task of additive composition. This comparison is of special theoretical interest given that it makes possible to explore the role of physicality/three-dimensionality in learning mathematics. In other words, the present research aims to investigate if it is indispensable that objects may be grasped, lifted, and explored or would it be enough to interact with virtual manipulatives? And specifically, we ask how the objects’ affordances (i.e., the possibility to grasp physical objects or drag virtual ones) will shape and constrain children’s composing strategies.

Material and Methods

Participants We recruited participants from one state school in Montevideo (Uruguay) with a medium-high sociocultural status consisting of 64 children (three classrooms) from first grade. All children had an informed consent form signed by their parents or legal guardians. A research protocol was approved by the Local Research Ethics Committee of the Faculty of Psychology, and is in accordance with the 2008 Helsinki Declaration. We employed a quasi-experimental design and each classroom became one of the following experimental groups: Control (CO), Virtual Interaction (VI), and Tangible Interaction (TI). Four children (two from the VI group and another two from the TI group) failed to correctly answer 25% of the trials in our training game. Therefore, we performed subsequent analyses with the remaining 60 children (33 girls and 27 boys). Group descriptive information is shown in Table 1. We examined the effect of age and sex by conducting separated t-tests on assessment scores, but we did not find any effect.

Table 1. Mean and standard deviations at pre- and post-tests by groups.

	n	Age (years)	Sex (*girls)	TEMA-3	
				Pre	Post
Passive Group (PA)	20	6.6 (0.3)	13	25.6 (5.7)	28.8 (4.6)
Virtual Interaction Group (VI)	20	6.8 (0.5)	11	31.8 (9.6)	35.1 (9.3)
Tangible Interaction Group (TI)	20	6.8 (0.6)	11	30.2 (10.3)	34.4 (10.5)

Procedure To evaluate the impact of both game modalities in the acquisition of mathematical abilities, we planned an intervention with three phases. A first and last phase of evaluations (Pre- and Post-Test), and a training of 13 days in between.

Pre-test To evaluate children’s mathematical abilities before and after training we used the third edition of the standardized Test of Early Mathematics Ability (TEMA-3, Bliss, 2006) for children between 3 and 8 years of age. The test was verbally administered and consisted of 72 items to assess: counting ability, number comparison facility, numeral literacy, mastery of number facts, basic calculation skills, and understanding of mathematical concepts. This test has high content validity (Baroody, 2003) and high reliability ranging from 0.82 to 0.97. Indeed, we found a high test–retest reliability measured by calculating TEMA 3 correlation between Pre-Test and Post-Test measures across children within each training group (TI: 0.94; VI: 0.94; CO: 0.78). We calculated scores by the sum of all the correct answers (taking into account ceiling and floor effects that are part of the test administration). Two trained evaluators conducted the evaluation and it took about 30 min per participant. This phase took one week, with 12 children evaluated per day.

Training/Playing The three classes selected to participate in the study continued with their regular formal learning activities as part of the school curriculum. Apart from the fact that each class had a different teacher, teachers followed the same program and protocol, and were committed to giving the same math curricula information for the three classes. Both the TI and VI group played over 13 days (3 weeks). Sessions had a duration of 20 min each, from Monday to Friday. Two researchers were present in every session to help with any technical problems that may have arisen. In the first session, we introduced the game dynamics and made explicit the relation between size and value of each tangible and virtual block to facilitate effective use of manipulatives. The CO group continued with their regular curricular activities while the other two groups had 20 min per day of training. The CO group only participated in the Pre- and Post-Tests assessments.

Post-test The same evaluators assessed the groups again with TEMA-3 and the scores were analyzed in the same manner as in the Pre-Test evaluation.

Training Game BrUNO The video game BrUNO was developed to give the learning activity a more attractive and playful format. We took gamification theory into consideration in order to incorporate some gamification elements in BrUNO, such as: microworlds, a main-character, a tutorial, several types of prizes, and funny sounds. During the development of BrUNO, we carried out two informal user tests to inform the game design (Marichal et al., 2017a).

BrUNO is a video game designed to work on additive composition. Children played BrUNO by using five types of blocks whose length and color were associated with their value (see Figure 1). The block of 1 represents the number "1"; the block of 2 represents the number "2," and so forth until 5. Each block has a different length which is proportional to the value that it represents).

Block	Block value	Digital block dimensions	Tangible block dimensions	Color
	1	40px x 40px	16mm x 16mm	yellow
	2	80px x 40px	32mm x 16mm	green
	3	120px x 40px	48mm x 16mm	blue
	4	160px x 40px	64mm x 16mm	orange
	5	200px x 40px	80mm x 16mm	red

Figure 1. Block values, dimensions, and color.

To facilitate visual recognition of the location of the number required to build, a horizontal or vertical number line (depending on the scenario) is shown on the screen (see **Figure 2**). It is known that as numerosity develops, a hierarchical mental representation of how numbers should be ordered arises in the form of a number line. This line, which is based on a spatial analogy, represents the numbers from lowest to highest and locates them according to their cardinality. Thus, to reinforce this mental representation and to facilitate the additive composition task, we presented a number line to guide the players while they compose the required number. It helps to count the missing/spare units and deduce how the target number can be correctly composed. If the child has to build the number 4 and she has already put one block of 3, she can observe that the game character is 1 unit away from the prize and compose the target number by adding the block of 1. This way, the child can learn that 3 + 1 = 4. Additionally, the game helps to demonstrate that, for example, the distance between 1 and 3 is the same as between 21 and 23—a fact that is not so obvious for young children (Siegler and Booth, 2004).

Figure 2. Fully virtual version of BrUNO. Prize placed in number three (as indicated by the orange color). The player has already introduced 1 block of value 2. To reach the prize, he must add one block of value 1. In this example, a horizontal number line is present to help children locating numbers and to help in adding and subtracting operations.

We developed two conditions for the evaluation of manipulatives: the Tangible Interaction Group (TI) and the Virtual Interaction Group (VI). In both cases, children played BrUNO, but the interaction with the blocks differed. In the first case, children manipulated technology-enhanced tangible blocks, and in the second case, virtual blocks.

Tangible Interaction Device We designed a low cost tangible interaction device named CETA (Marichal et al., 2017a), with three main components (see Figure 3): a mirror that changes the webcam's viewing direction, allowing the system to detect objects over the table; a wooden holder that keeps the tablet vertically in portrait orientation; and a set of tangible blocks of different sizes similar to Cuisenaire Rods (representing numbers from 1 to 5; see Figure 1)

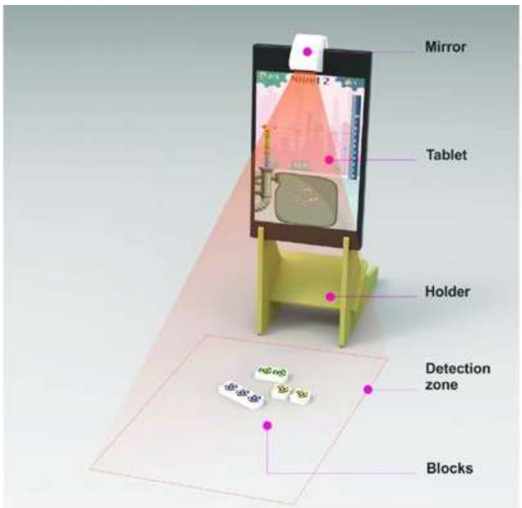


Figure 3. Tangible setting for BrUNO. Figure reproduced with author's permission (Marichal et al., 2017a).

We used the webcam of the tablet and a mirror to capture the image of the surface in front of the tablet holder in real-time. This image is constantly analyzed to detect blocks in the detection zone (for more details see Marichal et al., 2017b). The limits of the detection zone are determined by the webcam hardware and height of the holder. Blocks outside the detection zone are not visible to the computer vision system.

We designed a set of 25 blocks for 3D printing. The handling capabilities of the children at target age, the dimensions of the detection zone of the computer vision system, and the numeric quantities required by the different game challenges determined the dimensions of the blocks. All blocks contain magnets at their extremities, providing an affordance that increases the probability of joining blocks imitating the number line representation. Every block has a positive and a negative extremity. The concave and convex block's terminations constrain the way it can be joined. On the top face of each block we placed a set of colored markers (TopCodes; Horn, 2012) used by the computer vision system. The number of markers on each block corresponds to the block value.

Virtual Interaction Device The virtual version allows to play BrUNO without CETA device. The blocks are virtual and the child has to place them in the detection zone to submit its answer to the system (Figure 2).

Data Collection

We recorded the children's actions to trace the quantity and the type of blocks employed in children's solutions over time. This allowed us to analyze the game strategies developed by each group and follow the performance of every single participant. After each response our system recorded the following data: (1) the number required to form, (2) the number actually formed, and (3) the blocks used to form the number.

We assumed that if the child wanted to respond with two blocks but put the first block in the detection zone while looking for the other, then we should develop a strategy to avoid considering this incomplete answer as a child's final solution. Thus, to avoid recording partial solutions we implemented what we call "action submit," which consists of two steps. The first step is to wait for a stable solution. By stable solutions, we mean invariant responses by children for 1.5 s meaning that the blocks placed in the detection zone were not moved for 1.5 s and no blocks were added or removed. If this condition was completed, then we move to the second step in which the game character prepares itself for 1 s to execute the movement. If, during this time the child changed his or her answer, the time counter resets and "action submit" starts over again. If the answer did not change, the game character moves and the system records the blocks that composed the child's solution. To avoid duplicate responses (e.g., the child leaves the blocks in the detection zone and goes to the bathroom) we only registered the solutions that differed from the last recorded solution.

Results

Differences Between Groups To test the effect of playing our training game over 13 sessions, we assessed the children's mathematics performance using TEMA-3 before and after training or without training as in the case of the CO group.

While we had a quasi-experimental design in which the groups were non-randomized at baseline, there were no significant differences between groups on Pre-Test, $p = 0.84$. To test for conditional differences, we used an ANCOVA with the Post Test scores as the dependent variable, the Pre-Test as the covariate, and the Group as the independent variable. ANCOVA is advocated in this type of context because it controls for minor variations in the Pre-Test scores (Oakes and Feldman, 2001; Schneider et al., 2015). The assumptions of the ANCOVA were satisfied (as noted above, the covariate levels did not differ between conditions, and homogeneity of slopes held, as verified by running an ANOVA and customizing the model to include the interaction between the covariate and independent variable, $p = 0.5$). The ANCOVA identified a significant effect of Group, $F_{(2, 54)} = 20.9$, $p < 0.001$, $r = 0.44$. We followed up this analysis with pairwise comparisons between Post-Test scores adjusted by the ANCOVA with the baseline Pre-Test scores. Both experimental groups obtained higher Post-Test scores than the control group (VI_{Mean}: 32.54, $VI_{SD} = 0.77$; TI_{Mean}: 33.27, $TI_{SD} = 0.74$ and CO_{Mean}: 30.93, $CO_{SD} = 0.86$). However, only Post-Tests scores significantly differed when comparing TI vs CO ($p = 0.044$). We found no other significant effects between groups.

Virtual and Tangible Interaction Groups and the Minimum Blocks Coefficient (MBC) We focused on the possible problem-solving strategies employed by the children when resolving the number composition task, and how the type of interaction could have affected their actions. To do so, we carried out exploratory analysis using participants' log files. It allowed us to observe which blocks were used to compose each number by all the participants, at every successful trial.

Firstly, we analyzed whether the number of blocks used to build the correct solution was different across groups. For example, to build the number 3, it is possible to use three blocks of 1 ("1-1-1"), one block of 1 and one block of 2 ("1-2"), or directly use one block of 3 ("3"). To evaluate how close the child was to using the minimum number of blocks that were necessary to build a number (one block in the case of numbers from 1 to 5, two blocks in case of numbers from 6 to 10, or three blocks if the number is greater than 10), we developed a score called the "Minimum Blocks Coefficient" (MBC). MBC is a metric that allows us to observe the different solutions in composing numbers while training additive composition. We aim to explore how children compose numbers using different types of manipulatives. For each correct solution it takes the minimum number of blocks necessary to build the number requested, and divides it by the number of blocks actually used. For example, in the case of number 3 the variant "1-1-1" becomes the score 1/3 = 0.33, because just one block is necessary to build the number (block of 3), and in reality, three blocks were used. The combination "1-2," becomes 1/2 = 0.5, and "3," becomes the score of 1.0. To calculate the MBC for one particular number and one particular group (TI or VI), we take all the correct solutions of the number formed by the participants of the group and calculate the mean value. Error rates were not analyzed because we observed that the tangible system required more time for the physical manipulation and during that time some partial solutions were recorded as errors before the child's final answer. For example, if the child wanted to respond with two blocks, but he or she put the first block in the detection zone while looking for the other and n changes occur in the detection zone for 2.5 s, the system registered the child's uncompleted solution as a response (error in this case). The algorithm is explained with more detail in the section "2.3.3." For the aforementioned reasons we decided to only analyze the correct answers, so we were confident that we analyzed explicitly correct answers rather than random solutions.

Minimum Blocks Coefficient by Numbers (1–13) We applied a two-way ANOVA considering the MBC as the dependent variable and Group and Numbers as the independent variables. Numbers is the variable that represents the number the child is asked to build. We divided all the Numbers that appear in the game (1–13) into three ranges based on the theoretical MBC that could be used for those numbers. Specifically, the theoretical MBC for numbers ranged from 1 to 5 is one block (i.e., they have the possibility to respond with a minimum of one block); for the numbers ranged 6–10 is two (i.e., they have the possibility to respond with a minimum of two blocks) and for the numbers ranged from 11 to 13 is three blocks (i.e., they have the possibility to respond with a minimum of three blocks).

The results showed that the type of manipulatives (TI or VI group) [$F_{(1, 126)} = 6.21$, $p = 0.014$, $r = 0.076$] and the Number [$F_{(2, 126)} = 10.8$, $p < 0.001$, $r = 0.060$] (see Figure 4) significantly influenced the MBC. We found no further interaction. The TI group used significantly more pieces (lower MBC) comparing with the VI group ($TI_{Mean} = 0.65$, $TI_{SD} = 0.19$, $VI_{Mean} = 0.72$, $VI_{SD} = 0.15$). These differences between TI and VI may be a result of the diverse composing strategies used when solving the number composition task.

Considering the variable Number, the number of blocks used were significantly fewer for the numbers ranging from 1 to 5 compared to the numbers ranging from 6 to 10 ($p = 0.0002$) and also compared to the numbers ranging from 11 to 13 ($p = 0.0003$).

Minimum Blocks Coefficient Over Time Participants reduced the number of blocks used during the 13 sessions that our intervention lasted (see Figure 5). We found a significant positive correlation ($ps < 0.0001$) between the MBC and sessions for VI (0.84) and for TI (0.87) groups. We also explored whether the number of blocks employed was significantly different at different moments of our intervention by analysing

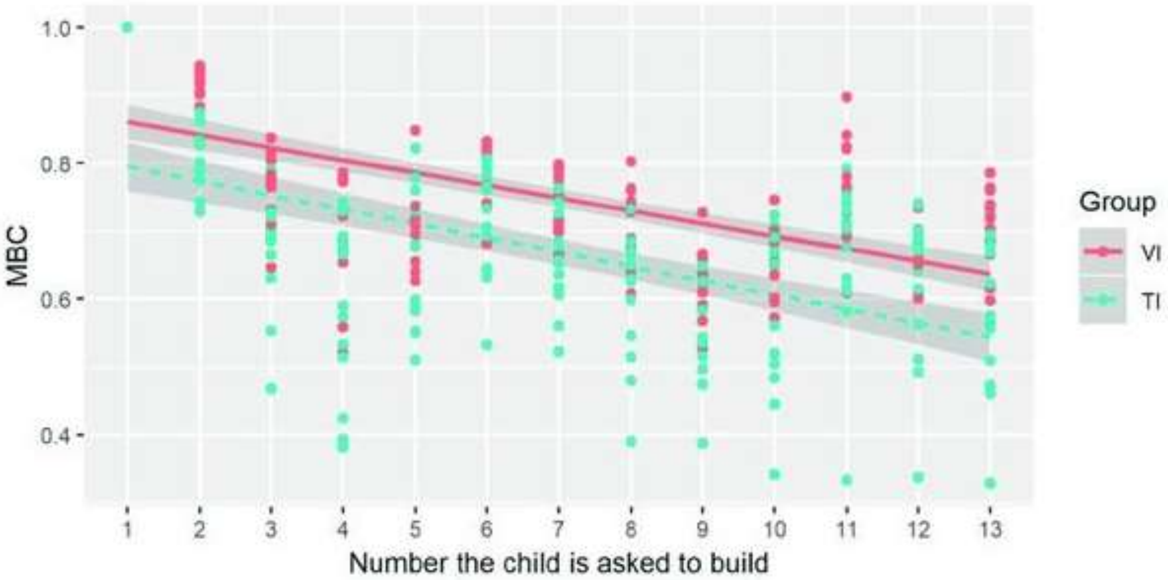


Figure 4. The Minimum Blocks Coefficient (MBC) for each number the child was asked to build. We applied a linear model to data points with a 95% confidence level for each Experimental Group: Virtual Interaction (VI) and Tangible Interaction (TI).

the MBC Mean for the first and last three sessions for both groups. Interestingly, in the first three sessions, the MBC was greater for the VI group, i.e., children used fewer blocks ($p < 0.0001$). In contrast, when analysing the last three sessions, the MBC did not differ between either group.

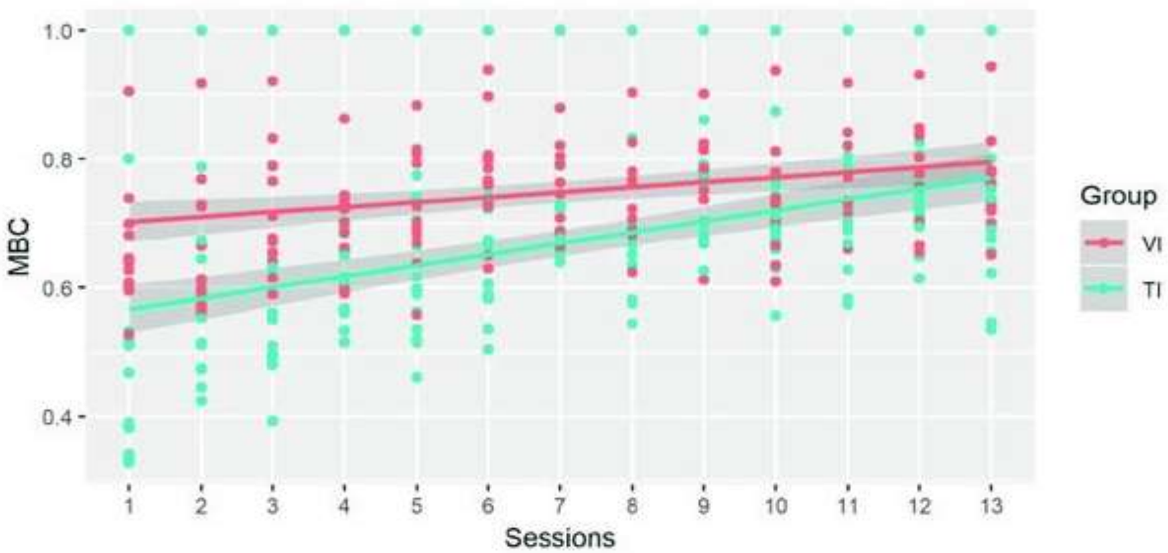


Figure 5. Minimum Blocks Coefficient (MBC) for each session and experiment group. We applied a linear model to data points with 95% confidence level.

Minimum Blocks Coefficient and Mathematics Improvement
We explored the relationship between the number of blocks employed during the intervention (measured by MBC) and the amount of mathematical improvement (dScores: Post-Test scores − Pre-Test Scores) and found no correlation ($p > 0.05$). Neither TI nor VI groups showed a significant correlation between MBC and Score when analyzed separately ($p > 0.05$).

Further, we decided to analyze the differences in the number of blocks employed comparing the performance of the Better and Worse Improvers. Thus, we divided all participants by the median of the dScore comprising two groups. The Better Improvers were the children with a dScore above the median, while the Worse Improvers were the ones whose dScore was below the median (see Figure 6). We found a significant negative correlation between MBC and dScores for the Better Improvers ($cor = -0.50$, $p = 0.021$), but not for the Worse Improvers. In conclusion, the children that had a greater improvement were the children using more blocks than the minimum blocks necessary to build the numbers required by the game. In contrast, we did not observe any change in the number of blocks used by the children who did not improve in mathematics.

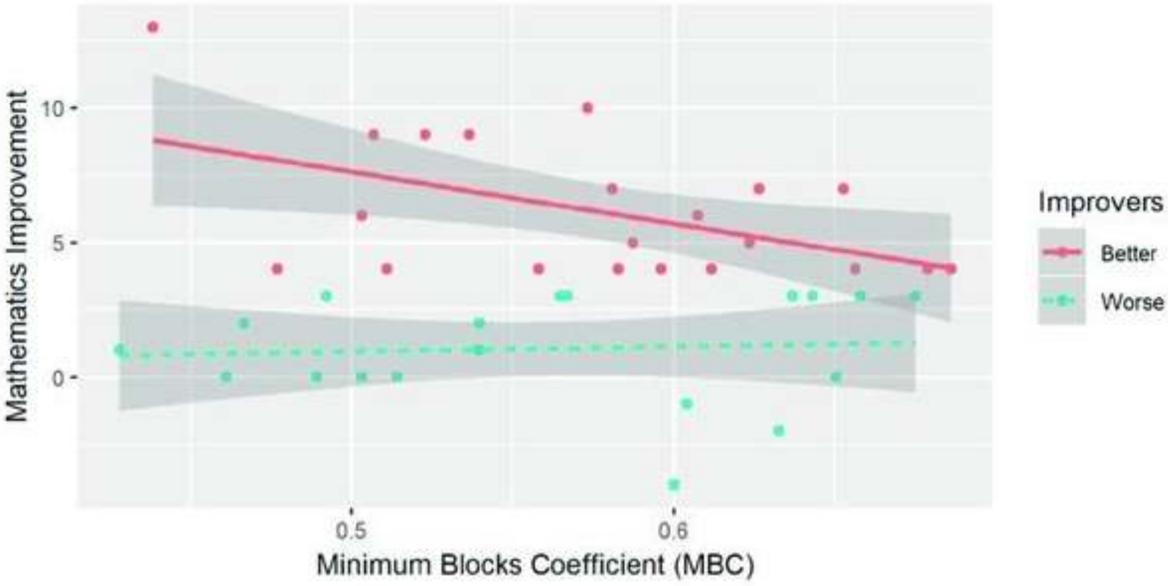


Figure 6. Minimum Blocks Coefficient by mathematics improvement for better and worse improvers.
We applied a linear model to data points with a 95% confidence level.

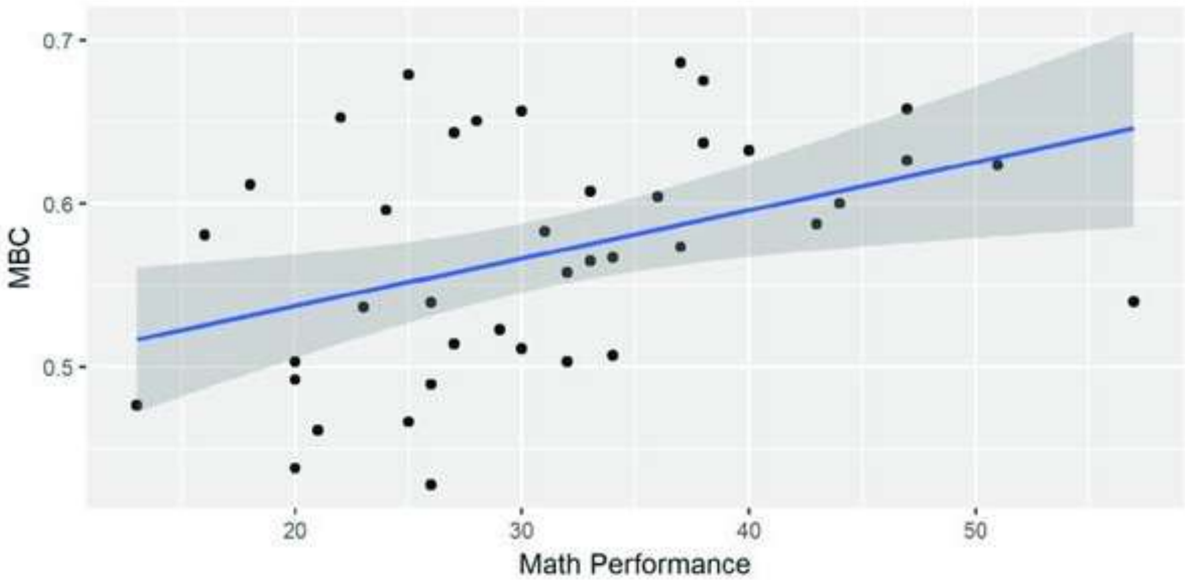


Figure 7. Minimum Blocks Coefficient (MBC) by math performance (pre-test scores).
We applied a linear model to data points with a 95% confidence level.

Minimum Blocks Coefficient and Mathematics Performance
We were also interested in the relationship between the Minimum Blocks Coefficient (MBC) and mathematical performance (Pre-Test scores). Analysis indicated that Pre-Test scores were positively correlated with the MBC (cor = 0.41, $p = 0.009$; see Figure 7). Children who had greater Pre-Test scores at the beginning of this study had the tendency to use less number of blocks during the game.

Discussion

Impact of Manipulatives on Mathematical Learning

Our results indicate that the tangible manipulative group showed an advantage in mathematics scores after training compared to the control group. Our findings highlight the possibility of improving mathematical ability by practicing implicit number composition tasks assisted by tangible manipulatives.

We did not find significant differences either between the two types of manipulatives (virtual and tangible), or between virtual manipulatives and the control group when considering mathematical improvement tested by TEMA-3. It may be the case that virtual tangibles also have an impact in Post-Test scores, which was not observed due to the lack of statistical power of the present study.

Virtual and Tangible Manipulatives Led to Different Strategies in Number Composition

We analyzed children's behavior during our intervention to look for possible differential profiles in their evolution during training. Our tablet-based intervention allowed us to record the children's responses every time they submitted a block to compose a number. Our results enabled us to reflect on the role of specific actions performed by children affecting the learning process, and how learning could be influenced by the interactive properties of the blocks rendered as a representational assistance (Manches and O'Malley, 2016).

It was observed that the TI and VI groups significantly differed in the numbers of blocks used to compose a number. VI employed significantly fewer blocks compared with TI, showing that the different type of manipulatives could have led to different problem solving strategies. TI children opted to compose numbers using more varied combination of blocks, i.e., they used more number composition strategies. This suggests that the affordances of physical objects do trigger more diverse solutions (Manches and O'Malley, 2016), which have been advocated to prompt better learning experiences in numerosity knowledge (Alibali and Goldinmeadow, 1993; Chi et al., 1994; Siegler and Shipley, 1995) and specifically foster mastery of basic number combinations (Baroody and Tiilikainen, 2003; Sarama and Clements, 2009).

Our results are in accordance with Manches et al. (2010) results that found that children employed a significantly greater number of solutions when they used plastic blocks as manipulatives, comparing with a condition in which children were aided with a visual representation drawn on paper. For instance, it is easier to detect the "reversion" strategy (5-2, 2- 5) when you can hold and displace objects representing these quantities (2 and 5). This finding supports the view that objects affordances implicitly carry information that could be relevant to reflect on abstract concepts, through conceptual metaphors. In our study, we compared tangible blocks (TI group) against virtual blocks (VI group). The use of virtual blocks allowed the children to drag, transform, and move blocks which allows a richer interaction compared to blocks drawn on paper. However, when compared to virtual blocks, tangible blocks enabled a more diverse combination of blocks to compose numbers as also observed elsewhere (Manches et al., 2010).

Strategies Evolution in Number Composition

When we analyzed strategies during training sessions we found that at the beginning of the training both groups employed more blocks to compose numbers with a tendency to diminish in the last sessions. This tendency to diminish may represent an approach to optimal performance (when the number is composed by the minimal quantity of possible blocks), probably reflecting learning toward increasing efficient and fastest strategies in number composition (Baroody and Dowker, 2003).

This is in line with the fact that composing and decomposing strategies becomes semiautomatic or automatic with effective and faster answers to basic number combinations. Children may automatize some combinations of a number through practice, resulting in an association with their counting knowledge. This association encourages efficiency, preventing children from repeatedly practicing all the possible combinations (Baroody, 2006). In our study, children at the beginning started by practicing various combinations of numbers. For instance, in the first sessions to form the number 5 children might use several combinations as 1+1+1+1+1, 2+2+1, 2+1+1+1, reflected by low MBC scores. Nevertheless, at the end of the training sessions children were able to answer more effectively, reflected by high MBC scores. For instance, to form the number 5 they answered with the block 5 or by adding just two blocks as 2+3 or 4+1, which is quicker and more direct.

Analyses showed that the mean of blocks used in the first three sessions was significantly smaller for the VI group, whereas both groups employed the same number of blocks in the last three sessions. This suggests that besides the tendency of both groups to optimize responses, they presented a different profile in their evolution during training. Children who used tangible manipulatives had the tendency to use more blocks and showed a more pronounced decrease in the number of blocks used during the intervention compared to children who used virtual manipulatives. This finding may be connected to the observed improvement in maths scores (measured by TEMA 3) for the TI group. The number of combinations used in the TI may have contributed to achieving mastery in mathematical knowledge, since mastery in basic number composition is enriched by experiencing more varied possibilities (Markman, 1978; Bowerman, 1982; Karmiloff-Smith, 1992). In this study, physical object affordances offered the user a richer set of action possibilities, and most probably also a more comprehensive understanding of the phenomenon explored.

Strategies in Additive Composition Task and Mathematical Improvement

We did not find a correlation between the number of blocks employed by children and mathematical improvement in general (all children analyzed together). Nevertheless, when children were divided according to their improvement in mathematics (Post-Test − Pre-Test) after the intervention, it was observed that the greater improvement group showed a positive correlation between number of blocks employed and gain in mathematical knowledge, which was not found for the Worse Improvers.

Therefore, children who showed a greater improvement tended to use more blocks. This outcome may suggest that an optimal performance in number composition (understood as fewer pieces used to form a number equals better performance) would not necessarily lead to a better learning experience. Another hypothesis would be that children who do not already have this mastery in number combinations, i.e., efficient, fast and accurate responses, would benefit more from employing manipulatives to solve additive composition and this might be the case for the "Better Improvers." Children who improved at maths during training were the ones using more varied block combinations. This is connected to the fact that the use of a greater variety of strategies can result in a better learning outcome (Markman, 1978; Bowerman, 1982; Karmiloff-Smith, 1992).

Strategies in Additive Composition Task and Mathematics

Interestingly, a negative correlation was found between mathematical scores at the Pre-Test (how good the children were at the beginning of the study) and the number of blocks employed. That is, being better at mathematics at Pre-Test implied the use of fewer manipulative blocks, probably due to a better knowledge of retrieval strategies while composing numbers (Rathmell, 1978; Steinberg, 1985; Kilpatrick et al., 2001). Children who were good at maths at the beginning of the training will not necessarily use more strategies because they already have a deeper knowledge in number concept and composition. That is to say, children who have already learned basic combinations of numbers have the ability to use such knowledge to answer quickly and efficiently in a familiar and unfamiliar learning context (Baroody, 2006).

It may seem contradictory that children who obtained the best scores at TEMA-3 (better at mathematics at baseline) used fewer blocks whereas the Better Improvers tended to employ more. However, according to Sarama and Clements (2009), despite seeming paradoxical, those who are better at solving problems with objects, fingers or counting are less likely to persist in these strategies in the future—as already reported by Siegler (1993)— but this is because they trust their answers and therefore move toward more precise strategies based on the retrieval of number combinations, leaving behind what once served as a scaffolding.

These results also suggest that children who will benefit more from the use of manipulative blocks are the children who do not have already mastery in number combinations. The use of enhanced manipulatives may be more suitable for younger children who need to practice and automatize simple number combinations.

Limitations

The present study has several limitations that should be considered when interpreting the results. It may lack statistical power since the number of participants in each group is small and for such reason, a larger confirmatory study is needed to strengthen the conclusions of the present study. The quasi experimental design of the current study has more ecological validity (children were kept in their school groups), but it is susceptible to threats on internal validity compared to controlled experimental designs and for that reason we consider our results as exploratory and conclusions are drawn carefully.

Conclusions

Current findings indicate that the use of tangible manipulatives had a positive impact on mathematical learning. We were able to observe interesting relationships between the level of mathematics and the kind of manipulative strategies chosen by the children when solving number composition tasks. Our results suggest that tangible manipulatives increase action possibilities and may also contribute to a deeper understanding of core mathematical concepts. Playing the game BrUNO with tangible manipulatives promotes meaningful practice of more varied number combinations by encouraging children to focus on patterns and relationships in basic number combinations. In addition, we were able to observe how their responses pattern changed throughout the training leading to the use of less but efficient strategies in the last sessions which may reflect that they achieved mastery in doing such combinations. Thus, training in this basic combinations led to an improvement in mathematics and hopefully may lead children to effectively apply this knowledge in new and unfamiliar number combinations.

From an interaction design perspective (for more details regarding this research and perspective, see Marichal et al., 2017a), the most relevant observation is how the objects' affordances (i.e., the possibility to grasp physical objects or drag virtual ones) somehow shape and constrain users' strategies. In our study, tangible blocks meant a richer interaction, providing the opportunity to explore more number composition possibilities. This possibly led to an improvement in mathematical performance. Thus, depending on the learning task objective (context), we might take advantage of this phenomena, by choosing either tangible, virtual or mixed learning environments. The current study invites researchers to delve deeper in the exploration of the potential for designing interactive activities aimed at fostering learning of specific target content.

Author Contributions

AP: substantial contributions to the conception or design of the work, analysis and interpretation of data for the work, drafting and revising it critically for important intellectual content. FG: drafting the work or revising it critically for important intellectual content, interpretation of data for the work. EB: substantial contributions to the conception and design of the work and data acquisition. BF: drafting the work or revising it critically for important intellectual content, analysis, and interpretation of data for the work. GS: substantial contributions to the design of the work. SM: substantial contributions to the design of the work, drafting and revising it critically for important intellectual content.

Ethics Statement

All children that participated in this research had the informed consent form signed by their parents or legal guardians. The intervention current protocol was approved by the Local Research Ethical Committee of the Faculty of Psychology, and is in accordance with the 2008 Declaration of Helsinki.

Funding

This work was supported by Agencia Nacional de Investigación e Innovación (ANII); Fundación Ceibal, Espacio Interdisciplinario, and by Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA), Universidad de la República, Uruguay; and by Fundação para a Ciência e a Tecnologia (FCT), I.P., through funding of project mIDR (AAC 02/SAICT/- 2017, project 30347, cofunded by COMPETE/FEDER/FNR), and of the LASIGE Research Unit (UID/CEC/00408/2013), Portugal.

Acknowledgments

We would like to show our gratitude to the children and educators that participated in this work from Escuela Panamá, Montevideo, Uruguay. We also want to thank our colleagues Rita Soria, María Pascale, Mariana Rodriguez, Leonardo Secco, Matías Correa, Leandro Fernández, Dilva Devita, Mariana Borges, Gonzalo Tejera, Alvaro Cabana, Fulvio Capurso, and Rodrigo López who provided insight and expertise that greatly assisted the research and design of CETA; and Catarina Tome-Pires and reviewers for suggestions and comments that greatly improved the manuscript.

References

Alibali, M. W., and Goldin-meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: what the hands reveal about a child's state of mind. *Cogn. Psychol. 25*, 468–523.

Anderson, S. E. (2003). The school district role in educational change: a review of the literature. Int. Centre Educ. Change. Available online at: http://fcis.oise. utoronto.ca/~icec/workpaper2.pdf

Anstrom, T. (2006). Supporting students in mathematics through the use of manipulatives. *Am. Inst. Res*. 1–15. Available online at: https://studylib.net/doc/ 18105288/supporting-students-in-mathematics-through-the-use-of-man

Antell, S. E., and Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Dev. 54*, 695–701. https://doi.org/doi: 10.2307/1130057

Antle, A. N., and Wise, A. F. (2013). Getting down to details: using theories of cognition and learning to inform tangible user interface design. *Interact. Comput. 25*, 1–20. https://doi.org/doi: 10.1093/iwc/iws007

Baroody, A. J. (2003). "The development of adaptive expertise and flexibility: the integration of conceptual and procedural knowledge," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise, Studies in Mathematical Thinking and Learning*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 1–33.

Baroody, A. J. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teach. Child. Math. 13*, 22–31. Available online at: https://www.semanticscholar.org/ paper/Why-Children-Have-Difficulties-Mastering-the-Basic-Baroody/753415c8d3fe2d6d3089bb0da0d9637dff02af32

Baroody, A. J., and Dowker, A. (eds.). (2003). *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Baroody, A. J., and Tiilikainen, S. H. (2003). "Two perspectives on addition development," in The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise, Studies in *Mathematical Thinking and Learning*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 75–125.

Bennett, S., Maton, K., and Kervin, L. (2008). The "digital natives" debate: a critical review of the evidence. *Br. J. Educ. Technol. 39*, 775–786. https://doi.org/doi: 10.1111/j.1467-8535.2007.00793.x

Bliss, S. (2006). Test of early mathematics ability-third edition. *J. Psychoeduc. Assess. 24*, 85–98. https://doi: 10.1177/0734282905282839

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni Semant. 3*, 5–66.

Canobi, K. H., Reeve, R. A., and Pattison, P. E. (2002). Young children's understanding of addition concepts. *Educ. Psychol. 22*, 513–532. https://doi.org/doi: 10.1080/0144341022000023608

Carpenter, T. P., and Moser, J. M. (1982). "The development of addition and subtraction problem-solving skills," in *Addition and Subtraction: A Cognitive Perspective*, eds J. M. Moser, T. A. Romberg, and T. P. Carpenter (Hillsdale, NJ: Lawrence Erlbaurn Associates), 9–24.

Carpenter, T. P., and Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *J. Res. Math. Educ. 15*, 179–202.

Chao, S.-J., Stigler, J. W., and Woodward, J. A. (2000). The effects of physical materials on kindergartners' learning of number concepts. *Cogn. Instruct. 18*, 285–316. https://doi.org/doi: 10.1207/S1532690XCI1803_1

Chi, M. T., De Leeuw, N., Chiu, M.-H., and LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cogn. Sci. 18*, 439–477. https://doi.org/doi: 10.1207/s15516709cog1803_3

Clements, D. H. (1999). Teaching length measurement: research challenges. *School Sci. Math. 99*, 5–11. https://doi.org/doi: 10.1111/j.1949-8594.1999.tb17440.x

Cuisenaire, G. (1968). Les Nombres en Couleurs. Bruxelles: Calozet. Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*, Revised and Updated Edition New York, NY: Oxford University Press.

Deng, Z., Gopinathan, S., and Lee, C. (2013). *Globalization and the Singapore Curriculum*. From policy. New York, NY: Springer.

Dienes, Z. P. (1961). *Building Up Mathematics*. London: Hutchinson Educational.

Feigenson, L., and Carey, S. (2003). Tracking individuals via object files: evidence from infants' manual search. *Dev. Sci. 6*, 568–584. https://doi.org/doi: 10.1111/1467-7687.00313

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci. 8*, 307–314. https://doi: 10.1016/j.tics.2004.05.002

Fuson, K. C. (1992). "Research on whole number addition and subtraction," in *Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics*, ed D. A. Grouws (New York, NY: Macmillan Publishing Co, Inc.), 243–275.

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. Cognition 44, 43–74. https://doi: 10.1016/0010-0277(92)90050-R

Geary, D. C., Bow-Thomas, C. C., and Yao, Y. (1992). Counting knowledge and skill in cognitive addition: a comparison of normal and mathematically disabled children. J. Exp. Child Psychol. 54, 372–391. https://doi.org/doi: 10.1016/0022-0965(92)90026-3

Gelman, R., and Gallistel, C. R. (1978). *The Child's Understanding of Number.* Cambridge: Harvard University Presss.

Groen, G., and Resnick, L. B. (1977). Can preschool children invent addition algorithms? J. Educ. Psychol. 69:645. doi: 10.1037//0022-0663.69.6.645 Horn, M. (2012). Topcode: Tangible Object Placement Codes. Available online at: http://users.eecs.northwestern.edu/~mhorn/topcodes/

Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A. 106*, 10382–10385. https://doi.org/doi: 10.1073/pnas.0812142106

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. London: MIT Press; Taylor and Francis.

Kaufman, E. L. and Lord, M. W. (1949). The discrimination of visual number. Am. J. Psychol. 62, 498–525. https://doi.org/doi: 10.2307/1418556

Kilpatrick, J., Swafford, J., and Findell, B. (eds.). (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academies Press.

Klahr, D., Triona, L., Strand-Cary, M., and Siler, S. (2008). "Virtual vs. physical materials in early science instruction: transitioning to an autonomous tutor for experimental design," in *Beyond Knowledge: The Legacy of Competence*, eds J. Zumbach, N. Schwartz, T. Seufert, and Kester L. (Dordrecht: Springer), 163–172.

Levine, S. C., Jordan, N. C., and Huttenlocher, J. (1992). Development of calculation abilities in young children. *J. Exp. Child Psychol. 53*, 72–103. https://doi.org/doi: 10.1016/S0022-0965(05)80005-0

Manches, A. (2011). Digital manipulatives: tools to transform early learning experiences. Int. *J. Technol. Enhanc. Learn. 3*, 608–626. https://doi.org/doi: 10.1504/IJTEL.2011.045451

Manches, A., and O'Malley, C. (2016). The effects of physical manipulatives on children's numerical strategies. *Cogn. Instruct. 34*, 27–50. https://doi.org/doi: 10.1080/07370008.2015.1124882

Manches, A., O'Malley, C., and Benford, S. (2010). The role of physical representations in solving number problems: a comparison of young children's use of physical and virtual materials. *Comput. Educ. 54*, 622–640. https://doi.org/doi: 10.1016/j.compedu.2009.09.023

Marichal, S., Rosales, A., Perilli, F. G., Pires, A. C., Bakala, E., Sansone, G., et al. (2017a). "Ceta: designing mixed-reality tangible interaction to enhance mathematical learning," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17* (New York, NY: ACM), *29*:1–29:13.

Marichal, S., Rosales, A., Sansone, G., Pires, A. C., Bakala, E., Perilli, F. G., et al. (2017b). "Ceta: open, affordable and portable mixed-reality environment for low-cost tablets," in *Proceedings of the 19th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '17* (New York, NY: ACM), 47:1–47:7.

Markman, E. M. (1978). Empirical versus logical solutions to part-whole comparison problems concerning classes and collections. *Child Dev.* 49, 168– 177. https://doi: 10.2307/1128605

Martin, T., and Schwartz, D. L. (2005). Physically distributed learning: adapting and reinterpreting physical environments in the development of fraction concepts. *Cogn. Sci. 29*, 587–625. https://doi: 10.1207/s15516709cog0000_15

McGuire, P., Kinzie, M. B., and Berch, D. B. (2012). Developing number sense in pre-k with five-frames. *Early Childh. Educ. J. 40*, 213–222. https://doi: 10.1007/s10643-011-0479-4

Miura, I. T., and Okamoto, Y. (2003). "Language supports for mathematics understanding and performance," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise, Studies in Mathematical Thinking and Learning*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 229–242.

Montessori, M. (1917). *The Advanced Montessori Method: Scientific Pedagogy as Applied to the Education of Children From Seven to Eleven Years*. London: W. Heinemann.

Morin, J. E., and Franks, D. J. (2009). Why do some children have difficulty learning mathematics? Looking at language for answers. *Prevent. School Fail. 54*, 111–118. https://doi.org/doi: 10.1080/10459880903217861

Moyer-Packenham, P. S., and Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *IJVPLE 4*, 35–50. https://doi.org/doi: 10.4018/jvple.2013070103

Oakes, J. M., and Feldman, H. A. (2001). Statistical power for nonequivalent pretest-posttest designs: the impact of change-score versus ancova models. *Eval. Rev. 25*, 3–28. https://doi.org/doi: 10.1177/0193841X0102500101

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci. 14*, 542–551. https://doi.org/doi: 10.1016/j.tics.2010.09.008

Rathmell, E. C. (1978). "Using thinking strategies to teach the basic facts," in *Developing Computational Skills*, ed M. N., Suydam (Reston: NCTM Yearbook), 13–38.

Sarama, J., and Clements, D. H. (2009). Early Childhood Mathematics Education Research: Learning Trajectories for Young Children. New York, NY: Routledge. Sarama, J., and Clements, D. H. (2016). "Physical and virtual manipulatives: what is "concrete"?" in *International Perspectives on Teaching and Learning Mathematics With Virtual Manipulatives*, ed P.S. Moyer-Packenham (Cham: Springer International), 71–93.

Schneider, B. A., Avivi-Reich, M., and Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Front. Psychol. 6*:474. https://doi.org/doi: 10.3389/fpsyg.2015.00474

Schwartz, D. L., Martin, T., and Pfaffman, J. (2005). How mathematics propels the development of physical knowledge. *J. Cogn. Dev. 6*, 65–88. https://doi.org/doi: 10.1207/s15327647jcd0601_5

Siegler, R., and Jenkins, E. A. (2014). *How Children Discover New Strategies*. New York, NY: Psychology Press.

Siegler, R. S. (1984). "Strategy choices in addition and subtraction: how do children know what to do?" in *Origins of Cognitive Skills*, ed C. Sophian (Hillsdale, NJ:Erlbaum), 229–293.

Siegler, R. S. (1993). "Adaptive and nonadaptive characteristics of low-income children's mathematical strategy use," in *The Challenge in Mathematics and Science Education: Psychology's Response*, eds L. A. Penner, G. M. Batsche, H. M. Knoff, and D. L. Nelson (Washington, DC: American Psychological Association), 341–366.

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev. 75*, 428–444. https://doi.org/doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Shipley, C. (1995). "Variation, selection, and cognitive change," in *Developing Cognitive Competence: New Approaches to Process Modeling*, eds T.J. Simon and G. S. Halford (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 31–76.

Sophian, C., and McCorgray, P. (1994). Part – whole knowledge and early arithmetic problem solving. *Cogn. Instruct. 12*, 3–33. https://doi.org/doi: 10.1207/s1532690xci1201_1

Steffe, L. P., and Cobb, P. (1988). Construction of Arithmetical Meanings and Strategies. New York, NY: Springer-Verlag.

Steinberg, R. M. (1985). Instruction on derived facts strategies in addition and subtraction. *J. Res. Math. Educ. 16*, 337–355. https://doi.org/doi: 10.2307/749356

Verschaffel, L., Greer, B., and De Corte, E. (2007). "Whole number concepts and operations," in *Second Handbook of Research on Mathematics Teaching and Learning*, ed F. Lester (Charlotte, NC: Information Age Publishing Inc.), 557–628.

Wang, J., and Goldschmidt, P. (2003). Importance of middle school mathematics on high school students' mathematics achievement. *J. Educ. Res. 97*, 3–17. https://doi.org/doi: 10.1080/00220670309596624

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev. 9*, 625–636. https://doi.org/doi: 10.3758/BF03196322

Wong, K. Y. (2009). *Mathematics Education: The Singapore Journey*, Vol. 2. Singapore: World Scientific Publishing Co.

Wong, K. Y., and Lee, N. (2009). "Singapore education and mathematics curriculum," in *Mathematics Education: The Singapore Journey*, eds K. Y. Wong, P. Y. Lee, B. Kaur, P. Y. Foong, and S. F. Ng (World Scientific), 13–47.

Yerushalmy, M. (2005). Functions of interactive visual representations in interactive mathematical textbooks. Int. *J. Comput. Math. Learn. 10*, 217–249. https://doi.org/doi: 10.1007/s10758-005-0538-2

Taza decorativa de lectura con soporte digital
Decorative mug of reading in digital media

Autores: **Eleonora Achugar** y **Álvaro Cabana.** Tesis de Maestría en Ciencias Cognitivas (en curso): *Estudio de lectura digital*. Financiación: Beca Agencia Nacional de Investigación e Innovación (ANII). Universidad de la República, Uruguay. Tutor de Tesis: Álvaro Cabana. Este estudio profundizó en el análisis del efecto de los soportes digitales en la comprensión de textos y la producción de sentido, así como en el tamaño del vocabulario y la cultura general.

Authors: **Eleonora Achugar** and **Álvaro Cabana**. Master's thesis in Cognitive Sciences (in progress): Digital reading study. Funded by: ANII Grant. Universidad de la República, Uruguay. Thesis Tutor: Álvaro Cabana. This study delved into the analysis of the effect of digital media on text comprehension and meaning production, as well as on vocabulary size and general culture.

Influence of time orientation on food choice: Case study with cookie labels

Giuliana Tórtora[1] & Gastón Ares[1,2]

Gastón Ares https://orcid.org/0000-0002-0565-8835

1 **Centro de Investigación Básica en Psicología (CIBPsi)**. Facultad de Psicología. Universidad de la República. Tristán Narvaja 1674. CP 11200. Montevideo, Uruguay
2 Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República. By Pass de Rutas 8 y 101 s/n. C.P. 91000. Pando, Canelones, Uruguay

Correspondence concerning this manuscript should be addressed to Gastón Ares, Facultad de Química, Instituto Polo Tecnológico de Pando. E-mail: **gares@fq.edu.uy**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Tórtora G. & Ares G. (2018). Influence of time orientation on food choice: Case study with cookie labels. *Food Research International*, *106*, 706-711. Available at **https://doi.org/10.1016/j.foodres.2018.01.045**

Abstract

Time orientation can influence health-related behaviors, including food consumption. The aim of the present work was to study the influence of time orientation on food choice, using cookie labels as case study. A choice-conjoint task was designed using labels differing in type of cookie (chocolate chips vs. granola), front-of-pack nutrition information (nutritional warnings vs. Facts Up Front system) and nutritional claim (no claim vs. "0% cholesterol. 0% trans fat"). An online study was conducted, in which 155 participants evaluated 8 pairs of cookie labels and selected the one they would buy if they were in the supermarket. Then, they were asked to complete a consideration of future consequences scale (CFC) adapted to eating habits, as well as a questionnaire about socio-demographic characteristics. Time orientation influenced participants' choices of cookies labels; particularly the importance attached to type of cookie. Participants with greater consideration of future consequences preferred the granola cookies, associated with health, while those who prioritized immediate consequences preferred chocolate chip cookies. In addition, nutritional warnings discouraged choice regardless of participants' time orientation. Results from the present work provide additional evidence of the influence of time preferences on food choices and suggest that strategies to stimulate and generate a more future-oriented perspective on eating habits could contibute to more healthful food choices.

Keywords: consideration of future consequences; food choice; time orientation; time preferences; eating habits; labels; nutritional warnings.

Introduction

Overweight and obesity are one of the most important health problems worldwide (World Health Organization, 2017). They are associated with an increased risk of non-communicable diseases, and increasing health care and social costs (Abajobir et al., 2017; World Health Organization, 2014). Considering that eating habits are one of the main causes of this situation, public policies and interventions aimed at promoting more healthful food choices have been extensively reccommended (Beaglehole et al., 2011; World Health Organization/Food and Agriculture Organization, 2003). To this end, it is crucial to understand the motives that underlie food chioces and particulary why people make unhealthul food choices (Sheperd, 2006).

The current food market is characterized by the high variety and availability of food products (Rozin, 2005). Within a category, there is a huge variety of products that differ in price, brand, quality, taste, packaging, labels, nutritional characteristics, among others. On a daily basis, consumers must make decisions about which foods to consume, and these choices involve trade-offs between immediate consequences (e.g., the pleasure of eating a tasty food) and future consequences (e.g., adverse health effects) (van Beek, Antonides & Handgraaf, 2013).

Consideration of future consequences and food choices

Individual differences in consumers' motivations and personal interests strongly influence their food choices (Köster, 2009). In particular, there are marked individual differences in the extent to which people consider future outcomes of their current behavior. Some individuals consider the future consequences of their actions and believe that certain behaviors are worthwhile because of their future desirable results, even though this entails a certain cost and effort (Strathman, Gleicher, Boninger, & Edwards, 1994). On the other hand, there are individuals who focus on maximizing the immediate benefits of their actions by giving them a high priority, without worrying about future consequences that may arise (Strathman et al., 1994). These differences between individuals are referred to as *time orientation*, which is defined as a preference towards past, present or future thinking, which influences behaviour and thought (Hulbert & Lens, 1988). This general concept is closely related to other concepts such as *temporal discounting* (Frederick, Loewenstein & O' Donoghue, 2002), *time preferences* (Bishai, 2001), *time perspective* (Zimbardo & Boyd, 1999) and *consideration of future consequences* (Strathman et al., 1994).

The consideration of future consequences (CFC) has been defined as "*the extent to which people consider the potential distant outcomes of their current behaviors and the extent to which they are influenced by these potential outcomes*" (Strathman et al., 1994). These authors developed the CFC scale to evaluate this construct, which has been applied in multiple studies that relate temporal orientation to several important domains such as work behavior, environmental and financial decision making, health behavior (Joireman & King, 2016). The CFC scale has been shown to be a predictor of various health-related behaviors such as smoking and alcohol consumption (Adams & Nettle, 2009; Beenstock, Adams, & White, 2011), doing physical exercise (Adams & Nestle, 2009), sleeping habits (Peters, Joireman, & Ridgway, 2005), getting vaccinated against the H1N1 virus (Nan & Kim, 2014), and sexual behavior (Appleby et al., 2005).

More recently, the CFC scale has been applied to eating behaviour (Dassen et al., 2015, Joireman, Shaffer, Balliet, & Strathman, 2012; Lawless, Drichoutis & Nayga Jr, 2013; van Beek et al., 2013). These studies have shown that time orientation can affect daily decisions about food consumption, suggesting that it can provide valuable insights to undertand individual differences in eating patterns. Making the decision to maintain a healthful diet often involves a greater expenditure of time, money and effort, as well as being willing to avoid some food products that provide immediate gratification (Huston & Finke, 2003). Therefore, CFC may play a key role in determining individuals' willingness to have a healthful diet. It can be hypothesized that people who consider the future consequences of their decisions may have a more healthful diet than those who are oriented towards the immediate consequences of their decisions (Huston & Finke, 2003). In addition, Cavaliere, De Marchi & Banterle (2014) recognized time orientation as a crucial aspect in explaining weight gain or maintaining a healthy weight condition. These authors reported that healthy body mass index was associated with a high orientation towards the future, with a high interest in nutritional information, a high level of education and a low attention to health-related claims.

Despite widespread recognition of the importance of CFC and time orientation in relation to health, few studies have investigated its influence on the choice of specific products. Recently, De Marchi, Caputo, Nayga Jr & Banterle (2016), demonstrated that those individuals who consider the future consequences of their actions show greater interest in organic foods and low calorie content when making their food choices than those with greater orientation to the present, who prefered taste or other characteristics which offer a more immediate gratification.

Food labels and food choice

Food labels are an important marketing strategy for food companies, attracting consumers' attention at the time of purchase and providing information that influences their perceptions and buying decisions (Ares et al., 2013). Within a category, labels make it possible to differentiate products by brand, industry or manufacturer, ingredients, net content, graphic design, and nutritional information. The graphic design of the label plays a key role in generating sensory and hedonic expectations about the product, which affect consumers' decisions both consuciously and subconsciously (Ares et al., 2011; Deliza & MacFie, 1996; Moskowitz, Porretta, & Silcher, 2005).

In addition to the compulsory information included on labels, food companies often use various types of nutritional and health claims as a marketing strategy to convey the concept of healthfulness (Lähteenmäki, 2013). These claims have been repoted to generate a healthy perception in some food products that have an unfavorable nutritional profile, as their precise and concrete wording often discourages consumers from looking for further nutritional information (Williams, 2005).

Nutritional information is one of the strategies that can be used to encourage consumers to make healthful food choices (Drichoutis, Lazaridis, & Nayga, 2009). However, conventional or back-of-pack nutrition information is often difficult to find and to understand for consumers, who do not usually take it into account when making their purchasing decisions (Grunert, Fernández-Celemín, Wills, Storcksdieck genannt Bonsmann & Nureeva, 2010). For this reason, the inclusion of front-of-pack (FOP) nutrition information has received great attention in recent years (Hawley et al., 2013).

In order to quickly capture consumers' attention and facilitate the understanding of nutritional information, simple formats schemes have been developed (EUFIC, 2017). Nutritional warnings are one of the most recent FOP nutrition labelling schemes aimed at discouraging consumers to purchase unhealthful products (Corvalán, Reyes, Garmendia, & Uauy, 2013) This scheme has been recently implemented in Chile, where packages should include octagonal black signs with the expression "High in" if the content of calories, sugars, saturated fat and sodium exceeds the established limits (Ministerio de Salud, 2015). These warnings have been reported to facilitate the identification of unhealthful products and to discourage their consumption (Arrúa, Machín et al., 2017; Arrúa, Curutchet et al., 2017; Ministerio de Salud, 2016).

Objectives and general description of the study

The aim of the present work was to study the influence of consideration of future consequences on food choices, using cookie labels as case study. This product category was chosen considering its frequent consumption by Uruguayan consumers and the wide range of products available in the marketplace, which stress either their sensory characteristics or health-related aspects. A choice-conjoint study using cookie labels was implemented online. Labels differed in the type of cookie, inclusion of nutritional warnings and nutrition claims. Participants were asked to evaluate a series of pairs of labels and to answer a consideration of future consequences scale, related specifically to eating habits (Beek et al., 2017).

Materials and methods

Participants A total of 155 participants aged between 18 and 60 years old from Montevideo (Uruguay) participated in the study. They were recruited through social networks. Participants were 84% women and their ages were distributed as follows: 46% from 18 to 25 years, 21% from 26 to 35 years, 15% from 36 to 45 years and 18% from 46 to 60 years. Participants filled an informed consent form before the study and did not receive any compensation.

Experimental design Cookie labels were designed using a 2^3 factorial design with three 2-level variables: type of cookie, nutrition claim and front-of-pack nutrition information. A summary of the variables and the levels considered in the study is shown in Table 1.

Two types of cookies were selected based on the characteristics of the products available in the Uruguayan market: granola and chocolate chip cookies. Based on the positioning of commercial products, granola cookies were expected to generate health associations, whereas chocolate chip cookies were expected to be associated with hedonics and immediate gratification. Type of cookie was communicated in the product name and label design, which included a picture of the cookies (Figure 1). In addition, the two types of cookies had different nutritional composition.

Front-of-pack nutrition information was included on the labels using the Facts Up Front system and nutritional warnings. The Facts Up Front system is being voluntarily included by food companies in several products commercialized in the Uruguay. In the present study, information about calories, sugars, fat and sodium was included. Nutritional warnings were included given that they are currently considered for their compulsory incorporation in the country (Ministerio de Industria, Energía y Minería, 2017). Black octogonal signs for high sugar and total fat content were included (Figure 1).

Finally, the inclusion of nutritional claims was considered due to its frequent use by food companies to communicate product healthfulness. The most frequently used expressions in commercial products was used in the present study "0% cholesterol, 0% trans fat".

Table 1. Variables and levels considered in the design of the cookie labels

Variables	Levels	
Type of cookie	Chocops	Granolinas
Nutrition claim	No claim	0% cholesterol, 0% trans fat
Front-of-pack nutrition information	Nutritional warnings	Facts Up Front system

Using the 8 cookie labels (corresponding to all possible combinations of the levels of the three considered variables), 8 sets of 2 labels were designed following the mix-and-match procedure (Johnson, Kanninen, Bingham & Ozdemir, 2007). The labels were designed by a graphic designer with previous experience in the design of labels of this type of food. To ensure that participants saw the labels for the first time during the experiment and to prevent prior knowledge from influencing decision-making, label design did not correspond to cookies currently available in the Uruguayan market. The position of the labels within the set was balanced. Figure 1 shows an example of one of the choice sets.

Figure 1. Example of a choice set involving granola cookies with the Facts Up Front system and no nutritional claim (above) and chocolate chip cookies with nutritional warnings for sugar and total fat and the nutritional claim “0% cholesterol, 0% trans fat” (below).

Experimental procedure A questionnaire composed of three parts was administered online. The first part consisted of a choice experiment using the 8 sets of pairs of cookie labels. The sets were presented one by one, following Latin squares experimental design. Participants were instructed to look at each of the labels and to indicate which of the two labels they would buy if they were in a supermarket.

In the second part of the study, participants completed an adapted versión of the consideration of future consequences scale specific for food habits (van Beek et al., 2013). This scale is used to evaluate the way in which people consider the most distant or proximal consequences of their eating habits. The main advantage of this domain-specific scale is that it is tailored to the type of behavior under investigation, eating behaviour, which then would result in better predictive capacity (Beek et al., 2017). The 14 items of the CFC-Food scale are shown in Table 2. The items were translated to Spanish by the authors and checked by a professional translator. Participants were asked to rate each of the items using a 7-point Likert scale (1 = “very uncharacteristic of me” and 7 = “very characteristic of me").

Finally, participants provided some personal data: age, gender, educational level, occupation, consumption frequency of cookies, type of cookies usually consumed and frequency with which they were in charge to make the purchases at home. The questionnaire was administered in Spanish using Compusense Cloud software (Compusense Inc., Guelph, Canada). The total duration of the task was approximately 15 minutes.

Data analysis

Exploratory factor analysis of CFC-Food scale An exploratory factor analysis of the CFC-Food scale was performed considering the maximum likelihood and promax rotation method. Two factors were identified, as expected: one related to future consequences (CFC-Future) and one related to immediate consequences (CFC-Immediate). The number of factors was determined considering the scree plot and the proportion of explained variance of each individual factor (at least 5%) (Cattell, 1966). Items with loadings above 0.40 onto one factor were considered (Hair, Anderson, & Tatham, 1987). In the case of items with loadings higher than 0.4 in more than one factor, a minimum difference of 0.15 was required (Worthingon & Whittaker, 2006)

For each of the factors identified in the scale, the sum of the scores of the items belonging to each factor was calculated for each participant. The analysis was performed using the psych package in R language (R Core Team, 2016).

Choice conjoint task The data of the choice experiment was analyzed using a logistic model with random parameters (Scarpa, Ferrini & Willis, 2005). The model considered the main effect of the variables included in the experimental design: type of cookie, FOP nutrition information and nutritional claim, as well as their interactions with the two factors of the CFC-Food scale (CFC-Future and CFC-Immediate) observed for each participant. This model allows to evaluate the influence of the attributes that characterized the cookie labels in participants’ choices (main effects) and how their influence varied according to the future or present orientation of each participant (interaction effects). The analysis was performed using the mlogit package in R language (R Core Team, 2016).

Cluster analysis Hierarchical cluster analysis was used on the participants’ scores in each of the factors of the CFC-Food scale to identify groups of participants with different time orientation. Euclidean distances and the Ward agglomeration method were considered. The choices of each group of participants were analysed separately using the logistic model described in the previous section considering only the main effects. Differences in the sociodemographic characteristics of the participants were evaluated using the chi-square test and analysis of variance.

Results

Exploratory factor analysis of the CFC-Food Scale The exploratory factor analysis carried out on data from the CFC-Food scale identified two factors (Table 2). One of the factors, CFC-Future, was related to 8 items associated with future consequences (1, 2, 6, 7, 8, 12, 13 and 14), whereas the other, CFC-Immediate, was related to 6 items associated with immediate consequences (3, 4, 5, 9, 10 and 11) (Table 2). Although item 12 presented cross-loading (factor loading higher than 0.4 on both factors), the difference between the loadings was higher than the threshold (Worthington & Whittaker, 2006), it was retained on the CFC-Future factor. However, this item has been previously related to the CFC-Immediate factor (Beek et al., 2017).

The sum of the scores of the items related to each of the factors of the scale was calculated to estimate the consideration of future and immediate consequences related to eating behaviour of each of the participants.

Table 2. Results of the exploratory factor analysis performed on the consideration of future consequences (CFC) scale, adapted to eating behavior. Factor loadings of each of the items on the two identified factors.

Items	CFC-Future	CFC-Immediate
1. I consider how my health might be in the future, and try to influence my health with my day to day eating behavior.	**0.70**	-0.24
2. Often I engage in a particular eating behavior in order to achieve outcomes that may not result for many years.	**0.69**	-0.10
3. I only choose my food to satisfy immediate needs, figuring the future will take care of itself.	-0.07	**0.77**
4. My eating behavior is only influenced by the immediate (i.e., a matter of days or weeks) consequences of my actions.	-0.07	**0.77**
5. My convenience is a big factor in the food I choose or my eating behavior.	-0.18	**0.43**

Note: For each item, the factor with the highest factor load is indicated with bold.

Table 2 (cont). Results of the exploratory factor analysis performed on the consideration of future consequences (CFC) scale, adapted to eating behavior. Factor loadings of each of the items on the two identified factors.

Items	CFC-Future	CFC-Immediate
6. I am willing to sacrifice the immediate happiness or wellbeing I derive from my eating behavior in order to achieve future outcomes.	**0.62**	0.00
7. I think it is important to take warnings about negative consequences of my eating behavior seriously even if the negative consequence will not occur for many years.	**0.69**	-0.04
8. I think it is more important to perform eating behavior with favorable distant consequences than eating behavior with less favorable immediate consequences.	**0.66**	0.01
9. I generally ignore warnings about possible future consequences of my eating behavior because I think they will be resolved before they reach crisis level.	-0.01	**0.78**
10. I think that sacrificing particular food now is usually unnecessary because future outcomes can be dealt with at a later time.	-0.21	**0.52**
11. I only choose my food to satisfy immediate needs, figuring that I will take care of future problems that may occur at a later date.	-0.05	**0.82**
12. Because my day to day eating behavior has specific consequences, it is more important to me than behavior that has distant consequences.	**0.63**	0.45
13. When I choose my *food* I think about how it might affect me in the future.	**0.76**	-0.18
14. My *eating behavior* is generally influenced by future consequences.	**0.83**	-0.08

Note: For each item, the factor with the highest factor load is indicated with bold.

Influence of CFC-Food on cookie label's choices A logistic model was applied to estimate the effect of the characteristics of cookie labels, as well as the influence of the consideration of future and immediate consequences related to eating behaviour, on participants' choices. As shown in Table 3, choices were significantly influenced by the variables type of cookie ($p<0.01$) and FOP nutrition labeling ($p<0.05$). As shown, on average participants preferred chocolate chip cookies over granola cookies. Meanwhile, the inclusion of nutritional warnings on the labels discouraged consumers' choice compared to the Facts Up Front system.

An interaction effect between the two factors of the CFC scale adapted to eating behaviour, CFC-Future and CFC-Immediate, and type of cookie was observed, indicating that future and present time orientation affected the importance attached to the type of cookie. The significant and negative coefficient of the effect of the interaction between type of cookie and CFC-Future, suggests that participants with greater future orientation in relation to their eating behaviour presented a lower preference for chocolate chip cookies than less future-oriented participants. On the contrary, the significant and positive coefficient of the interaction effect between type of cookie and the CFC-Immediate factor, points out that participants who focused more on the immediate consequences of their eating behaviour, had a higher preference for chocolate chip cookies (vs. granola cookies).

Table 3. Results of the effect of each variable on the cookies' choices, in relation to the CFC-Food Scale.

Main effect	Coefficient
Type of cookie (chocolate)	2.21**
Front-of-pack nutrition informaiton (warnings)	-1.70*
Nutrition claim (0% Cholesterol, 0% Trans Fat)	0.65
Interaction effects	
Front-of-pack nutrition information: CFC-F	-0.01
Type of cookie: CFC-F	-0.10***
Nutrition claim: CFC-F	0.00
Front-of-pack nutrition information: CFC-I	0.03
Type of cookie: CFC-I	0.10***
Nutrition claim: CFC-I	-0.00

Note: The significance of the coefficients is indicated by the following codes: ns $p>0.05$; * $p<0.05$; ** $p\leq 0.01$; *** $p<0.001$

Identification of groups of participants with different CFC score Cluster analysis was performed on the scores of the CFC-Future and CFC-Immediate factors to identify groups of participants with different time orientation. As shown in Table 4, two groups of participants were identified. Group 1, composed of 83 participants, had higher scores on the CFC-Future factor and lower scores on the CFC-Immediate factor, which indicates a greater consideration of future consequences of their eating behaviour. Group 2, composed of 72 participants, had higher scores on the CFC-Immediate factor and lower scores on the CFC-Future factor, indicating a greater consideration of the immediate consequences of their eating behaviour of these participants. The groups did not differ in gender ($\chi^2 = 0.00$, $p = 0.96$), age ($\chi^2 = 6.80$, $p = 0.07$), educational level ($\chi^2 = 6.29$, $p = 0.39$), or in the consumption frequency of cookies ($\chi^2 = 2.92$, $p = 0.40$).

The choices of the two groups were analyzed separately. In the case of Group 1, composed of future-oriented participants regarding their eating behaviour, the three variables significantly affected their choices. The significant and negative coefficient of the variable type of cookie indicates that these participants preferred granola cookies over choocolate cookies. In addition, the warning system decreased likelihood of choosing a label, whereas nutritional claim increased likelihood of choosing a label. The coefficients of the variables FOP nutrition labeling and type of cookie were similar, suggesting that they had a smilar influence on the choices of participants in Group 1.

For participants in Group 2, characterized by a greater orientation towards the present regarding their eating behaviour, the three variables also had a significant effect on their choices. Although the warning system discouraged the choices and the nutritional claim stimulated them, differences from group 1 were found regarding the influence of they variable type of cookie. This variable had the largest coefficient (Table 4), suggesting that it was the most important variable for their choices. In addition, the negative coefficient indicates that they preferred the chocolate chip cookies over the granola cookies.

Discussion

Understanding the influence of individual differences in time orientation on daily food choices can allow the use of knowledge about time orientation to positively influence public health (Lawless et al., 2013). Results from the present work added empirical evidence about the influence of the consideration of future consequences on food choice. Using cookie labels as case study, results showed that time orientation modulated the influence of product characteristics on the choice of specific products, in agreement with results reported by De Marchi et al. (2016). A greater consideration of future consequences in relation to eating behaviour was associated with preference for granola cookies over chocolate chip cookies, whereas greater consideration of immediate consequences was associated with preference of chocolate chip cookies. This difference can be explained considering that people who usually consider the future consequences of their decisions selected the cookies with healthy image, whereas those who mainly take into account their immediate consequences of their food-related decisions preferred the cookies with the greatest hedonic connotation. In addition, groups of consumers with different time orientation gave different relative importance to product characteristics. Consumers with a temporal orientation towards the present attached the greatest importance to the type of cookie for making their food choices, whereas those with a temporal orientation towards the future gave similar relative importance to the type of cookie and front-of-pack nutrition information. These results are consistent with the results reported by Dassen et al. (2015), which indicate that people who are focused on the immediate consequences of their eating behaviour report a less healthy diet, while those who are more concerned about health in the future report a more healthful diet. Similar results have been reported by De Marchi et al. (2016), Lawless et al. (2013), Huston & Finke (2003) and Cavaliere et al. (2014).

The inclusion of nutritional warnings on the labels discouraged consumer choice, which confirms the potential of this front-of-pack nutrition labelling scheme to discourage consumption of unhealthful products. Similar results have been reported by Arrúa, Curutchet, et al. (2017) for children's choice of waffle cookies and orange juice. Interestingly, nutritional warnings discouraged consumer choice regardless of their time orientation, which suggests that they could be an efficient public policy to improve the eating patterns of the population. Simplified nutrition labelling schemes have been reported to encourage people to take into account nutrition information in their decision making process, regardless of their thinking style (Mawad, Trías, Giménez, Maiche & Ares, 2015; Ares, Mawad, Giménez, & Maiche, 2014). In this sense, preliminary studies conducted one year after the implementation of nutritional warnings in Chlie report that 92% of the consumers take nutritional warnings into account when making their food choices (Valdebenito Verdugo, Labrín Elgueta, León Perath & Fierro Kalbhenn, 2017).

In the present work, a domain-specific version of the consideration of future consequences scale was used to evaluate the influence of temporal orientation on food choice. This is the first work to consider such scale in the context of food choice as previous studies have used the general version of the scale. Although a food-related CFC scale may often a more accurate estimation of temporal orientation in the context of food choices, it should be taken into account that it is still not clear what participants have in mind when they think of the future or inmediate consequences of their eating behaviour (van Beek et al., 2013). For example, when thinking of the future consequences of their eating behavior they can imagine either positive or negative consequences on their health. Further research should explore in depth how consumers conceptualize the future consequences of their eating behavior and how they understand the items of the scale.

Limitations of the study The main limitation of the present work was the relatively small and non-representative sample of participants. Further research should extend the findings of the present work by using a larger sample and looking at the relationship between CFC related to eating behaviour and sociodemographic variables, such as gender, age and socio-economic status. Moreover, the present study used cookie labels as model foods, which are usually regarded as an unhealthful product, and the choices were made in a hypothetical context as consumers did not have to consume the cookies they selected. Further research could include other types of foods and involve realistic consumption contexts.

Conclusions

The present study showed the existence of an interaction between the consideration of the future consequences and consumers' choice of cookie labels, providing additional empirical evidence that time orientation influences food consumption. Using the consideration of future consequences scale adapted to eating behavior, it was shown that those consumers who consider the future consequences of their eating behaviour tend to make more healthful choices than those who focus on the immediate consequences of this behavior. These results suggest that time orientation may be an interesting target for intervention, in agreement with Dassen et al. (2015). Interventions targeted at promoting future orientation, such as stimulating future episodic thinking, could encourage to more healthful eating behaviours and, as a consequence, contribute to improve the health status of the population (Daniel, Stanton & Epstein, 2013).

Acknowledgments

The authors are indebted to Espacio Interdisciplinario (Universidad de la República, Uruguay) for financial support.

References

Abajobir, A.A., Abate, K.H., Abbafati, C., Abbas, K.M., Abd-Allah, F., Abdulkader, R.S., et al. (2017). Measuring progress and projecting attainment on the basis of past trends of the health-related Sustainable Development Goals in 188 countries: an analysis from the Global Burden of Disease Study 2016. *The Lancet*, 390, 1423 – 1459.

Adams, J., & Nettle, D. (2009). Time perspective, personality and smoking, body mass, and physical activity: An empirical study. *British Journal of Health Psychology,* 14, 83–105.

Appleby, P. R., Marks, G., Ayala, A., Miller, L. C., Murphy, S., & Mansergh, G. (2005). Consideration of future consequences and anal intercourse among men who have sex with men. *Journal of Homosexuality*, 50, 119–133.

Ares, G., Giménez, A., Bruzzone, F., Vidal, L., Antúnez, L., & Maiche, A. (2013). Consumer visual processing of food labels: results from an eye−tracking study. *Journal of Sensory Studies*, *28*(2), 138-153.

Ares, G., Piqueras-Fiszman, B., Varela, P., Marco, R. M., López, A. M., & Fiszman, S. (2011). Food labels: Do consumers perceive what semiotics want to convey? *Food Quality and Preference*, 22(7), 689-698.

Ares, G., Mawad, F., Giménez, A., & Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference,* 31, 28–37.

Arrúa, A., Curutchet, M.R., Rey, N., Barreto, P., Golovchenko, N., Sellanes, A., et al. (2017). Impact of front-of-pack nutrition information and label design on children's choice of two snack foods: Comparison of warnings and the traffic-light system. *Appetite*, 116, 139-146.

Arrúa, A., Machín, L., Curutchet, M. R., Martínez, J., Antúnez, L., Alcaire, F., et al. (2017). Warnings as a directive front-of-pack nutrition labelling scheme: comparison with the Guideline Daily Amount and traffic-light systems. *Public Health Nutrition*, 20, 2308-2317.

Beaglehole, R., Bonita, R., Horton, R., Adams, C., Alleyne, G., Asaria, P., ... Cecchini, M. (2011). Priority actions for the

non-communicable disease crisis. *The Lancet*, *377*(9775), 1438-1447.
Beek, J., Handgraaf, M. J., & Antonides, G. (2017). Time orientation and construal level: effects on eating and exercising behaviour and preferences. *International Journal of Consumer Studies*, *41*(1), 54-60.
Beenstock, J., Adams, J., & White, M. (2011). The association between time perspective and alcohol consumption in university students: Cross-sectional study. *The European Journal of Public Health, 21*, 438–443.
Bishai, D. (2001). Life-cycle changes in the rate of time preference: Testing the theory of endogenous preferences and its relevance to adolescent substance abuse. *Economic Analysis of Substance Abuse*, 155-77.
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research,* 1, 245–276.
Cavaliere, A., De Marchi, E., & Banterle, A. (2014). Healthy–unhealthy weight and time preference. Is there an association? An analysis through a consumer survey. *Appetite, 83*, 135–143.
Corvalán, C., Reyes, M., Garmendia, M.L., Uauy, R. (2013). Structural responses to the obesity and non-communicable disease epidemic: The Chilean law of food labelling and advertising. *Obesity Reviews,* 14, 79-87.
Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013). The future is now: reducing impulsivity and energy intake using episodic future thinking. *Psychological science*, *24*(11), 2339-2342.
Dassen, F. C., Houben, K., & Jansen, A. (2015). Time orientation and eating behavior: Unhealthy eaters consider immediate consequences, while healthy eaters focus on future health. *Appetite*, *91*, 13-19.
De Marchi, E., Caputo, V., Nayga Jr., R.M., & Banterle, A. (2016). Time preferences and food choices: Evidence from a choice experiment. *Food Policy, 62*, 99-109.
Deliza, R., & MacFie, H. J. H. (1996). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: A review. *Journal of Sensory Studies,* 11, 103–128.
Drichoutis, A. C., Lazaridis, P., & Nayga Jr, R. M. (2009). On consumers' valuation of nutrition information. *Bulletin of Economic Research*, *61*(3), 223-247.
EUFIC. (2017). Global Update on Nutrition Labelling. The 2017 edition. Brussels: EUFIC.
Frederick, S., Loewenstein, G., y O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351-401.
Grunert, K. G., Fernández-Celemín, L., Wills, J. M., Storcksdieck genannt Bonsmann, S., y Nureeva, L. (2010). Use and understanding of nutrition information on food labels in six European countries. *Journal of Public Health*, 18(3), 261–277.
Hair, J. F., Anderson, R. E., & Tatham, R. L. (1987). *Multivariate analysis with readings* (2nd ed.). London: Macmillan Publishing Company.
Hawley, K.L., Roberto, C.A., Bragg, M.A., Liu, P.J., Schwartz, M.B., & Brownell, K.D. (2013). The science on front-of-package food label. *Public Health Nutrition*, 16, 430-439.
Hulbert, R. J., & Lens, W. (1988). Time perspective, time attitude, and time orientation in alcoholism: A review. *International Journal of the Addictions,* 23(3), 279-298.
Huston, S. J., & Finke, M. S. (2003). Diet choice and the role of time preference. *Journal of Consumer Affairs,* 37(1), 143-160.
Johnson, F. R., Kanninen, B., Bingham, M., & Ozdemir, S. (2007). Experimental design for stated-choice studies. In K. B. Kanninen (Ed.), *Valuing environmental amenities using stated choice studies* (pp. 159–202). Dordrecht: Springer.
Joireman, J., Shaffer, M., Balliet, D., & Strathman, A. (2012). Personality promotion orientation explains why future-oriented people exercise and eat healthy: Evidence from the two-factor Consideration of Future Consequences-14 Scale. *Social Psychology Bulletin, 38*, 1272–1287.
Joireman, J., & King, S. (2016). Individual differences in the consideration of future and (more) immediate consequences: A review and directions for future research. *Social and Personality Psychology Compass*, 10(5), 313-326.
Köster, E. P. (2009). Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference*, *20*(2), 70-82.
Lähteenmäki, L. (2013). Claiming health in food products. *Food Quality and Preference,* 27, 196-201.
Lawless, L., Drichoutis, A. C., & Nayga Jr, R. M. (2013). Time preferences and health behaviour: a review. *Agricultural and Food Economics*, 1(1), 1-19.
Mawad, F., Trías, M., Giménez, A., Maiche, A., Ares, G. (2015). Influence of cognitive style on information processing and selection of yogurt labels: Insights from an eye-tracking study. *Food Research International,* 74, 1-9.
Ministerio de Industria, Energía y Minería (2017). *Consulta pública sobre proyecto de decreto relativo a Rotulado de alimentos envasados*. Available at: http://www.miem.gub.uy/-/consulta-publica-sobre-proyecto-de-decreto-relativo-a-rotulado-de-alimentos-envasados (Accessed August 17th, 2017).
Ministerio de Salud. (2015). *Decreto número 13, de 2015.* Santiago de Chile: Biblioteca del Congreso Nacional de Chile.
Ministerio de Salud. (2016). Ley de Alimentos – Nuevo etiquetado de alimentos. Santiago de Chile: Ministerio de Salud. Available at: http://web.minsal.cl/ley-de-alimentos-nuevo-etiquetado-de-alimentos/
Moskowitz, H. R., Porretta, S., & Silcher, M. (2005). Concepts as combinations of graphics. In H. R. Moskowitz (Ed.), *Concept research in food product design and development* (pp. 105–121). Ames: Blackwell Publishing.
Nan, X., & Kim, J. (2014). Predicting H1N1 vaccine uptake and H1N1-related health beliefs: The role of individual difference in consideration of future consequences. *Journal of Health Communication*, 19, 376–388.
Peters, B.R., Joireman, J., & Ridgway, R. L. (2005). Individual differences in the consideration of future consequences scale correlate with sleep habits, sleep quality, and GPA in university students. *Psychological Reports,* 96, 817–824.
R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rozin, P. (2005). The meaning of food in our lives: a cross-cultural perspective on eating and well-being. *Journal of Nutrition Education and Behavior,* 27, S107-S112.
Scarpa, R., Ferrini, S., & Willis, K. (2005). Performance of error component models for status-quo effects in choice experiments. In: R Scarpa, A Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics* (pp. 247–273). Dordrecht, The Netherlands: Springer Publisher.
Sheperd, R. (2006). The use of the stages of change model with dietary behaviours. In: R Sheperd, M Raats (Eds.), *The psychology of food choice* (pp.345-356). Wallingford: CABI.
Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*, 742-752.
Valdebenito Verdugo, M., Labrín Elgueta, J.M., León Perath, V., & Fierro Kalbhenn, S. (2017*). Informe de resultados: Descripción de las percepciones y actitudes de los/as consumidores respecto a las medidas estatales en el marco de la implementación del Decreto 13/15.* Santiago de Chile: Demoscopica.
van Beek, J., Antonides, G., & Handgraaf, M. J. (2013). Eat now, exercise later: The relation between consideration of immediate and future consequences and healthy behavior. *Personality and Individual Differences*, *54*(6), 785-791.
Williams, P. (2005). Consumer understanding and use of health claims for foods. *Nutrition reviews*, *63*(7), 256-264.
World Health Organization. (2014). *Global status on noncommunicable diseases 2014.* Geneva: World Health Organization
World Health Organization. (2017). Obesity and overweight. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/ (Last accessed October 16th, 2017).
World Health Organization/Food and Agriculture Organization (WHO/FAO). (2003). *Expert report: diet, nutrition and prevention of chronic diseases. Report of a Joint WHO/FAO*

Expert Consultation. WHO Technical Report Series 916. Rome: WHO/FAO.

Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist,* 34, 806-838.

Zimbardo, P.G & Boyd, J. (1999). Putting Time in Perspective: A valid, reliable, individual-differences metric. *Journal of Personality and Social Psychology,*

Taza decorativa de metáfora temporal
Time metaphor's decorative mug

Autores: **Maria Noel Macedo** y **Jhonny Rodriguez.** Financiación: Programa de Apoyo a la Investigación Estudiantil (PAIE): *Activación de la metáfora temporal en verbos con esquemas de imagen horizontal y vertical,* Universidad de la República, Uruguay. La intención en este estudio fue evaluar si existe algún tipo de interacción entre la representación imaginística de los verbos (p. e., CORRER como un verbo que se puede representar con un movimiento horizontal hacia adelante) y la de la ubicación de las acciones o estados que ellos representan en una línea espacial del tiempo izquierda-derecha. El estudio confirmó la asignación normativa de los rasgos esquemáticos e imaginísticos de los verbos estudiados y su ubicación en una línea mental del tiempo que se desplaza de izquierda a derecha para hablantes de español**.**

Authors: **Maria Noel Macedo** and **Jhonny Rodriguez.** Funding: Programa de Apoyo a la Investigación Estudiantil (PAIE): *Activación de la metáfora temporal en verbos con esquemas de imagen horizontal y vertical,* Universidad de la República, Uruguay. The intention in this study was to evaluate whether there is any kind of interaction between the imagistic representation of verbs (e.g., RUN as a verb that can be represented with a horizontal forward movement) and that of the location of the actions or states they represent in a left-right spatial time line. The study confirmed the normative assignment of the schematic and imagistic features of the verbs studied and their location on a mental time line moving from left to right for Spanish speakers.

Time Attitude Profiles and Health-Related Behaviors: Validation of Scores on a Spanish Version of the Adolescent and Adult Time Inventory–Time Attitudes (AATI-TA)

Alejandro Vásquez-Echeverría[1], Lucía Alvarez-Nuñez[1], Zena R. Mello[2], Frank C. Worrell[3]

Alejandro Vásquez-Echeverría https://orcid.org/0000-0002-9538-4857, Lucía Alvarez-Nuñez https://orcid.org/0000-0003-1579-8759, Zena R. Mello https://orcid.org/0000-0001-8218-9801, Frank C. Worrell https://orcid.org/0000-0002-7122-527X

1 Instituto de Fundamentos y Métodos en Psicología. Facultad de Psicología. Universidad de la República, Tristán Narvaja 1674. CP 11200. Montevideo, Uruguay

2 Department of Psychology. San Francisco State University. 1600 Holloway Avenue. San Francisco, CA 94132

3 Graduate School of Education. 2121 Berkeley Way. University of California, Berkeley. Berkeley, CA 94720-1670

Correspondence concerning this manuscript should be addressed to Alejandro Vásquez-Echeverría, Facultad de Psicología, Tristán Narvaja 1674, Montevideo 11200, Uruguay. Telephone +5982400855 int 340. Email: **avasquez@psico.edu.uy**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Vásquez-Echeverría, A., Álvarez-Núñez, L., Mello, Z., & Worrell, F. C. (2020). Time attitude profiles and health-related behaviors: Validation of a Spanish version of the Adolescent and Adult Time Inventory–Time Attitudes (AATI-TA). *The Spanish Journal of Psychology*, *23,51, 1-14*. Available at **https://doi.org/10.1017/SJP.2020.51**

Abstract

Temporal psychology constructs are an individual difference variable related to behavioural outcomes. Recent research has shown that there are different time attitude profiles based on different configurations of the six Adolescent and Adult Time Inventory-Time Attitude (AATI-TA) subscales. The objective of this study was to analyze the psychometric properties of AATI-TA scores in Uruguay and determine the existence of temporal profiles in this context. Participants were a convenience sample of 446 (36.5% males) adults in Uruguay with a mean age of 34.53 years (*SD* = 13.17, range 18–75 years). Participants completed a sociodemographic questionnaire, the AATI-TA, and questionnaires on intentions, behaviours, and attitudes towards healthy food consumption and physical activity. AATI-TA scores had good reliabilities (> .70). The six-factor solution was supported and invariance by gender and age group was established. We identified five profiles – Resilients, High Positives, Negatives, Present Negatives, and Moderate Positives – which were associated differently with healthy food consumption patterns. Negative profiles were related to higher levels of unhealthy food consumption.

Keywords: time attitudes, time perspective, adults, latent profile analysis, health-related behaviours

Introduction

In the past two decades, there has been an increase in the research that considers individual differences in temporal psychology constructs as predictors of many outcomes, such as health-related behaviours, educational achievement, and pro-environmental behaviour (e.g., Andre et al., 2018). One key concept in this field is time perspective, defined as a multidimensional construct that comprises the thoughts, attitudes, and behaviours towards the three temporal regions of one's life (Zimbardo & Boyd, 1999). Although many researchers have focused on the future time period (e.g., Lang & Carstensen 2002), Zimbardo and Boyd (1999) developed the Zimbardo Time Perspective Inventory (ZPTI) to assess cognitive, emotional, motivational, and social aspects of the past (past positive, past negative), present (present hedonistic, present fatalistic), and future (focus on planfulness). Another innovation of the ZTPI was its focus on both positively (past positive, present hedonistic, future [planning]) and negatively (past negative, present fatalistic) valenced aspects of time. Since its introduction in 1999, the ZTPI has become the most widely used scale in the literature.

However, concerns have been raised about robustness of ZTPI scores. In early studies, some ZPTI subscale scores have yielded low internal consistency estimates (e.g., Keough et al., 1999; Zimbardo et al., 1997), and the hypothesized five-factor structure has not yielded acceptable fit indices (e.g., Zimbardo et al., 1997). Moreover, these types of psychometric issues have been found with regard to scores on multiple versions of the ZTPI for the past two decades (e.g., McKay, Worrell, et al., 2015; Milfont et al., 2008; Sircova et al., 2014; Temple et al., 2019; Worrell & Mello, 2007; Worrell, Temple, et al., 2018). Thus, although research with the ZTPI has indicated that time perspective is related to a range of psychological constructs and adaptive behaviors (e.g., McKay et al., 2014; McKay, Perry, et al., 2018), these findings can be questioned on the basis of the robustness of ZTPI scores.

The psychometric issues with ZTPI scores may be the result of (a) measuring multiple aspects of time (e.g., attitudes, behaviors, cognitions) on the same subscale, (b) including constructs other than time (e.g., hedonism, planning) on several subscales, or (c) both of these concerns. Nonetheless, the publication of the ZTPI had a profound impact on the time perspective literature and contributed to the development of several other instruments assessing all three time periods. One of these new instruments was the Adolescent Time Inventory (Mello & Worrell, 2007), which assesses time attitudes, the construct used in this study. The goals of the present study were (a) to assess the psychometric properties of time attitude scores and (b) to assess the association between time attitude scores and health related behaviours in a sample of Uruguayan adults.

The Adolescent and Adult Time Inventory – Time Attitudes Scale

The Adolescent Time Inventory (Mello & Worrell, 2007) is a measure that assesses five aspects of time perspective: time meaning (how do individuals define the past, present, and future), time frequency (how often do individuals think about the three time periods), time relation (how do individuals think about the relationship among the three time periods), time orientation (which time period or periods is most salient to an individual), and time attitudes (what are individuals' attitudes toward the three time periods). Time attitudes are the only aspect of the Adolescent Time Inventory that are conceived of as latent variables, and they are similar to time perspective as assessed by the ZTPI, including having scales with both positive and negative valence. In developing the Adolescent Time Inventory-Time Attitudes Scale (ATI-TA; Mello & Worrell, 2007), the authors tried to avoid the psychometric problems of the ZTPI (e.g., Zimbardo et al., 1997) by using items that assessed attitudes only and items only focusing on time (see Worrell et al., 2013). Originally developed to measure time attitudes in adolescents, ATI-TA scores were subsequently validated in adult samples (Cole et al., 2017; Mello et al., 2016) and the instrument was renamed the Adolescent and Adult Time Attitude Scale (AATI-TA). The AATI-TA assesses negative and positive attitudes toward the past, present, and future: Past Positive, Past Negative, Present Positive, Present Negative, Future Positive, and Future Negative.

The multi-stage development process used to develop the AATI-TA is documented in Worrell et al. (2013), where the English and German versions of the scale are introduced. The process began with over 200 items, which were eventually reduced to 53, 49, and ultimately 30 items – five items on each of the six subscales – based on theory and psychometric analyses. There is strong evidence in support of AATI-TA scores. Internal consistency estimates ≥ .70 have been consistently reported for five of the six AATI-TA scores, with future negative scores having lower estimates in some samples (Worrell et al., 2020). Additionally, evidence in support of the hypothesized six factor structure has been reported for samples from the United States (Mello et al., 2016; Worrell et al., 2013), Germany (Buhl & Lindner, 2009; Worrell et al., 2013), New Zealand (Alansari et al., 2013), the United Kingdom (Cole et la., 2017; McKay, Cole, et al., 2015), Turkey (Çelik et al., 2017; Şahin-Baltaci et al., 2017), Albania (Worrell et al., 2020), Nigeria (Mello et al., 2019), Italy (Donati et al., 2019; Worrell et al., 2020), and Japan (Chishima et al., 2019).

Unlike the ZTPI, the AATI-TA includes a future negative scale, which has been the most psychometrically challenging, with future negative scores raising concerns similar to those raised with some ZTPI scores. For example, future negative scores have yielded very low reliability estimates in Turkey, Albania, Italy, and Nigeria, but not in the US, Germany, New Zealand, Japan, and the UK. Scores on AATI-TA short versions (i.e., 24 items) used in Germany (Buhl & Lindner, 2009), Slovenia (Juriševič et al., 2017), and Spain (Konowalczyk et al., 2018) have yielded acceptable to close fit with most factor coefficients > .50 across studies and acceptable to strong internal consistency estimates, including for future negative (αs ≥ .75) scores.

Scores on the six-factor structure have shown scalar invariance between younger and older adolescents in Spain on the short version (Konowalczyk et al., 2018), between adolescents and young adults in Italy (Donati et al., 2019) using the full 30-item scale, and between young and middle-aged adults in the US for five scales (excluding future negative; Mello et al., 2016). The six-factor have also been shown scalar invariance by gender (Konowalczyk et al., 2018; Worrell, McKay, et al., 2018) and configural invariance over a 2-year period (Worrell, McKay et al., 2018). In short, reliability and structural validity evidence for the six AATI-TA subscale scores is generally strong to date, with less support for the future negative scores. Convergent validity evidence is reported in the next section, and as with scales of this type, establishing validity for scores is an ongoing endeavour (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014, 2018).

Person-Centered Analyses

In addition to introducing the ZTPI, Zimbardo and Boyd (1999) recommended another change – advocating for "temporal profiles (p. 1273) – which has had important implications for the field of temporal psychology. Indeed, one goal of having an instrument that assessed all three time periods was to "profit from the use of combined 'profile patterns' [i.e., person-centered analyses] of the five ZTPI factors instead of independent examination of ZTPI subscales [i.e., variable-centered analyses]" (Zimbardo & Boyd, 1999, p. 1284). Support for the utility of ZTPI profiles was provided in a study by McKay et al. (2014). Using a sample-specific short version of the ZTPI, McKay et al. (2014) identified four profiles – Balanced, Future, Past Negative, and Present Hedonistic – and found that adolescents in the Balanced and Future profiles were more likely to be abstainers and adolescents in the Past Negative and Present Hedonistic profiles were more likely to be problematic drinkers.

Researchers using the AATI-TA have also used both person-centered and variable-centered analyses. AATI-TA scores have been found to be associated with self-esteem, perceived stress, perspective-taking, and self-efficacy (Andretta et al., 2014; Buhl & Lindner, 2009; Cole et al., 2017) and several other constructs, providing evidence of convergent validity (Worrell & Mello, 2009). Research also shows that time attitude *profiles* are also predictive of outcomes. For instance, Buhl and Lindner (2009) applied latent profile analysis to the short version of AATI-TA scores and identified six temporal profiles in a sample of German adolescents. These profiles were associated with self-efficacy, perspective taking, and trust in schools, with positive profiles reporting higher scores on these variables than negative profiles. Similar time attitude profiles have been identified in the US, the UK, and New Zealand, and the research indicates that time attitude profiles are stronger and more reliable predictors of a variety of outcomes (educational constructs, cultural constructs, psychological constructs) than individual time attitude scores (e.g., Worrell & Andretta, 2019). Two objectives of the current study were to see (a) if interpretable time attitude profiles would be found in a sample of Uruguayan adult participants and (b) if these profiles would be related to differences in health behaviors.

Physical Exercise and Eating Behaviours

Since prevention of some chronic pathologies depends on behavioural change, temporal psychology constructs may play a key role in design of interventions and framing of public campaigns. Temporal psychology constructs have been documented to be related to health protective behaviours in many studies (Andre et al., 2018; Joireman et al., 2012; Murphy & Dockray, 2018), as health protective behaviours often posit a conflict between present and future rewards. Time attitudes have been found to be associated with perceived stress, hopelessness, anxiety, depression, alcohol use, and the agentic aspect of hope (Andretta et al., 2014; Buhl & Lindner, 2009; McKay, Perry et al., 2018; Worrell & Andretta, 2019). Whereas time attitude scores have yielded near-zero correlations with exercise, smoking, cannabis use, and visits to the dentist, time attitude *profiles* are associated with all of these behaviours (McKay, Andretta, et al., 2018). Konowalczyk et al. (2018) also found that time attitude scores were modestly associated with being a member of a sports clubs, sports significance, and self-concept related to physical ability and physical appearance.

To the best of our knowledge, no one has explored the association between time attitudes and attitudes toward healthy food consumption, and only McKay, Andretta et al. (2018) have looked at time attitudes and physical exercise. We were also interested in analyzing *intentions* to eat healthy and engage in physical exercise as well as the actual behaviours within the framework of planned action theory. Although attitudes are an important determinant of behavioural intention and actual behaviour, these three constructs are not interchangeable (Ajzen, 2005). We expected that time attitudes would have stronger correlations with attitudes toward health and healthy food consumption than with intentions and behaviours (see Andre et al., 2018, for a meta-analysis).

The Current Study

The current study is the first examination of AATI-TA scores in a Uruguayan sample, with the aim of assessing internal consistency evidence, structural validity evidence, and convergent validity evidence (AERA, APA, & NCME, 2014, 2018). We were also interested in seeing if interpretable time attitude profiles, found in other studies using the AATI-TA, would be found in Uruguay. In sum, we asked three questions in this study. First, we examined the internal consistency and structural validity of AATI-TA scores translated into Rio de la Plata Spanish in a sample of Uruguayan adults who were diverse in age and educational attainment. We hypothesized that AATI-TA scores would be reliable and the theorized 6-factor structure would yield acceptable fit estimates. We also hypothesized that the scores would be invariant across gender and developmental period as reported in Italy (Donati et al., 2019) and Spain (Konowalczyk et al., 2018).

Second, we assessed the existence of interpretable latent temporal profiles in this sample, as well as their similarities to profiles found in other samples. We hypothesized that we would find several profiles, with some with some being more adaptive than others with higher scores on the positive than the negative attitudes. Lastly, we explored how time attitude scores and time attitude profiles were related to healthy food consumption and physical exercise attitudes, intentions, and behaviours. Although we speculated that positive scores and profiles would be associated with healthier attitudes and behaviours, we did not make specific hypotheses as this question has not yet been addressed in the literature.

Methods

Participants and Procedures Participants consisted of 446 (36.5 % males; n = 163) adults in Uruguay with a mean age of 34.53 years (SD = 13.17, range 18–75 years) who completed the ATTI-TA. Two hundred and ten participants were young adults (aged 18–29 years) and 236 were middle or older adults (aged 30 or more); two participants did not report their age. A subsample of 328 adults (34.8% males; M_{age} = 33.97 years, SD = 13.32, range 18–75 years) also completed physical activity and healthy eating questionnaires. Participants were a convenience sample recruited by the researchers and their students in Montevideo and the metropolitan area of Uruguay. For the full sample, in terms of education, 13.5% had not completed high school education, 16.8% had a high school diploma, 39% were university students, 18.4% were university graduates, 5.1% had completed or were in advanced graduate studies, 5.6% had other tertiary education, and 1.6% did not report their educational level. Participants signed an informed consent in a manner that was approved by the local university's institutional

Table 1. Descriptive Statistics for AATI-TA Subscales

	Mean (SD)	α (95% CI)	ω^h	2	3	4	5	6
1. Past P.	3.93 (0.86)	.84 (.81, .86)	.84	-.61*	.31*	-.28*	.31*	-.30*
2. Past N.	2.03 (0.93)	.82 (.80, .85)	.83	-	-.31*	.44*	-.29*	.43*
3. Present P.	4.05 (0.81)	.84 (.82, .87)	.85		-	-.74*	.55*	-.42*
4. Present N.	2.12 (0.90)	.79 (.75, .82)	.79			-	-.46*	.45*
5. Future P.	4.07 (0.76)	.82 (.79, .85)	.82				-	-.59*
6. Future N.	1.70 (0.73)	.70 (.66, .75)	.71					-

Note. P. = Positive; N. = Negative; SD= Standard Deviation; ω^h = hierarchical omegas were computed using the coefficients of the six-factor model; *$p < .01$.

research review board and they completed the measures in a paper-and-pencil format during classes or at home.

Measures

In addition to self-reporting information on age, sex, and educational level, participants also completed questionnaires on time attitudes, health behaviours, and eating behaviours.

Time Attitudes As noted in the Introduction, the AATI-TA consists of six subscales that assess positive and negative attitudes toward the past (Past Positive, Past Negative), present (Present Positive, Present Negative), and future (Future Positive, Future Negative). The instrument has been translated into several different languages,

Table 2. Fit Indices Using WLMSV Estimator for Time Attitude Scores

Models	χ^2	df	CFI	TLI	RMSEA (90% CI)
2 factors (valence)	3240.36*	404	.787	.771	.125 (.121- .129)
3 factors (temporal)	1131.06*	402	.945	.941	.064 (.059- .068)
6 factors (theory)	766.39*	390	.971	.968	.047 (.042-.051)

*$p < .001$.

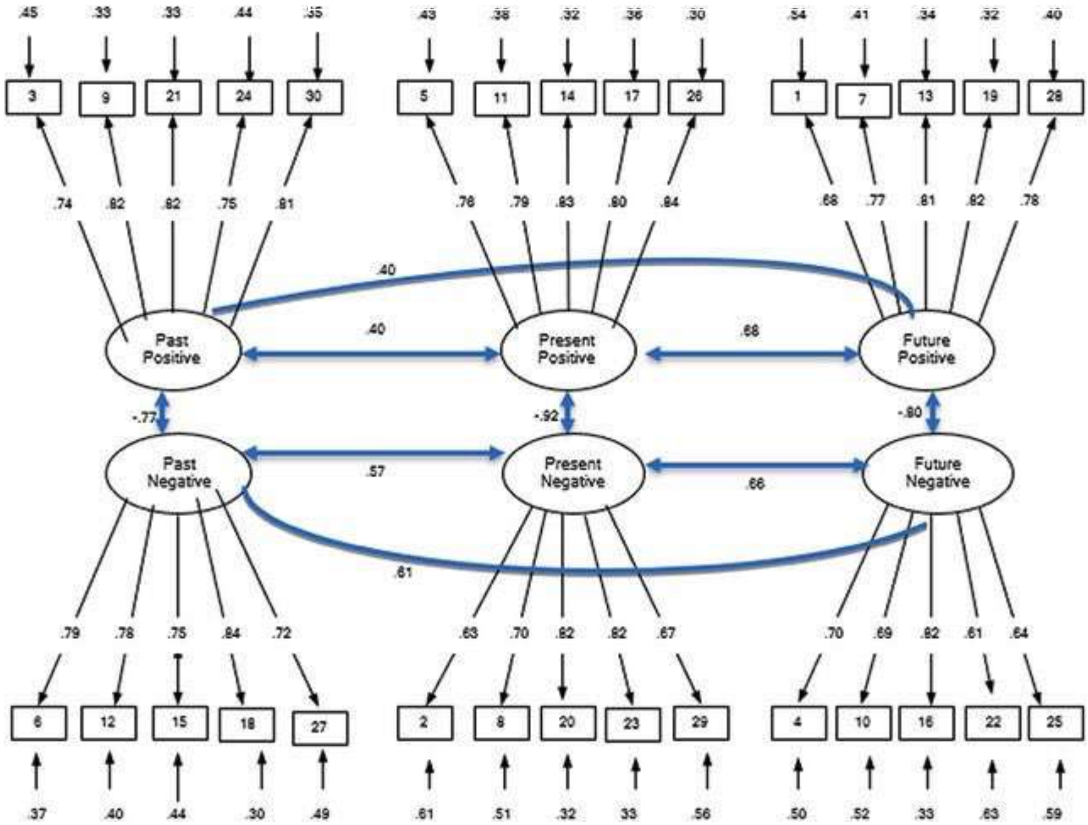


Figure 1 Factor Coefficients and Residual Variances for the Six-Factor Solution

including Spanish, and the Spanish version has been used in other studies (Konowalczyk et al., 2018). After looking the Spanish version of the AATI-TA (see Mello et al., 2010), the authors decides to start with the English version of the scale (Mello & Worrell, 2007) and do a translation to the Spanish dialect of the Rio de la Plata, that is used mostly in Uruguay and the Buenos Aires region of Argentina. A forward-translation/back-translation procedure was done beginning with the original English scale. Four English-language professionals contributed to these process (two for translations from English to the Spanish dialect of the Rio de La Plata and two for back translation). Discrepancies were resolved by discussion. Two cognitive interviews were also carried out to ensure comprehensibility by the intended sample, yielding excellent results for the version approved by the translators. The items for the final version used in this study are presented in the Supporting Information section. AATI-TA items are scored on a 5-point Likert scale (1 = *totally disagree*; 5 = *totally agree*).

Also as noted in the Introduction, there is considerable, supportive psychometric evidence for AATI-TA scores, suggesting that the scale could be useful in Uruguay. There is strong support for the six-factor structure (Worrell et al., 2013; Worrell, McKay, et al., 2018), and evidence of convergent validity (Alansari et al., 2013; Andretta et al., 2013; Buhl & Lindner, 2009; Chishima et al., 2019; Worrell & Mello, 2009). The subscale scores have generally been found to be internally consistent (Worrell, McKay et al., 2018), with some exceptions for Future Negative scores (e.g., Şahin-Baltaci et al., 2017). AATI-TA scores have demonstrated scalar invariance across males and females, younger and older adolescents, and adolescents and young adults (Donati et al., 2019; Konowalczyk et al., 2018).

Health Behaviours We used questionnaires about physical activity with questions related to attitudes, intentions, and actions from Craig et al. (2003) and Joireman et al. (2012). Physical activity attitudes were assessed by averaging three questions scored on a Likert scale of five values (1 = *Strongly disagree*; 5 = *Strongly agree*): "Regular physical activity is essential for good health," "Regular physical activity makes me feel better," and "I enjoy physical activity." The mean intercorrelation among these items was $r = 0.61$. Physical activity intentions were assessed with one question: "During the next week, how many times do you plan on doing exercises?" The question had eight response options (0 = *no days*, 7 = *every day of the week*). Physical activity actions were also assessed with a single question and the same eight response options (0 = *no day*, 7 = *every day of the week*): "In the last 7 days, how many days have you engaged in physical activity for a total of 60 minutes per day?" As can be seen, all of the variables were coded such that higher values reflected more positive attitudes and healthier attitudes intentions, and behaviours.

Eating Behaviours We used survey questions related to attitudes, intentions, and actions related to healthy food consumption from Joireman et al. (2012) and van Beek et al. (2013). Attitudes were assessed with two questions using a 5-point Likert format (1 =*Totally disagree*; 5 = *Totally agree*): "Healthy eating is essential for my health and well-being" and "I enjoy meals." Behavioural intention was assessed with a single question with 10 response options ranging from 1 (*Very unhealthy) to* 10 (*Very healthy*): "Consider the foods you will eat in the next week and evaluate how healthy you think they will be." We assessed the actions related to healthy food consumption using three questions with five response options (0 = *Never*, 4 = *Once per day*): "How often do you eat fruits," "How often do you eat vegetables," and "How often do you drink sugary drinks (e.g., soda)?"

Data Analysis

In this study, SPSS was used to calculate descriptive statistics, and missing values were imputed through the expectation maximization procedure for these analyses. MPlus 8.1 was used to examine the factor structure and factorial invariance and to create latent profiles. For the analyses using Mplus, we used the original data file and missing values were estimated using the full information maximum likelihood procedure. One participant was excluded from analyses due to a repetitive pattern of answers (i.e., giving the same response 25 times on the AATI-TA). No multivariate outliers were identified (using Mahalanobis distance procedure excluding participants with $p < .001$), and Item 9 on the AATI-TA (see Appendix) had the highest frequency of missing values (2.7%). Although the amount of missing data was low, using robust imputation techniques does not bias scores and is considered better practice than listwise elimination of participants with missing data, which can sometimes yield biased estimates (Graham, 2009).

We used confirmatory factor analysis (CFA) to examine model fit of the factor solution of the AATI-TA. All CFA models were analyzed using Mplus 8.1 (Muthén & Muthén, 2017) and fitted using the weighted least squares robust (WLMSV) estimator recommended for use with ordinal data (Svetina et al., 2020). Goodness of fit was evaluated with the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). As has been done with previous studies of the AATI-TA, we considered values ≥ .90 for the CFI and TLI as acceptable and ≥ .95 as excellent. For the RMSEA, values ≤ .08 were considered acceptable and values ≤ .05 were considered excellent (Marsh et al., 2004). Measurement invariance was evaluated by age and sex; in these analyses, we used the same criteria to evaluate fit. To assess if constraints worsened model fit, we used ΔCFI, where the most restrictive model should not reduce the CFI value by more than .002. (Meade et al., 2008) or as a minimum, .01 (Cheung & Rensvold, 2002).

Latent profile analysis (LPA) was conducted to explore the existence of time attitude profiles. LPA is a categorical latent variable approach that tries to identify hidden groups within a population based on a certain set of variables, in this case, ATTI-TA subscales scores. This analysis used Mplus (maximum-likelihood robust estimator). Models estimating two-profile to seven-profile solutions were compared. Better fit indices were determined by lower values in Akaike information criteria (AIC), Bayesian information criteria (BIC) and sample-size adjusted BIC (aBIC) but higher Entropy (< .80). Fit was also evaluated using the *p* value of Vuong–Lo–Mendell–Rubin likelihood ratio test (VLMR), the Lo-Mendell-Rubin test (LMR), and bootstrapped log-likelihood ratio tests (BLRT) where non-significant results signal that k - 1 profile solution is more parsimonious. We also considered the theoretical relevance of each latent profile (Jung & Wickrama, 2008; Muthén & Muthén, 2017).

Interpretations of correlations and group differences were based on effect sizes rather than statistical significance, using criteria suggested by Ferguson (2009). According to Ferguson, correlations of .20 represent an effect size of minimal practical significance, .50 a medium effect size, and .80 a strong effect size; equivalent values for r^2 are .04 .25, and .64, respectively. With regard to group differences, the comparable values for Hedges *g* are .41, 1.15, and 2.70, respectively.

Results

Preliminary Analyses Table 1 contains descriptive statistics for the six AATI-TA subscale scores and intercorrelations among the scores. In keeping with the extant literature, means for the positive subscales were higher than means for the negative subscales. Also as expected, correlations between positive and negative subscales were negative and moderate to strong and between subscales of the same valence were positive and moderate. Internal consistency coefficients (alphas and omegas) are also presented in Table 1. Alpha coefficients based on the raw scores ranged from .70 to .84, and omega coefficients based on the factor coefficients from the six-factor model in Table 2 ranged from .71 to .85. The lowest estimates were for scores on the future negative subscale. All other internal consistency estimates were ≥ .79.

In keeping with prior examinations of the factor structure of AATI-TA scores, three models were examined. The first model consisted of two factors based on valence, with all of the positive items on one factor and all of the negative items on the other factor. Model 2 was based on time period, with the three factors consisting of past, present, and future items. The final model examined was the hypothesized six-factor model. The fit statistics for the three models are presented in Table 2. The two-factor solution yielded unacceptable fit. The three-factor model (grouping items by temporal period) showed improvements in model fit, with two fit indices in the acceptable range and one in the excellent range. The three fit indices for the theorized six-factor model were all in the excellent range. Given these findings, which matched the theory and results of previous studies, this model was chosen for further analysis. The model with factor coefficients and residual variances is presented in Figure 1.

We then conducted invariance analyses on the six-factor model In Table 3, we present the results of measurement invariance by sex and age, although these groups were smaller than ideal for this type of analysis. Following the recommendations for measurement invariance analyses, with ordered categorical data (Svetina et al.,

Table 3. Invariance Analyses (WLMSV) for Time Attitude Scores

Model	χ^2 s-b	df	CFI	TLI	RMSEA (90 % CI)	MC	ΔCFI
Six factor- age							
1. Configural	1210.99*	780	0.966	0.962	0.050 (0.044 - 0.055)		
2. T & L	1289.65*	864	0.966	0.966	0.047 (0.042 - 0.053)	2 - 1	.000
3. T, L & I.	1300.99*	888	0.967	0.968	0.046 (0.040 - 0.051)	3 - 2	.001
Six factor - sex							
1. Configural	1182.51*	780	0.968	0.964	0.049 (0.043 - 0.054)		
2. T & L	1256.97*	864	0.969	0.969	0.046 (0.040 - 0.051)	2 - 1	.001
3. T, L & I.	1254.53*	888	0.971	0.972	0.043 (0.038 - 0.049)	3 - 2	.002

Note. MC = Models compared; T = Thresholds; L = Loading; I = Intercept.
*$p < .01$.

2020; Wu & Estabrook, 2016), we tested for configural/threshold invariance, threshold and loading invariance, and threshold, loading, and intercept invariance, in that order. As can be seen in the table, model fit did not decrease substantially with increasing constraints, supporting intercept invariance.

Latent Profile Analyses Latent profile analyses of AATI-TA scores were done with factor scores saved from the six-factor model. Model fit indices from the three to seven latent profile solutions are presented in Table 4. As can be seen, there was not a clearly acceptable solution based solely on fit indices, as the 3-, 4-, and 5-profile solutions could be considered good. We selected the 5-profile solution for interpretation based on the fit indices, parsimony, and theoretical considerations, such as content, size, and plausibility of the profiles generated (Jung & Wickrama, 2008). For example, the 3-profile solution was very broad, grouping just positives, negatives and intermediates, and the 4-profile solution was very similar to the 5-profile solution, but with no distinction between the High Positives and Moderate Positives. The five-profile model has, comparatively, good fit indices and parallels previous findings in the literature using time attitudes scores (Morgan et al., 2017). The five profiles, presented in Figure 2, are described below.

Individuals in Profile 1 reported a negative evaluation of the past and positive evaluations of the present and future. Both their present positive and future positive attitudes were high, whereas future negative attitudes were substantially lower than past and present negative attitudes. This group, which made up 10.4% of the sample, was labelled Resilients. Profile 2, called *High Positives*, had scores on the three time positive attitudes subscales substantially above the mean and scores on the three negative time attitudes substantially below the mean. This group was the smallest with 8.8% of the sample. The third profile constituted 20.8% of the sample. Labelled *Negatives*, this profile was the inverse to the Positives profile, and characterized by substantially above average scores on the negative time attitudes and substantially below average scores on the positive attitudes. Participants in the fourth profile, labelled *Present Negatives*, represents people with an ambivalent evaluation of the past and future and a negative evaluation of the present; at 34.5% of the sample, this group was the largest. The fifth group was called *Moderate Positives* (25.5%); their scores are conceptually Figure 2. *Latent Profiles Found in Current Study* analogous to the *High Positives*, but with scores closer to the mean. There were no significant profile differences in terms of education, sex, or age. Descriptive statistics for the five time attitude latent profiles are presented in Table 5.

Correlations with Outcome Measures We examined the associations between AATI-TA subscale factor scores and several physical and healthy food consumption variables. These results are presented in Table 6, and correlations ≥ .20 are interpreted (Ferguson, 2009). Healthy attitudes toward physical activity correlated positively with Present Positive and Future Positive scores, with the correlations in keeping with theory; that is, future positive scores were positively correlated with these healthy attitudes toward physical activity and future negative scores were inversely correlated with these attitudes. The pattern of correlations between healthy attitudes toward physical activity and past and present time attitudes was similar, but these correlations were neither statistically nor practically significant. No statistically significant correlations were observed between the six AATI-TA subscale scores and physical activity intentions or actions; indeed, these correlations were close to zero.

There were no statistically significant or meaningful correlations between time attitudes and healthy food consumption. However, positive dimensions of present and future of AATI-TA correlated positively with enjoying meals, and the negative dimensions were negatively associated with this variable; although

Table 4. Model Summaries for Three to Seven Latent Profiles Solutions

Model	AIC	BIC	aBIC	VLMR *p*	aLMR *p*	BLRT *p*	Entropy	No. of free parameters
3- Profile	5615,43	5722,04	5639,53	0,000	0,000	0,000	0,896	26
4- Profile	5384,28	5519,59	5414,87	0,015	0,016	0,000	0,879	33
5- Profile	**5266,48**	**5430,49**	**5303,55**	**0,027**	**0,029**	**0,000**	**0,884**	40
6- Profile	5154,94	5347,65	5198,49	0,408	0,416	0,000	0,910	47
7- Profile	5055,01	5276,43	5105,06	0,471	0,475	0,000	0,891	54

Notes. AIC = Akaike information criteria; BIC= Bayesian information criteria; a = adjusted; VLMR= Vuong–Lo–Mendell–Rubin likelihood ratio test; LMR= Lo, Mendell, and Rubin test; BLRT = bootstrapped log-likelihood ratio tests.

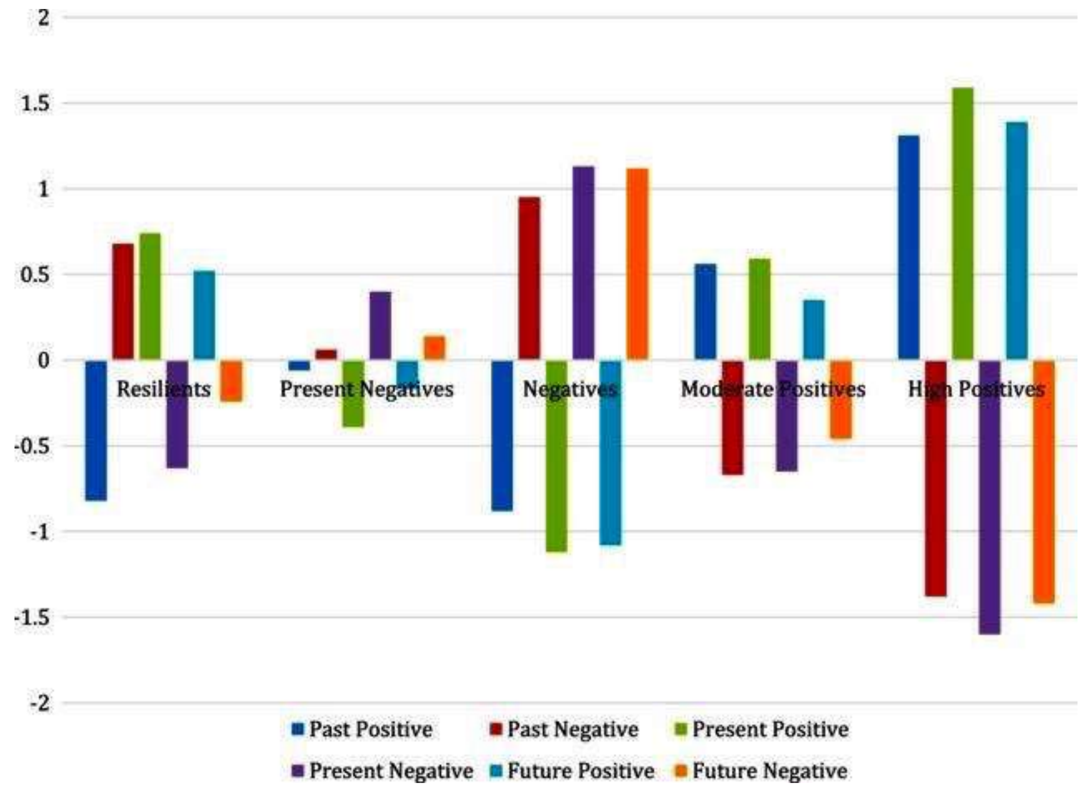


Figure 2. Latent Profiles found in current study

Table 5. Means for Time Attitude Factor Scores and Outcome Variables by Latent Profile

	1 Resilients, n = 47 (10,4%)	2 High Positives, n = 39 (8,8%)	3 Negatives n = 93 (20,8%)	4 Present Negatives n = 153, (34,5%)	5 Moderate Positives, n= 114 (25,5%)
Age in years	35.34	37.18	36.86	32.64	33.91
% of men	26.1	35.9	38.0	40.5	35.4
Factor scores					
Past P.	-.82	1.31	-.88	-.04	.56
Past N.	.68	-1.38	.95	.06	-.67
Present P.	.74	1.59	-1.12	-.39	.59
Present N.	-.63	-1.60	1.13	.40	-.65
Future P.	.52	1.39	-1.08	-.17	.35
Future N.	-.24	-1.42	1.12	.14	-.46
Physical activity					
Attitudes	4.49 (.48)	4.67 (.48)	4.14 (.90)	4.42 (.68)	4.49 (.64)
Intentions	2.66 (1.81)	2.83 (2.36)	2.64 (2.08)	2.64 (2.06)	2.53 (1.83)
Action	1.88 (2.01)	2.43 (2.30)	2.02 (1.84)	2.19 (2.00)	2.20 (1.85)

Note. Standard deviations are offered between parenthesis. P = Positive; N = Negative; HFC = Healthy food consumption.

Table 5. (vCont.) Means for Time Attitude Factor Scores and Outcome Variables by Latent Profile

	1 Resilients, n = 47 (10,4%)	2 High Positives, n = 39 8,8%)	3 Negatives n = 93 (20,8%)	4 Present Negatives n = 153, (34,5%)	5 Moderate Positives, n= 114 (25,5%)
Eating behavior					
Attitude – HFC	4.45 (.71)	4.29 (1.04)	4.27 (.85)	4.48 (.73)	4.36 (.78)
Attitude – Enjoy meals	4.73 (.52)	4.61 (.84)	4.22 (.89)	4.51 (.71)	4.72 (.51)
Intentions	6,85 (1.66)	6.94 (2.02)	6.29 (1.80)	6.78 (1.72)	6.86 (1.62)
Action – Fruits	3.33 (.96)	3.30 (1.06)	3.11 (1.03)	3.16 (1.07)	3.24 (.95)
Action – Vegetables	3.64 (.60)	3.32 (.83)	2.83 (1.02)	3.31 (.83)	3.51 (.78)
Action - Sugary drinks	1.85 (1.23)	2.00 (1.17)	1.86 (1.13)	2.04 (1.15)	1.85 (1.13)

Note. Standard deviations are offered between parenthesis. P = Positive; N = Negative; HFC = Healthy food consumption.

the correlation with future negative was below the threshold for interpretation.

Correlations between time attitudes and fruit consumption were generally close to zero (rs ≤ |.14|) as were correlations between time attitudes and drinking sugary drinks (rs ≤ .03). However, associations of time attitudes to vegetable consumption were statistically significant, although still relatively modest (|.15| < r < |.20|), and only the correlations for the present attitudes met Ferguson's (2009) criterion for minimum practical significance (i.e., |.20|). All of these correlations were in the theoretically congruent directions.

Differences among Profiles In the final set of analyses, we ran ANOVAs to examine if individuals with different temporal profiles differed in their intentions, attitudes, and behaviours towards physical exercise and healthy eating. Based on the previous literature on profiles, we hypothesized that individuals with the Positive profiles (High Positives, Moderate Positives, Resilients) would report more adaptive outcomes (i.e., endorsing healthy attitudes, intentions, and behaviours) and those with negative profiles (Negatives, Present Negatives) would report the least adaptive outcomes. Moreover, we hypothesized that the High Positives would report the most adaptive outcomes and the Negatives would report the least adaptive outcomes.

Physical Activity Profile means and standard deviations on the outcome variables are presented in Table 5. The hypotheses with regard to physical activity were only partially supported. Positives reported the highest scores with regard to physical activity attitudes, intentions, and actions, but Negatives only had the lowest scores with regard to attitudes. Indeed, attitudes toward physical activity showed a main effect, F (4, 318) = 3.77, p < .01. In keeping with best practice, post hoc comparisons were conducted using Hedges g, an effect size measure (Ferguson, 2009). High Positives (g =

Table 6. *Correlations between TAS Subscale Factor Scores and Outcome Measures*

	Physical activity			Eating behaviour					
				Attitudes			Action		
AATI-TA	Attitudes	Intentions	Action	HFC	EM	Intentions	F	V	SD
Past P.	.11	.01	.02	.04	.16	.12	.14	.15*	.01
Past N.	-.12	.00	-.04	-.01	-.12	-.11	-.11	-.17*	.03
Present P.	.17*	-.02	.05	.04	**.23***	.15	.09	**.20***	.00
Present N.	-.16*	.01	-.07	-.02	**-.20***	-.14	-.08	**-.20***	.03
Future P.	**.21***	-.02	-.01	.05	**.26***	.14	-.01	.19*	.00
Future N.	**-.20***	.02	.01	-.02	-.19*	-.14	.00	-.19*	.03

Note. AATI-TA = Adolescent and Adult Time Inventory–Time Attitudes; P = Positive; N = Negative; HFC = Healthy food consumption; EM = Enjoy meals; F = Fruits; V= Vegetables; SD = Sugary Drinks. Bolded coefficients meet Ferguson's (2009) minimal effect size for interpretation.
*p < .01.

0.66), Moderate Positives, (g = 0.45), and Resilients (g = 0.44) reported meaningfully (i.e., $g \geq .41$) healthier attitudes toward physical exercise than Negatives. No other comparisons on attitudes toward physical activity met the 0.41 threshold. High Positives reported healthier attitudes toward physical activity than Present Negatives (g = 0.39), Resilients (g = 0.37) and Moderate Positives (g = 0.30), and Present Negatives (g = 0.36) also reported healthier attitudes toward physical activity than Negatives. Profile differences in intention to do physical exercise, F (4, 319) = 0.13*,* p > .05*,* as well as for action or engaging in physical exercise, F (4, 321) 0.42, p > .05, were both non-significant, with differences between the groups also having small effect sizes (gs < 0.25).

Healthy Food Consumption The hypotheses with regard to eating behaviors were also only partially supported. Negatives had the lowest scores on five of the six eating behaviors variables, and the three positive profiles had higher means than the two negative profiles on four variables; attitudes toward healthy food consumption and consuming sugary drinks were the exceptions. However, for several of these variables, there were no statistically or practically significant differences among profiles. Attitudes toward healthy food consumption did not differ significantly among profiles (p > .05, g < 0.18), but enjoyment of meals differed significantly among profiles, F (4, 319) = 5.45, p < .001. High Positives (g = 0.44), Moderate Positives, (g = 0.71), and Resilients (g = 0.65) reported meaningfully higher scores on enjoying meals than did Negatives; Present Negatives (g = 0.37) also reported enjoying meals more than Negatives with a lower effect size. All other comparisons on this variable yielded effect sizes less than 0.34.

Intentions showed non-significant effects, F (4, 319) = 1.36, p > .05, although Positives and Negatives were the farthest apart, g = 0.35); all other differences on intention were low, gs < 0.34. The ANOVAs for differences in fruit consumption, F (4, 318) = 0.41, p > .05, gs < 0.22, and sugary drinks, F (4, 317) = 0.41, p > .05, gs < 0.17, were also neither statistically nor practically significant. However, we found a main effect of temporal profile for vegetable consumption, F (4, 322) = 7.57, p < .001, with Resilients and Moderate Positives, who did not differ meaningfully (g = 0.18), reporting the highest scores in this category; High Positives had the third highest score. Resilients reported eating meaningfully more vegetables than Negatives (g = 0.89), High Positives (g = 0.45), and Present Negatives (g = 0.42). Moderate Positives (g = 0.76), High Positives, (g = 0.50), and Present Negatives (g = 0.53) also reported eating meaningfully more vegetables than Negatives.

Discussion

The objectives of this article were (a) to determine the initial psychometric properties of AATI-TA scores translated into Rio de la Plata Spanish in a sample of Uruguayans, (b) to examine the correlations of time attitudes to physical activity and healthy food consumption, and (c) to identify temporal profiles and explore how these profiles relate to attitudes, intentions, and behaviours related to physical activity and healthy food consumption. The six AATI-TA scores had adequate to good internal consistency values and the scale's six-factor structure was supported in this sample. Evidence for strong invariance of the model between males and between young and middle-aged adults was also supportive. A few time attitude subscales had minimally interpretable correlations with attitudes toward physical activity, enjoyment of meals, and eating vegetables. Five interpretable time attitude profiles were found in this sample, and several meaningful differences among profiles were found on attitudes toward physical activity, enjoyment of meals and eating vegetables. Profiles did not differ meaningfully with regard to intentions to engage or actually engaging in physical activity, nor with regard to attitudes toward healthy food consumption, intentions to eat healthy foods and fruits, nor consuming sugary drinks. These findings are discussed below.

Psychometric Properties of AATI-TA Scores

The internal consistency estimates in the current study generally mirror results from other studies in the literature and all had acceptable-to-good reliability estimates, with future negative scores just above the acceptable threshold. Suboptimal reliabilities for future negative scores have been reported in a few studies (e.g., McKay, Cole et al., 2015). However, in a study using Peruvian Spanish, future negative scores had an alpha estimate of .74 (Worrell, Merino Soto, et al., 2018), and future negative scores obtained an alpha estimate of .85 in a sample of adolescents in Spain (Konowalczyk et al., 2018). Thus, although it seems as if the future negative construct may be harder to measure, the extant data suggest that it can be measured reliably in many contexts and samples. Researchers using the AATI-TA should always examine the internal consistency of the subscale scores to ensure that they are working consistently.

The six-factor solution yielded better fit to the data than the competing models based on valence and time periods. This finding replicates results reported in other studies differing in language and culture (e.g., Alansari et al., 2013; Buhl & Lindner, 2009; Çelik et al., 2017; Chishima et al., 2019; Cole et la., 2017; Donati et al., 2019; Juriševič et al., 2017; Konowalczyk et al., 2018; McKay, Cole, et al., 2015; Mello et al., 2016, 2019; Şahin-Baltaci et al., 2017; Worrell et al., 2013, 2020). The strength of coefficients is another highlight of the measure, as all but three items had coefficients higher than .65 on their latent variable. Invariance by gender and age was supported as model fit was not reduced after models with more restrictions were tested. In sum, this version of the AATI-TA is comparable across sexes and younger and middle adults.

AATI-TA Score Associations and Profiles

The strength of correlations between the AATI-TA subscale scores and the outcome variables were generally small, and differed by the intentions, attitudes, and actions with regard to physical activity and eating behaviour. Statistically significant correlations were found between time attitudes one the one hand and attitudes toward physical activity and enjoyment of meals on the other, but these associations were generally not practically significant. Correlations between time attitudes and intentions and behaviour were even smaller. This pattern resembles that found by Andre et al. (2018) where meta-analytic effect sizes of temporal psychology constructs were higher for associations with health-related attitudes than with intentions or actions. The correlational findings suggest that time attitude scores are not meaningfully related to physical activity or healthy food consumption.

However, we also found latent temporal profiles in this study, an important recent line of research in temporal psychology, replicating some of the profiles found with AATI-TA scores in other languages. We found latent profiles representing positives and negatives as in the Morgan et al. (2017) and Cole et al. (2017) studies, but we also found a couple profiles not yet described in the literature. One new profile, which we labelled Resilients, has some similarities to the Ambivalent profile found previously (Morgan et al., 2017) and shows how individuals may overcome negatives attitude to the past and have positive attitudes to the present and future. Two versions of the Positives profile – High Positives and Moderate Positives – found in many previous studies emerged in this sample as well. The differences in profiles may be due to sample differences in age or cultural context, or may reflect profiles that have just not yet been found.

Furthermore, there were meaningful differences among a few of the time attitude latent profiles suggesting that this type of analysis *may* be a better way to analyse the relationship of time perspective to other variables (Zimbardo & Boyd, 1999). Persons with a negative temporal profile reported (a) more negative attitudes towards physical exercise, (b) less enjoyment of meals and (c) less vegetable consumption, with effect sizes reaching practical significance. These attitudes and actions are related to higher risks of developing chronic illnesses, so targeting campaigns to individuals with a Negative profile may be a useful public health strategy, assuming that these findings are replicated. One important question to explain is why temporal profiles are associated with vegetable consumption and not with consuming fruit and sugary drinks. One possibility is that vegetables are more difficult to cook, so the preparation of them requires better attitudes towards eating and positive attitudes towards present and future (common to the three positive profiles). An alternative explanation could be that there are different temporal profiles have different associations with adaptive behaviours like healthy food consumption than with less adaptive, and nowadays, less socially desirable behaviours (e.g., sugar consumption).

In addition to replicating profiles found in other studies and cultural contexts (e.g., Alansari et al., 2013; Andretta et al., 2013; Buhl & Lindner, 2009; Cole et al., 2017; Worrell & Andretta, 2019) and thus confirming their generalizability, the findings in the current study confirm that some profiles are more adaptive than others, with the Negative profile being the least adaptive profile of the group. Worrell and Andretta (2019) found that Negatives not only had lower scores on a range of adaptive attitudes such as hope, optimism, and perceived life chances, but Negatives also reported higher hopelessness, and this group perceived more discrimination gender-based, ethnicity/race-based, and income-based discrimination than the positive groups, as well as lower school belonging and more barriers to college. The research to date suggests that Negatives are at risk for a variety of negative outcomes and will benefit from a variety of behavioural and mental health interventions.

Limitations of the Present Study and Future Directions

As with most research endeavours, this one had several limitations. The main limitation of this study was the representativeness of this sample. Given that almost half of the participants were university students at the time of data collection, they are more educated than the average Uruguayan population, so it is not possible to generalize our results to the larger population. Moreover, as noted before, validity is an ongoing endeavor (AERA, APA, & NCME, 2014, 2018). The findings in this study, including effect sizes, need to be replicated, and these effects and other types of validity evidence for AATI-TA scores will need to be collected on other samples (e.g., adolescents, older adults, non-college educated individuals) in Uruguay in future research. Additionally, this study used self-report measures for convergent validity analyses, which may explain why the many of the convergent validity findings involved attitudes. Direct observations of eating behaviour and physical activity as well as more robust indicators of outcomes may yield different results.

However, this study highlighted several future lines of research. First, it could be of interest to test to what extent temporal profiles are culture- or language-dependent in other Spanish speaking countries. Second, it could be of interest to broaden the study of the relation of temporal profiles with health behaviours and measure the extent to which those profiles are associated with rates of chronic illness. In studies of associations between time attitudes and other constructs (e.g., psychological wellbeing, educational outcomes), time attitude profiles have been found to have stronger associations than time attitude scores (Worrell & Andretta, 2019). However, this claim cannot be made about health behaviors on the basis of this study; this issue needs to be examined in future research. For instance, further research could use other sources of information and work with clinical populations. Lastly, examining possible precursors of temporal profiles (like for instance, personality traits, negative/positive affect, etc.) may contribute to our understanding of how these profiles are related to other psychological phenomena. The results suggest that the AATI-TA will be a useful tool for future research.

Funding Statement

This research received no specific grant from anyfunding agency, commercial or not-for-profit sectors.

Supplementary Materials

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/SJP.2020.51.

References

Ajzen, L. (2005). *Attitudes, personality, and behaviour*. McGraw-Hill Education.

Alansari, M., Worrell, F. C., Rubie-Davies, C., & Webber, M. (2013). Adolescent Time Attitude Scale (ATAS) scores and academic outcomes in secondary school females in New Zealand. *International Journal of Quantitative Research in Education, 1*(3), 251–274. https://doi.org/10.1504/IJQRE.2013.057687

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.

Andre, L., Van Vianen, A. E. M., Peetsma, T. T. D., & Oort, F. J. (2018). *Motivational power of future time perspective: Meta-analyses in education, work, and health*. *PLoS ONE*, *13*. https://doi.org/10.1371/journal.pone.0190492

Buhl, M. & Lindner, D. (2009). Time perspectives in adolescence: Measurement, profiles and links with personality characteristics and scholastic experience. *Diskurs Kindheits und Jungendforschung*, *4*(2), 197–216.

Çelik, E., Sahranç, Ü., Kaya, M., & Turan, M. E. (2017). Adolescent time Attitude Scale: Adaptation into Turkish. *Universal Journal of Educational Research*, *5*(2), 252–257. https://doi.org/10.13189/ujer.2017.050210

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Chishima, Y., Murakami, T., Worrell, F. C., & Mello, Z. R. (2019). The Japanese version of the Adolescent Time Inventory-Time Attitudes (ATI-TA) scale: Internal consistency, structural validity, and convergent validity. *Assessment, 26*(2), 181–192. https://doi.org/10.1177/1073191116683800

Cole, J. C., Andretta, J. R., & McKay, M. T. (2017). An exploratory examination of the viability and meaningfulness of time attitudes profiles in adults. *Personality and Individual Differences*, *106*, 146–151. https://doi.org/10.1016/j.paid.2016.10.046

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., ... & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381-1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Donati, M. A., Boncompagni, J., Scabia, A., Morsanyi, K., & Primi, C. (2019). Using the Adolescent Time Inventory–Time Attitudes (ATI) to assess time attitudes in Italian adolescents and young adults: Psychometric properties and validity. *International Journal of Behavioral Development*, *43*(5), 424–435. https://doi.org/10.1177/0165025418797020

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Henson, J. M., Carey, M. P., Carey, K. B., & Maisto, S. A. (2006). Associations among health behaviours and time perspective in young adults: Model testing with boot-strapping replication. *Journal of Behavioural Medicine, 29*, 127–137. https://doi.org/10.1007/s10865-005-9027-2

Joireman, J., Shaffer, M. J., Balliet, D., & Strathman, A. (2012). Promotion orientation explains why future-oriented people exercise and eat healthy: Evidence from the two-factor Consideration of Future Consequences-14 Scale. *Personality and Social Psychology Bulletin*, *38*(10), 1272–1287. https://doi.org/10.1177/0146167212449362

Jung, T. A., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*, 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Juriševič, M., Worrell, F. C., & Mello, Z. R. (2017). Measuring time attitudes in Slovenia: Psychometric properties of the Adolescent and Adult Time Attitudes Scale (AATI-TA). *Horizons of Psychology*, *26*, 89–97. https://doi.org/10.20419/2017.26.472

Konowalczyk, S., Mello, Z. R., Röske, L. A., Buhl, M., Heim, R., & Worrell, F. C. (2018). Adolescent and Adult Time Inventory-Time Attitude Scales: Validity and contributions to physical activity and self-concept in Spanish adolescents. *International Perspectives in Psychology: Research, Practice, Consultation*, *7*(2), 76–90. https://doi.org/10.1037/ipp0000084

Lang, F. R., & Carstensen, L. L. (2002). Time counts: Future time perspective, goals, and social relationships. *Psychology and Aging*, *17*(1), 125–139. https://dx.doi.org/10.1037/0882-7974.17.1.125

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McKay, M. T., Andretta, J. R., Cole, J C., Konowalczyk, S., Wells, K. C., & Worrell, F. C. (2018). Time attitudes profile stability and transitions: An exploratory study on adolescent health behaviours among high school students. *Journal of Adolescence, 69,* 44–51. https://doi.org/10.1016/j.adolescence.2018.09.002

McKay, M. T., Andretta, J. R., Magee, J., & Worrell, F. C. (2014). What do temporal profiles tell us about adolescent alcohol use? Results from a large sample in the United Kingdom. *Journal of Adolescence, 37*(8), 1319–1328. https://doi.org/10.1016/j.adolescence.2014.09.008

McKay, M. T., Cole, J. C., Percy, A., Worrell, F. C., & Mello, Z. R. (2015). Reliability and factorial validity of Adolescent Time Inventory-Time Attitude (ATI-TA) scores in Scottish and Northern Irish adolescents. *Personality and Individual Differences, 86,* 412–416. https://doi.org/10.1016/j.paid.2015.06.040

McKay, M. T., Perry, J. L., Cole, J. C., & Worrell, F. C. (2018). What time is it? Temporal psychology measures relate differently to alcohol-related health outcomes. *Addiction Research and Theory, 26*(1), 20–27. https://doi.org/10.1080/16066359.2017.1309032

McKay, M. T., Worrell, F. C., Temple, E. C., Perry, J. L., Cole, J. C., & Mello, Z. R. (2015). Less is not always more: The case of the 36-item short form of the Zimbardo Time Perspective Inventory. *Personality and Individual Differences*, *72*, 68–71. https://doi.org/10.1016/j.paid.2014.08.018

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Mello, Z. R., Oladipo, S. E., Paoloni, V. C., & Worrell, F. C. (2019). Time perspective and risky behaviors among Nigerian young adults*. Journal of Adult Development*, *26*(3), 161–171. https://doi.org/10.1007/s10804-018-9304-2

Mello, Z. R., & Worrell, F. C. (2007). *The Adolescent Time Inventory-English.* Graduate School of Education, University of California. http://faculty.sfsu.edu/~zmello/content/adolescent-and-adult-time-inventory-aati#English

Mello, Z. R., Worrell, F. C., Anguiano, R. M., & Mendoza-Denton, R. (2010). *The Adolescent and Adult Time Inventory – Spanish.* Unpublished scale. University of Colorado, Colorado Springs & Graduate School of Education, The University of California. http://faculty.sfsu.edu/~zmello/content/adolescent-and-adult-time-inventory-aati#Spanish

Mello, Z. R., Zhang, J. W., Barber, S. J., Paoloni, V. C., Howell, R. T., & Worrell, F. C. (2016). Psychometric properties of time attitude scores in young, middle, and older adult samples. *Personality and Individual Differences*, *101*, 57–61. https://doi.org/10.1016/j.paid.2016.05.037

Milfont, T. K., Andrade, P. R., Belo, R. P., & Pessoa, V. S. (2008). Testing Zimbardo Time Perspective Inventory in a Brazilian sample. *Interamerican Journal of Psychology*, *42*(1), 49–58.

Morgan, G. B., Wells, K. E., Andretta, J. R., & McKay, M. T. (2017). Temporal attitudes profile transition among adolescents: A longitudinal examination using mover−stayer latent transition analysis. *Psychological Assessment, 29*(7), 890–901. https://doi.org/10.1037/pas0000383

Murphy, L., & Dockray, S. (2018). The consideration of future consequences and health behaviour: A meta-analysis. *Health Psychology Review*, 1–55. https://doi.org/10.1080/17437199.2018.1489298

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Author.

Şahin-Baltaci, H, Tagay, Ö, Worrell, F. C., & Mello, Z. R. (2017). Psychometric properties of Turkish Adolescent Time Inventory-Time Attitudes (ATI- TA) scores. *International Perspectives in Psychology: Research, Practice, Consultation, 6*(1)*,* 47–59. https://doi.org/10.1037/ipp0000066

Svetina D., Rutkowski L, & Rutkowski, D (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling, 27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

van Beek, J., Antonides, G., & Handgraaf, M. J. (2013). Eat now, exercise later: The relation between consideration of immediate and future consequences and healthy behavior. *Personality and Individual Differences, 54*, 785-791. https://doi.org/10.1016/j.paid.2012.12.015

World Health Organization. (2018). *World health statistics 2018: Monitoring health for the sustainable development goals*. https://apps.who.int/iris/bitstream/handle/10665/272596/9789241565585-eng.pdf?ua=1

Worrell, F. C., & Andretta, J. R. (2019). Time attitude profiles in American adolescents: Educational and psychological correlates. *Research in Human Development*, *16*(2), 102–118. https://doi.org/10.1080/15427609.2019.1635860

Worrell, F. C., McKay, M. T., & Andretta, J. R. (2018). Psychometric properties of Adolescent Time Inventory-Time Attitude (ATI-TA) scores in three waves of longitudinal data. *Psychological Assessment, 30*(1), 106–115. https://doi.org/10.1037/pas0000457

Worrell, F. C., & Mello, Z. R. (2009). Convergent and discriminant validity of time attitude scores on the Adolescent Time Perspective Inventory. *Diskurs Kindheits und Jugendforschung*, *4*(2)*,* 185–196.

Worrell, F. C., Mello, Z. R., & Buhl, M. (2013). Introducing English and German versions of the Adolescent Time Attitude Scale (ATAS). *Assessment, 4*(4)*,* 496–510. https://doi.org/10.1177/1073191110396202

Worrell, F. C., Mello, Z. R., Laghi, F., Baiocco, R., & Lonigro, A. (2020). Time perspective constructs in Albanian and Italian adolescents: Exploratory analyses. *Psychological Reports.* Advance online publication. https://doi.org/10.1177/0033294120913493

Worrell, F. C., Merino Soto, C., Mello, Z. R., & Mendoza-Denton, R. (2018, May). *Examining Adolescent and Adult Time Inventory (AATI) scores in Peru* [Poster Presenteation]. Association for Psychological Science Annual Meeting, San Francisco, CA, United States.

Worrell, F. C., Temple, E., McKay, M. T., Zivkovic, U., Perry, J. L., Mello, Z. R., Musil, B., & Cole, J. C. (2018). A theoretical approach to the Zimbardo Time Perspective Inventory: Results from America, Australia, Slovenia, and the United Kingdom. *European Journal of Psychological Assessment, 34*(1)*,* 41–51. https://doi.org/10.1027/1015-5759/a000313

Zimbardo, P. G., & Boyd, J. N. (1999). Putting tie in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, *77*(6), 1271–1288. https://doi.org/10.1037/0022-3514.77.6.1271.

Wu H., & Estabrook R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*, 1014–1045 https://doi.org/10.1007/s11336-016-9506-0

Taza decorativa del proyecto Lexiland
Lexiland's project decorative mug

Autoras: **Camila Zugarramurdi** y **Lucia Fernández.** Este estudio formó parte del doctorado de la Dra. Camila Zugarramurdi y contó con el apoyo de la Administración Nacional de Educación Pública-Consejo de Educación Inicial y Primaria (ANEP-CEIP), el Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea (CEIBAL), el Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA) y el **CIBPsi**. El proyecto desarrolló una batería de tareas digitalizadas que permiten mejorar la predicción del desempeño lector en niños en edad preescolar de forma sistemática y semi-automatizada, así como, avanzar en la comprensión de sus mecanismos subyacentes.

Authors: **Camila Zugarramurdi** and **Lucia Fernández**. This study was part of Dr. Camila Zugarramurdi's PhD and was supported by the Administración Nacional de Educación Pública-Consejo de Educación Inicial y Primaria (ANEP-CEIP), the Plan de Conectividad Educativa de Informática Básica para el Aprendizaje en Línea (CEIBAL), the Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA), and the **CIBPsi**. The project developed a battery of digitized tasks that allow improving the prediction of reading performance in preschool children in a systematic and semi-automated way, as well as advancing the understanding of its underlying mechanisms.

Mind in the ortography: revisiting the contribution of pre-reading phonological awareness to reading acquisition

Camila Zugarramurdi[1,2], Lucía Fernández[1,3], Marie Lallier[2], Juan Carlos Valle-Lisboa[1,3], Manuel Carreiras[2,4,5]

1 Instituto de Fundamentos y Métodos en Psicología y **Centro de Investigación Básica en Psicología** (CIBPsi). Facultad de Psicología. Universidad de la República. Montevideo, Uruguay
2 Basque Center on Cognition Brain and Language (BCBL). Donostia, Spain
3 Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA). Universidad de la República. Montevideo, Uruguay
4 Ikerbasque. Basque Foundation for Science. Bilbao, Spain
5 Departamento de Lengua Vasca y Comunicación. University of the Basque Country. Bilbao, Spain

Camila Zugarramurdi https://orcid.org/0000-0003-1675-0722, Marie Lallier https://orcid.org/0000-0003-4340-1296, Juan Valle-Lisboa https://orcid.org/0000-0002-5662-5937, Manuel Carreiras https://orcid.org/0000-0001-6726-7613

Author note. Juan Valle Lisboa is now at Sección Biofísica y Biología de Sistemas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Correspondence concerning this manuscript should be addressed to Juan Valle-Lisboa, Facultad de Ciencias, Iguá 4225, Montevideo 11600, Uruguay. Email: **juancvl@fcien.edu.uy**

This is a free version -with the approval of the authors- of the pre-print of the manuscript Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the orthography: Revisiting the contribution of prereading phonological awareness to reading acquisition. *Developmental Psychology*, *58*(6), 1003-1016. Available at **https://doi.org/10.1037/dev0001341**

Abstract

Reading acquisition is based on a set of preliteracy skills that lay the foundation for future reading abilities. Phonological awareness —the ability to identify and manipulate the sound units of oral language— has been reported to play a central role in reading acquisition. However, current evidence is mixed with respect to its universal contribution to reading acquisition across orthographies. This longitudinal study examines the development and contribution of phonological awareness to early reading skills in Spanish, a transparent orthography. The results of a comprehensive battery of phonological awareness skills in a large sample of children (time 1 n = 616, 296 females, mean age 5.6, from middle to high socioeconomic backgrounds; time 2 n = 397) with no reading experience at study onset suggest that the development of phonological awareness is delayed in Spanish. Furthermore, our results show that phonological awareness does not contribute to the prediction of reading acquisition above and beyond other preliteracy skills. Letter knowledge indexes children's ability to identify phonemes and thus takes a more central role in the prediction of early reading skills. Therefore, we underscore the need to thoughtfully address the distinctive features of the reading acquisition process across orthographies, which should be taken into account in models of reading and learning to read.

Keywords: Phonological awareness, decoding, reading, Spanish, transparent orthography, longitudinal

Introduction

Reading is a fundamental ability in modern societies, yet many children and adults struggle with reading. This has far-reaching negative consequences for personal development and professional achievement (Arnold et al., 2005). Thanks to decades of research, we now have a fairly comprehensive picture of the preliteracy skills required for successful reading. There is broad consensus regarding three critical skills: knowledge of letter sounds (letter knowledge; LK); rapid and efficient access to lexical representations (rapid automatized naming; RAN); and an ability to consciously manipulate the constituent units of oral language, generally referred to as phonological awareness (Boets, Wouters, van Wieringen, & Ghesquière, 2007; Lyytinen et al., 2006; Muter, Hulme, Snowling, & Stevenson, 2004; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Phonological awareness (PA) is a central construct in most, if not all, models of reading acquisition (for a meta-analysis see Melby-Lervåg, Lyster, & Hulme, 2012). It is generally defined as a metalinguistic or metacognitive ability to identify and manipulate the sounds of a language. The term *sounds* in this context may refer to individual phonemes, syllables, or words. However, most research suggests phonological awareness at the phonemic level (*phonemic* awareness) is the most important component in reading acquisition (Castles & Coltheart, 2004). Typical PA tasks involve segmenting a word into its constituting syllables or phonemes, blending syllables or phonemes into a word, or replacing a syllable or phoneme within a given word or pseudoword. The important role of PA in reading is further confirmed by studies reporting that dyslexic children show a PA deficit (Melby-Lervåg et al., 2012; Vellutino, Fletcher, Snowling, & Scanlon, 2004), and that training in PA can improve reading skills (Bowyer-Crane et al., 2008). A fourth skill that taps into phonological processes—and also predicts early reading acquisition—but has received far less attention than PA, is verbal short-term memory. Verbal short-term memory (vSTM) has received a mixed treatment in the literature, sometimes treated as a foundational skill of interest (Moll et al., 2014; Ramus & Szenkovits, 2008; Torgesen, Wagner, & Rashotte, 1994) and sometimes as a covariate (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Puolakanaho et al., 2007; Vaessen et al., 2010). Interestingly though, the core phonological deficits described in children with dyslexia often involve vSTM, alongside PA and RAN (Torgesen et al., 1994). While all of the preliteracy skills mentioned above have been shown to play a role in predicting reading acquisition, PA is the most studied, both because of its central role, and because of its potential as a target for intervention. However, at least three aspects of the PA-reading relation remain unclear.

Current issues on the PA-reading relation

Universality

Most evidence comes from studies on English, whose orthography is atypical, in comparison with the orthographies of most other languages (Share, 2008). Orthographies can be characterized by the consistency of the mapping between graphemes and phonemes. In highly transparent orthographies, such as Spanish, Finnish, or Italian, the mapping between graphemes and phonemes is almost univocal, while in less transparent orthographies, such as English, this mapping depends heavily on the orthographic context in which the grapheme is embedded (Schmalz, Marinus, Coltheart, & Castles, 2015). Orthographic consistency, in turn, modulates the developmental trajectories of reading acquisition: high decoding levels are achieved faster in more transparent orthographies (Seymour, Aro, & Erskine, 2003). Moreover, several studies have shown that PA together with RAN and LK skills account for larger amounts of variance in English than they do in other languages with more transparent orthographies (Caravolas et al., 2012; Moll et al., 2014). Therefore, it is still debated whether the central role attributed to PA in reading acquisition in English can be generalized to more transparent orthographies (Castles & Coltheart, 2004; Share, 2008; Verhoeven & Keuning, 2017). Questioning the central role of PA does not necessarily mean that PA has no role to play in predicting reading acquisition in languages with more transparent orthographies, but rather that its role might be less central with respect to other preliteracy skills (Duncan et al., 2013; Landerl et al., 2019; Verhoeven & Keuning, 2017). While each preliteracy skill adds unique variance that helps explain early reading skills, these skills are also correlated. For example, a child with strong PA skills may use this knowledge as a scaffold to learn letter names and vice versa: learning letter names can aid in the development of PA (Kim, Petscher, Foorman, & Zhou, 2010; Piasta & Wagner, 2010; Treiman & Kessler, 2004).

Causality

Reciprocal influences between PA and reading acquisition throughout development makes it harder to address the issue of the PA-reading relation (Castles & Coltheart, 2004; Hulme, Snowling, Caravolas, & Carroll, 2005). While children with reading difficulties often show accompanying poor PA skills, it is possible that these observed deficiencies in PA result from reduced or suboptimal reading experience. For example, Huettig and colleagues (2018) note the importance of distinguishing cause and effect when establishing the main factors that contribute to reading difficulties. They argue that in order to determine whether a given skill plays a causal role in reading development, it needs to be assessed in prereaders before any reading skills have developed. This rules out the possibility that the observed effects are a consequence of suboptimal reading experience rather than their primary cause. Longitudinal studies that initially test prereaders before reading instruction, although not conclusive, are thus a primary source of evidence to assess the causal role of PA skills in future reading performance (see also Goswami, 2015).

Operationalization

All the above aspects of the PA-reading relation have been additionally obscured by a third factor: the operationalization of PA (McBride-Chang, 1995; Runge & Watkins, 2006; Vloedgraven & Verhoeven, 2009; Yopp, 1988). Tasks used to test PA vary in difficulty on many dimensions. First, they may differ in terms of the linguistic unit of analysis, which could be whole-words, syllables, intra-syllabic units (onset and rimes), or phonemes. While it has been shown across orthographies that children's sensitivity develops along a trajectory from larger to smaller units (Anthony et al., 2011; Duncan et al., 2013; Papadopoulos, Spanoudis, & Kendeou, 2009; Ziegler & Goswami, 2005), it is still debated whether sensitivity to phonemes can be attained prior to any literacy exposure (Castles & Coltheart, 2004; Landerl et al., 2019). Longitudinal studies have used either measures of *phonological* awareness (including both syllabic and phonemic items) or *phonemic* awareness (only phonemic items) measures, further complicating the matter. Second, they may differ in terms of the kind of cognitive operation involved in the task. Both in English and Spanish, children are able to blend linguistic units before they can segment them and identify them before they can manipulate them (e.g., detect identical onsets in two words vs. remove the onset) (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003). Different studies use a wide variety of different tasks, which makes it to compare them. Third, PA tasks vary in terms of the memory-load they impose. This has been identified as a crucial modulating factor in performance across phonological awareness tasks (Martinez Perez, Majerus, & Poncelet, 2012; Ramus & Szenkovits, 2008). Finally, PA tasks vary in their response format. PA is usually measured in tasks that require verbal responses, such as removing a given phoneme from a word and producing the resulting word/ nonword. Naturally, producing a verbal response adds an additional cognitive process, which may or may not be tapping into the PA construct directly. For example, Cunningham and colleagues (2015) showed that grain size and response format in PA tasks constitute independent factors, and that the production of a verbal response contributed unique variance to decoding above and beyond the linguistic component involved. However, production of a verbal response is not usually defined as part of the PA construct, but as a means for measuring it. In sum, the wide range of task properties commonly used has presented a further obstacle in trying to reconcile divergences in the existing literature (Ramus & Szenkovits, 2008).

The above factors —universality, causality, and operationalization— may explain why evidence for the unique (i.e., above and beyond other variables) contribution of PA to reading has been inconsistent across orthographies. In the last two decades, several studies have attempted to address these issues by assessing preliteracy skills in prereaders in less transparent orthographies using longitudinal designs as well as cross-language approaches. While some of these studies have found evidence that supports a universal role for PA in reading acquisition (Caravolas et al., 2012; Furnes, Elwér, Samuelsson, Olson, & Byrne, 2019; Puolakanaho et al., 2007; Vaessen et al., 2010), others have not (De Jong & Van der Leij, 2003; Defior, Serrano, & Marín-Cano, 2008; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 2012; Landerl et al., 2019; Mann & Wimmer, 2002; Schmitterer & Schroeder, 2019; Van Bergen et al., 2011).

There is an additional challenge in trying to make sense of this divergent evidence that has not, to the best of our knowledge, yet been systematically addressed: studies which show no evidence for a universal contribution of PA to reading acquisition also frequently report floor effects on PA measures (De Jong & Van der Leij, 2003; Georgiou et al., 2012; Landerl et al., 2019; Van Bergen et al., 2011). Floor effects are a form of scale attenuation encountered when measures are close to zero for most participants, thus providing an inaccurate measure of individual participant's ability. They generally result from tasks that are too difficult for the target participants, either due to poor item design or lack of adjustment for developmental stage. Most often, this is regarded as a methodological limitation in such studies.

Additionally, evidence in favour of a universal account has its own challenges. Some studies include a sample which already has some reading experience at study onset. As discussed above, this introduces a confound due to the reported reciprocal influences between phonological awareness and reading (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Vaessen et al., 2010). In other cases, not all relevant covariates are included (crucially, verbal short-term memory and letter knowledge may be left out), making it difficult to compare the unique contribution of each predictor hard to perform (Furnes & Samuelsson, 2010; Puolakanaho et al., 2007). Finally, sample composition in these studies is often enriched with children at risk of reading failure, usually due to family history of dyslexia (Puolakanaho et al., 2007). Naturally, when trying to achieve a final sample that contains at least some children with reading difficulties, this is a sensible approach. However, it limits the generalizability of results to a broader, unselected population.

Understanding the unique contribution of PA to early reading skills across orthographies is relevant for both practical and theoretical reasons. An important practical implication is the design and use of appropriate screening tools for children at risk of reading difficulties. Early screening is vital, since it has been shown that remediation programmes are more effective the sooner they begin (Ozernov-Palchik & Gaab, 2016). If the unique contribution of PA is orthography dependant, then screening should be orthographic-specific (see for example, Solheim, Torppa, & Uppstad, 2020). Adaptation of tools developed for English-speaking children would not be appropriate for Spanish-speaking ones. From a theoretical standpoint, it raises new questions concerning the universality of the current prevailing model of reading acquisition. If the contribution of PA is not unique, does this mean PA has no role to play in reading acquisition? Can this explain the floor effects often reported in more transparent orthographies? If so, why are floor effects in PA often observed in more transparent orthographies but not in less transparent ones? We claim here that floor effects could be explained by a delayed development of PA skills in more transparent orthographies, rather than by measurement error. If, in more transparent orthographies, PA skills during the kindergarten years are only primitively or not at all developed, then it stands to reason that PA will show no unique contribution to later reading acquisition. Further, it is possible that other preliteracy skills will take its place. We believe LK is a strong candidate. Since, in more transparent orthographies, letter sounds are virtually equivalent to the phonemes they represent, in such orthographies LK might index children's ability to identify phonemes, thus replacing PA as a main contributor to later reading acquisition.

The Present Study

In the present study we examined the unique contribution of pre-reading phonological awareness to early reading skills in a transparent orthography, Spanish. Our hypothesis was that, in more transparent orthographies: i) delayed development of PA skills explain the previously observed floor effects of PA, ii) LK indexes children's ability to identify phonemes and thus takes a more central role.

To test this hypothesis, it was critical to design tasks that were sensitive to the general PA abilities of children at the time of testing. In order to tackle this issue, we employed a comprehensive assessment of phonological awareness, involving the manipulation of syllables and phonemes, which included four different tasks consisting of 163 items. We longitudinally assessed an unselected sample of children at two time points: in kindergarten, before any reading instruction has taken place, and at the end of Grade 1. Crucially, we computed latent ability scores through an item-response theory (IRT) approach, which allowed us to control for measurement error and compare tasks scores across different scales (Cole & Preacher, 2014; Hjetland et al., 2019). Moreover, in order to examine the unique contribution of PA relative to other preliteracy skills and general cognitive factors, we also assessed LK, RAN, and vSTM, as well as several other relevant control variables. At the end of grade 1, we repeated K5 measures and additionally measured reading skills. In order to account for the fact in more transparent orthographies children achieve high accuracy levels at the end of first grade (Seymour et al., 2003), we assessed decoding accuracy in words and pseudowords, as well as fluency, and modelled these factors independently.

Methods

Participants and procedure Data was collected longitudinally in two consecutive school years starting with the last year of Kindergarten (K5), followed by 1st grade (G1). All children were native Spanish speakers. Children were recruited from 26 public schools in Montevideo, Uruguay, with middle to high socio-economic status, according to the National School Administration classification. Schools were either part-time or full-time. All children attending K5 level at time 1 were invited to take part in the study. Only those whose parents signed a consent form, in accordance with the Declaration of Helsinki, took part (N = 616, 296 females, M age = 5.6 years).

Sample size was estimated based on expected dropout, prevalence of reading difficulties, and power calculations. Power calculations using GPower3.1 (Faul, Erdfelder, Buchner & Lang, 2009) show that, for a multiple regression with 11 factors (see below), and an expected effect size $f^2 = 0.1$, a sample of at least 262 children is needed to obtain a power of .95 at .05 level. Dropout was estimated at 30%, given the study was conducted in the school setting, longitudinally; and prevalence of reading difficulties was estimated at 10%. Thus, an initial sample of approximately 600 children was targeted for.

The final sample, including children that completed both stages of data collection, consisted of 397 children (181 females, M age = 5.6). Dropout was mainly due to children that switched schools between time points. Nine children were additionally excluded because they did not complete at least half of the tasks. Dropout analysis showed significant differences in SES ($X^2(2, N = 616) = 6.82, p = .033$), with those dropping out showing a larger proportion of children from low SES. No other variable of interest showed any significant differences between groups (Age: $t(614)= 0.55, p = .58$; Gender: $X^2(1, N = 616) = .70, p = .40$, IQ: $t(614)= 1.29, p = .20$, letter knowledge: $t(614)= 1.80, p = .07$, phonological awareness: $t(614)= 0.32, p = .75$, and rapid automatized naming: $t(614)= -1.15, p = .25$).

Whenever justified, missing data was imputed among independent variables (i.e., excluding reading). The only exception being SES, since there are not enough social variables to make a valid imputation (see below). Demographic information (age, gender, and maternal education as a proxy for socioeconomic status) was obtained from a national database from the educational system (ANEP). Seventy-five children were excluded due to missing SES data. Missing data was present across measured variables (M = 14%, min = 10%, max = 18%) and was handled through random forest imputation via the *missforest* package in R (Eckert, Vaden, & Gebregziabher, 2018; Stekhoven & Buehlmann, 2012). This method predicts the missing data taking into account all the other data and t has been shown to be highly effective even when there is low or moderate correlation between variables (Tang & Ishwaran, 2017). No dependent variables (i.e., reading) were included in the imputed procedure. Normalized root mean square error of approximation for the imputation of continuous variables was 0.30 (Stekhoven & Buehlmann, 2012).

Assessment was carried out at the schools, in groups of four to five children, except for the reading and RAN tasks which were assessed individually. Sessions lasted between 10 and 15 minutes. All tasks were presented digitally through individual tablets, except for reading and RAN. In order to do so, we created an Android-based application in videogame format, called *Lexiland* (Zugarramurdi, Fernández, Lallier, Carreiras, Valle-Lisboa, in press). Instructions and stimuli were pre-recorded and delivered though headphones. All tasks began with two to three examples, followed by two to three practice trials with feedback. Children were allowed to repeat practice trials freely. Two research assistants monitored task performance and were available to clarify instructions on demand.

All the procedures were approved by the Ethics Committee of the School of Psychology at Universidad de la República; date of approval February 17th, 2016, under the study title: "Design of a digital assessment battery to detect reading difficulties".

Measures

Phonological awareness Phonological awareness was assessed by 4 different tasks: segmentation, blending, onset matching, and rhyming. In *Segmentation* (22 syllabic items, 28 phonemic items) children heard a word —with their corresponding image to reduce memory load— and had to segment it. In order to avoid verbal responses, illustrations of dices corresponding to number two to four for syllables, and three to five for phonemes appeared in the screen. The answer was given by tapping on the dice corresponding to the number of syllables or phonemes in the word. This task is similar to the task called tapping by other authors (see Bryant, MacLean, Bradley & Crossland, 1990), a name we avoid since *Lexiland* includes a different tapping task (tapping to a rhythm). In the *Blending* test (18 syllabic items, 16 phonemic items), children heard a sequence of syllables or phonemes and had to blend them into a word. The phonemes were sounded following usual procedures in phonics instruction (McGuinness, 2004). Children gave their response by pressing the image corresponding to the blended word on screen. The target image was presented along with three distractors (one semantically related, one phonologically related, one unrelated). The location of the correct response was randomized across trials. Within each grain size, items ranged from two to four syllables, and four to six phonemes. In the *Onset matching* test (27 syllabic items, 32 phonemic items) children heard pairs of words —with their corresponding images— and had to indicate whether both words started with the same phoneme/ syllable or not. Children gave their response by pressing on a tick or a cross button. Phonemic items with matching onsets shared the first phoneme for CV word onsets and the first two phonemes for CCV word onset (*flaco-flecha*). Items with non-matching onsets did not share any phonemes in the first syllable. Half of the items had matching onsets and half had non-matching onsets. In the *Rhyming* test (10-word items, 10-pseudowords items), children heard two words —with its corresponding images— or two pseudowords and had to indicate whether they rhymed or not. Children gave their response by pressing on a tick or a cross button. Half of the items rhymed, and half did not. For each task, the score was calculated as mean accuracy.

We used Item response theory (IRT) to estimate each subject's latent phonological awareness score from observed responses. We used a 2-parameter model where the hit probability on each item is modeled as a logistic curve defined by two item parameters: discrimination and difficulty. The model allowed us to estimate the discrimination and difficulty of each item and the subject latent ability. Item and subject parameters are estimated iteratively with Marginal Maximum Likelihood Estimation (MMLE), until reaching a model compatible with the observed data. Further details of estimation of model parameters can be found on Rizopolous (2006) and Baker (2001). In the IRT context, *information* is used to replace the concept of reliability. Information is conceptualized as a function of model parameters and varies with the level of ability. The information of the test is defined as the average information of its items.

Letter knowledge (LK) Letter knowledge was assessed separately for letter sounds and names. In each subtask, children heard the name/sound of a letter and had to choose the answer by tapping the screen to select one of three options: the target letter, a visually similar distractor (Boles & Clifford, 1989), or an unrelated distractor. There were 22 items of each type (for a total of 44). For each task, the score was calculated as mean accuracy.

Rapid Automatized naming (RAN) Children were presented with a 6 x 5 array of five items each repeated six times and were asked to name them as quickly and as accurately as possible. Items were either objects (*gato, jugo, mano, silla, queso* [cat, juice, hand, chair, cheese, respectively]) or colours (*azul, negro, rojo, verde, blanco* [blue, black, red, green, white]). All subtasks were preceded by a familiarization phase. The score was calculated as total response time.

Vocabulary (VOC) Receptive vocabulary was measured through the noun subset of the BEST vocabulary test (De Bruin, Carreiras, & Duñabeitia, 2017). Children heard a word and had to select one out of four images.

Short-term memory Verbal short-term memory (vSTM) was assessed through an adaptation of the task described in Martinez Peres, Majerus and Poncelet (2012)012). Children heard a sequence of monosyllabic words, followed by their corresponding images on the screen (*sol, pan, tren, rey, flor, pez* [sun, bread, train, king, flower, fish]). Children were asked to order the images according to the order of presentation of the words heard. The sequence ranged from two to six items; the test included three trials of each sequence length. The order of presentation in each trial was randomized. The score was the maximum number of items recalled. Non-verbal short-term memory (nvSTM) was assessed through an adaptation of the Corsi Block task (Corsi, 1972)orsi, 1972). Blocks were replaced by pictures of pigs to make the task more attractive to children. Sequences ranged from two to eight items, four trials per sequence length. Testing was interrupted if three errors were made on four consecutive trials of the same length. The score was the maximum number of items recalled.

Nonverbal IQ Nonverbal IQ was measured using the Matrix Reasoning subtest of the Spanish version of the WPPSI (Wechsler, 2001). Scores were computed as the maximum level achieved, following the WPPSI scoring system.

Phonological Awareness, Letter knowledge, RAN, Vocabulary, Short-term memory, and nonverbal IQ measures reported were used in both K5 and G1.

Reading At Time 1 (K5), a list of 15 words and 15 pseudowords was presented on paper and children were asked to read them aloud. All items consisted of 2-syllable high frequency words with varying syllabic complexity including CV and CCV onset items. Pseudowords were constructed from the list of words using Wuggy (Keuleers & Brysbaert, 2010).

At time 2 (G1), reading assessment included two tasks: decoding and fluency. *Decoding*. A list of 30 words and 30 pseudowords was presented digitally, one word per screen. Items consisted of two to three syllable words of medium frequency with simple and complex syllables. Mean accuracy was computed for each child. *Fluency*. A two-minute reading test. It consisted in reading as quickly and as accurately as possible a text of 278 words on paper, within a maximum of 2 minutes. This was a Spanish adaptation of the Alouette French test (Lefavrais, 2005). The number of correct words read per minute was computed for each child.

Data and materials are available upon request to the first author. This study's design and its analysis were not pre-registered.

Results

The rationale for the analysis was as follows. First, we studied the development of phonological awareness from K5 to G1, at the syllabic and phonemic level. In order to control for measurement error, we computed latent ability scores of phonological awareness for each child. Next, we assessed individual reading levels in K5 to identify children who could read and children who could not read at all, and compared the preliteracy skills of K5 readers and non-readers. Finally, we tested our main hypothesis regarding the role of PA in reading acquisition in a transparent orthography using mixed effects linear models of decoding and fluency. All analyses were performed using R software (R Core Team, 2018).

Descriptive statistics for K5 measures and G1 reading are reported in Table 1 (See Table S1 for other G1 measures). Chance denotes the chance level for each task that involved a multiple-choice response format. Composite measures were computed for the two RAN tasks, for the two LK tasks, and for the two decoding tasks (RAN r = 0.56, 95% CI [0.49 – 0.63], $p < .001$; LK r = 0.77, [0.73 – 0.81], $p < .001$; decoding r = 0.96, [0.95 – 0.96], $p < .001$). For PA, latent ability scores were computed from all tasks (See

Table 1. Descriptive statistics for K5 measures and G1 reading

Time		M	SD	min	max	skewness	kurtosis	reliability	chance
K5	Age	5.60	0.29	5.10	6.20	0.03	-1.21	-	-
K5	IQ	10.09	5.42	0.00	28.00	0.60	0.08	0.90	-
K5	Vocabulary	0.83	0.12	0.27	1.00	-1.66	4.18	0.87	0.25
K5	non-verbal STM	3.57	1.23	1.00	6.00	-0.26	-0.53	0.77	-
K5	verbal STM	3.62	1.04	1.00	6.00	-0.05	-0.44	0.75	-
K5	blending phonemes	0.31	0.18	0.00	0.94	1.36	1.87	0.64	0.38
K5	blending syllables	0.83	0.15	0.21	1.00	-1.46	2.37	0.79	0.37
K5	onset matching phonemes	0.54	0.12	0.31	0.97	1.46	2.19	0.64	0.50
K5	onset matching syllables	0.59	0.15	0.19	1.00	0.81	-0.04	0.74	0.50
K5	rhyme pseudowords	0.54	0.16	0.00	1.00	0.34	0.99	0.41	0.50
K5	rhyme words	0.57	0.17	0.10	1.00	0.57	0.53	0.44	0.50
K5	segmentation phonemes	0.27	0.12	0.07	0.96	2.00	7.28	0.62	0.25
K5	segmentation syllables	0.41	0.17	0.09	1.00	1.36	2.13	0.76	0.33
K5	RAN colours	56.64	19.05	20.17	125.10	1.22	1.91	-	-
K5	RAN objects	48.53	12.60	24.92	97.72	0.90	1.30	-	-
K5	letter name	0.60	0.23	0.09	1.00	0.14	-1.18	0.87	0.33

Units: Vocabulary, Blending, Onset matching, Rhyme, Segmentation, Letter and Decoding: mean accuracy; IQ, non-verbal STM, verbal-STM: maximum level achieved; RAN: total response time; Fluency: words read correctly per minute. Reliability is McDondald's omega (total). Reliability of the average Phonological Awareness score is 0.85 (when including all items) and for Letter Knowledge is 0.91. N = 388.

Table 1 (Cont.). Descriptive statistics for K5 measures and G1 reading

Time		M	SD	min	max	skewness	kurtosis	reliability	chance
K5	letter sound	0.55	0.21	0.14	1.00	0.32	-0.85	0.81	0.33
G1	decoding words	0.75	0.34	0.00	1.00	-1.32	0.22	-	-
G1	decoding pseudowords	0.68	0.32	0.00	1.00	-1.14	-0.09	-	-
G1	fluency	21.13	15.98	0.00	99.00	1.13	2.38	-	-

Units: Vocabulary, Blending, Onset matching, Rhyme, Segmentation, Letter and Decoding: mean accuracy; IQ, non-verbal STM, verbal-STM: maximum level achieved; RAN: total response time; Fluency: words read correctly per minute. Reliability is McDondald's omega (total). Reliability of the average Phonological Awareness score is 0.85 (when including all items) and for Letter Knowledge is 0.91. N = 388.

section on IRT analysis). Correlations among all variables measured in K5 and reading measured in G1 were studied to assess collinearity issues for model building (Table 2). The strongest correlations among K5 measures were between LK and PA, LK and vocabulary, and LK and verbal short-term memory. The strongest correlations between K5 variables and G1 reading were for LK, followed by RAN and non-verbal STM. All correlations were significant at the 99% level with p values corrected to through false-discovery rate.

Table 2. Pearson correlation coefficients for K5 variables and G1 reading measures

	Time		1	2	3	4	5	6	7	8
1	G1	decoding								
2	G1	fluency	0.67***							
3	K5	PA	0.26***	0.19***						
4	K5	IQ	0.21***	0.19***	0.28***					
5	K5	Voc	0.27***	0.16**	0.27***	0.27***				
6	K5	nvSTM	0.36***	0.28***	0.26***	0.26***	0.27***			
7	K5	vSTM	0.38***	0.32***	0.27***	0.26***	0.24***	0.35***		
8	K5	RAN	-0.38***	-0.34***	-0.15**	-0.24***	-0.28***	-0.31***	-0.28***	
9	K5	LK	0.50***	0.50***	0.36***	0.31***	0.38***	0.31***	0.44***	-0.34***

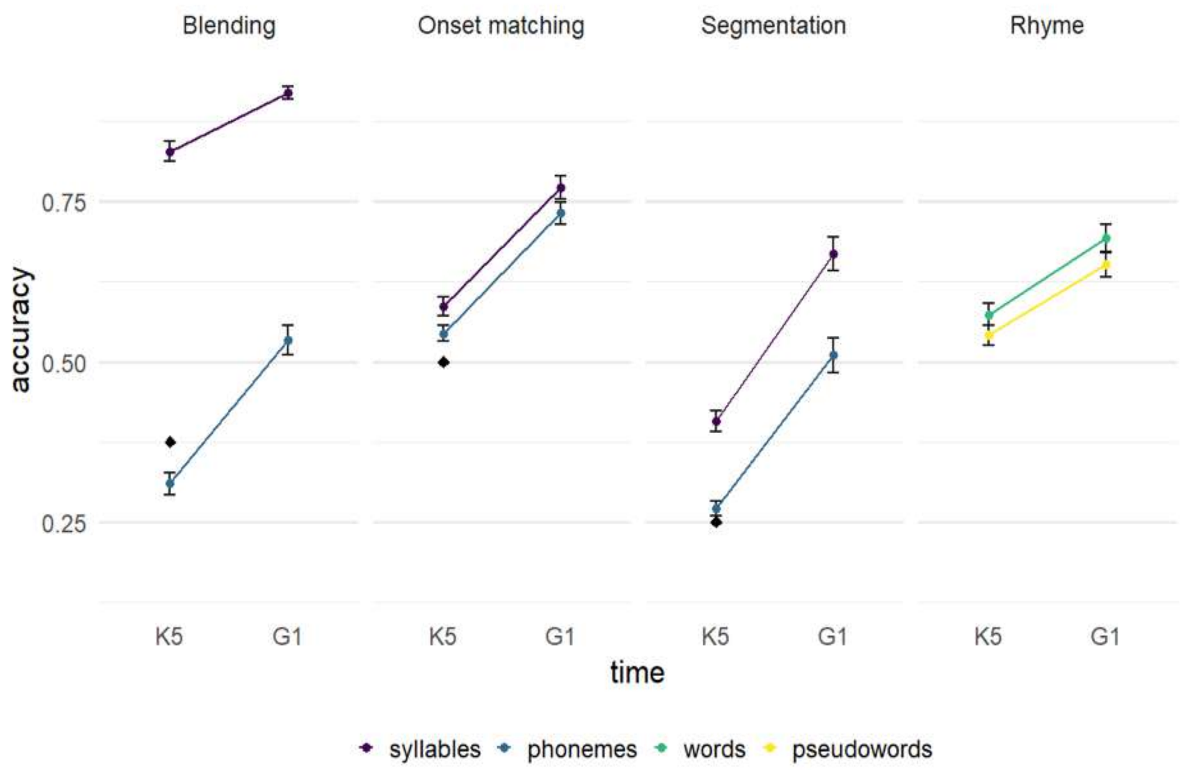
*** p < .001, ** p < .01 false discovery rate correction
Note. PA: phonological awareness, Voc: vocabulary, nvSTM: non-verbal short-term memory, vSTM: verbal short-term memory, RAN: rapid automatized naming, LK: letter knowledge. N=388.

Phonemic Awareness shows Floor Effects in Kindergarten In order to evaluate performance across measures and time points (Figure 1), we first performed one-sample t-tests of accuracy scores against chance, since all tasks were presented in a multiple-choice format. Children performed better than chance across all tasks (p < .001), except for blending phonemes, where average performance was significantly below chance level (M= 0.31, chance = 0.37, 95% CI [0.29, 0.33], t = -7.038, df = 387, p < .001). Notably though, performance in the other two PA tasks involving phonemes was barely above chance (segmentation phonemes = 0.27, chance = 0.25, 95% CI [0.26, 0.28]; onset matching phonemes = 0.54, chance = 0.5 [0.53, 0.55]).

Next, since PA skills were assessed both in K5 and G1, we could evaluate growth in PA skills across time (Figure 1). A linear mixed effect model with accuracy as the outcome and task, time, and task-time interaction as predictors showed significant effects for all predictor variables, including the time-task interaction. Post-hoc comparisons for each task across time points showed significant improvements in accuracy for all tasks (all p < 0.001, corrected through false discovery rate).

IRT Analysis shows Progression of Difficulty from Syllabic to Phonemic items Since the PA construct was measured by four tasks varying in difficulty and cognitive load, we estimated a latent ability score for each child by combining all tasks measured in K5. This served two ends. First, it enabled us to directly compare difficulty levels among tasks. Second, it controlled measurement error (Cole & Preacher, 2014; Hjetland et al., 2019). We estimated a 2PL model from the 163 phonological awareness items via the *ltm* package (Rizopoulos, 2006). Previous evidence shows that PA is a unitary construct, an assumption of IRT models (Anthony et al., 2011; Vloedgraven & Verhoeven, 2009). Model fit indices showed excellent fit (additional details on model fit available in Supporting information). Twelve items were excluded due to extreme difficulty parameters. The final, reduced, model included 151 phonological awareness items.

Person-level analysis Having established adequate model fit, latent ability scores were computed for each child from the reduced model via Empirical Bayes through the *factor.scores* function in the *ltm* package (Rizopoulos, 2006). Pearson correlation coefficient



Note. For all tasks, syllabic performance was significantly better than phonemic performance. Phonemic performance was barely above chance levels for Onset matching and Segmentation and significantly below chance level for Blending. All PA skills improve with time from K5 to G1. Error bars represent 95% confidence interval. Diamonds represent chance levels for tasks involving phonemes. N = 388.

Figure 1. Phonological Awareness Performance Across Tasks, Grain Size and Time Points

between latent scores obtained from the complete model and from the reduced model was .99. Obtained latent ability scores were normally distributed around 0 (min: -2.9, max: 2.7).

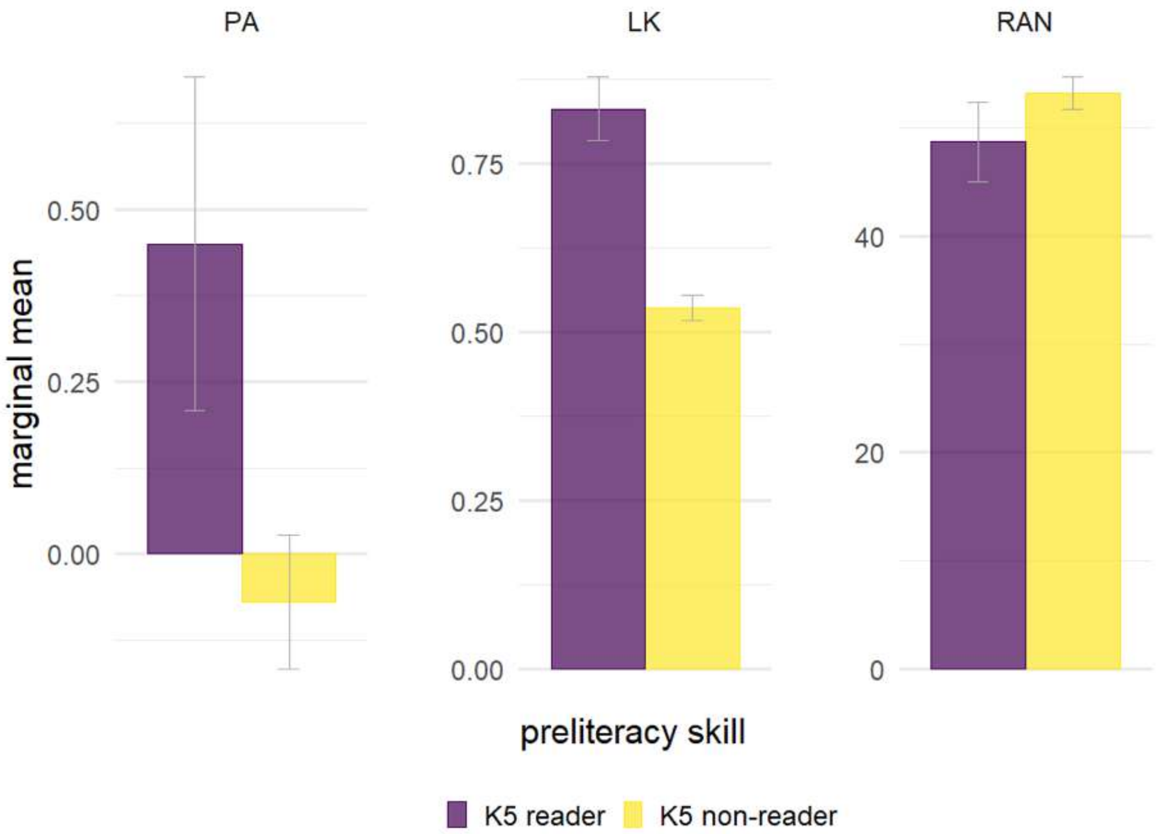
Item-level analysis Difficulty was examined for each task and grain size (that is, syllabic vs. phonemic items). Overall average difficulty for the reduced model was .5. Tasks arranged from less to more difficult were blending (M = −1.23) < rhyme (M = 0.33) < onset matching (M= 0.61) < segmentation (M= 1.87). Pairwise comparisons through two-sample t-test showed significant differences between blending and onset matching (t(147) = −184, *p* = .019) and between blending and segmentation (t(147) = −3.10, *p* < .001). With respect to grain size, syllabic items were significantly less difficult than phonemic ones (M syllables = −0.62, M phonemes = 1.65, t(129) = 2.27, *p* < .001). These results are consistent with the expected progression of development of phonological awareness from syllabic to phonemic units, and from blending to identifying to segmenting (Anthony et al., 2003, 2011; Ziegler & Goswami, 2005). Regarding discrimination parameters, average discrimination was 0.3, with tasks arranged from less to more discriminative: segmentation (M = 0.24) < blending (M = 0.35) < rhyme (M = 0.38) < onset matching (M = 1.16). Pairwise comparisons showed significant differences between segmentation and onset matching (t(147) = 0.92, p < .001), blending and onset matching (t(147) = −0.80, p < .001), and rhyme and onset matching (t(147) = 0.77, p < .001). No significant differences were observed in discrimination parameters between syllabic and phonemic items.

Finally, we computed test information in order to assess internal consistency. We found that the peak of the information function is located at 0.03, suggesting the estimation of latent ability scores for the PA construct are most precise at average latent ability score levels, with a standard deviation of 0.19. Thus, error of measurement was very low for most children with scores around average, and increased towards the edges, as expected, with a standard deviation of approximately 0.4 for latent ability scores of 2 and -2. For approximately 95% of the sample, which falls between -2 and 2 standard deviations of the mean, the error in estimation was very low.

Preliteracy Skills differ between K5 Readers and Non-readers In order to assess the unique contribution of PA to reading *before any reading experience*, children were tested on their reading levels in K5 through a list of 15 words and 15 pseudowords. Children are not expected to have reading skills at this stage as reading is not explicitly taught in kindergarten. Accordingly, 86.3% of the sample could not decode any pseudowords, while only 11.3% correctly decoded more than 10 pseudowords. In order to make sure that our results refer to pre-readers avoiding reciprocal effects, we conservatively defined *K5 readers* as those that decoded one or more pseudowords correctly, which constituted 13.9% of the sample. We used pseudoword decoding as a criterion for reading because it excludes the use of any familiar whole-word recognition strategies, which do not imply the decoding ability.

Following the vast literature on the role of preliteracy skills in reading acquisition, we compared K5 readers vs. non-readers in each preliteracy skill using one linear regression model per task, with task score (for PA and LK) or response time (for RAN) as outcomes, and Age, IQ and group (K5 reader vs. non-reader) as predictors (Figure 2). In all models, the group coefficient was significant at the 99% confidence level. Planned comparisons of marginal means showed that *K5 readers* outperformed non-readers in all preliteracy skills. All K5 readers were removed from further analysis in order to avoid reciprocal effects of PA and reading.

RAN and LK, but not PA, Uniquely Contribute to Reading Skills In order to evaluate the unique contribution of PA to early reading abilities, that is, variance explained while controlling for relevant covariates, we run linear mixed effects regression models with preliteracy skills measured in K5 as predictors (LK, RAN and PA), and two outcome variables: decoding (composite of words and pseudowords accuracy) and fluency (words read per minute) measured in G1. The inclusion of the PA variable in the regression models presented a challenge due to a. the floor effects observed for phonemic items, and b. the low reliability for rhyme tasks. In order



Note. Marginal means, controlling for Age and IQ. Error bars represent 95% confidence interval. Marginal means represent latent ability scores for PA, mean accuracy for LK, and response times in seconds for RAN (smaller scores mean better performance). K5 readers outperform K5 non-readers across all measures. N readers = 54, N non-readers = 334.

Figure 2 . Preliteracy Skills Performance of K5 Readers vs. Non-Readers at Time 1

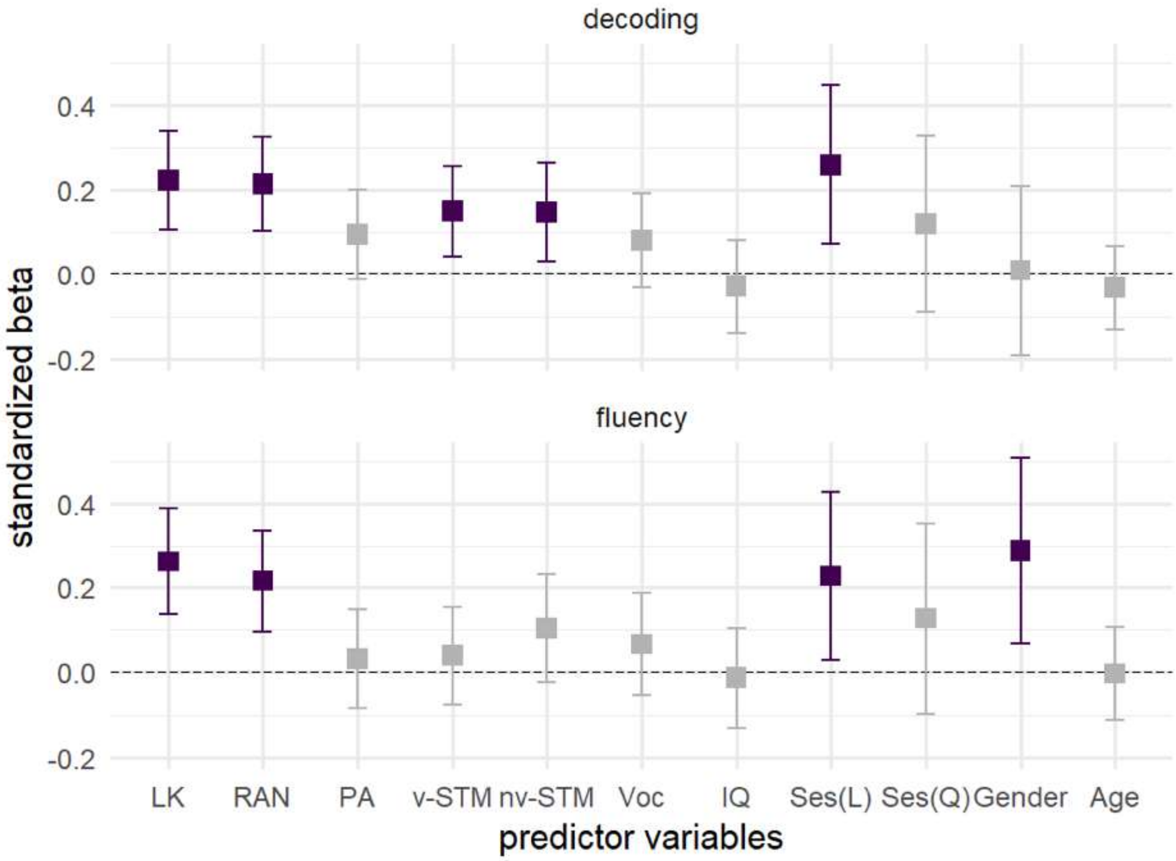
to overcome these challenges, we used latent ability scores obtained from the IRT analysis (See previous section about IRT Analysis), excluding the rhyme tasks. Additionally, all reported models were also fit with latent ability scores for phonemic awareness and syllabic awareness items separately. Results for the full model remained the same and are thus not reported.

School was included as random intercept to account for the nesting of children across schools. Age, Gender, IQ, Vocabulary, vSTM, nvSTM, and Maternal Education, as a proxy for socioeconomic status (SES), were included as control variables. vSTM was treated as a control variable in order to focus on the core component of the PA construct and because of the large memory load involved in some of the PA tasks. Since PA and reading have shown reciprocal effects (Castles & Coltheart, 2004), all children that showed any reading skill in K5 were excluded from the analysis. For this reason, we refer to PA skills in these children as PPA (pre-reading phonological awareness).

For model specification and selection, we followed Meteyard and Davies (2020) recommended practices on linear mixed-effects models. First, a *null model* containing only a random intercept for School was fitted. No random slopes were added since the number of children by school was low for estimation purposes. Model building continued from minimal to maximal. In the next step we computed the *preliteracy model*, adding three preliteracy skills of interest as fixed effects: PPA, LK, and RAN. Finally, we ran the *full model*, in order to assess the unique contribution of preliteracy skills *after controlling for relevant covariates,* adding all covariates as fixed effects (Age, Gender, SES, IQ, Vocabulary, vSTM and nvSTM). Model details are reported in Table S2.

The *null models*, containing only the random effect for School explained approximately 10% of the variance in decoding and 6% in fluency. In the *preliteracy models*, LK, RAN, and PPA all contributed uniquely to decoding. LK and RAN, but not PPA, contributed uniquely to fluency. All variables combined explained 39% of the variance in decoding, and 31% of the variance in fluency. Both models (accuracy and fluency) significantly improved model fit as compared to the null model. In the *full models* (Figure 3), which included all relevant covariates in addition to preliteracy skills, LK and RAN still contributed unique variance among preliteracy skills (Table S3). Crucially, PPA no longer contributed unique variance to decoding. Among covariates, vSTM, nvSTM, and SES all contributed unique variance to decoding. For fluency, SES and Gender were unique predictors (with boys outperforming girls). Overall, the full models accounted for 45% of the variance in decoding and 38% of the variance in fluency. As for variance explained by each predictor of interest while keeping all other variables constant, for decoding, PPA contributed 2.2% of additional unique variance, LK contributed 6.4% and RAN 6.0%. For fluency, PPA contributed 0.3%, LK 7.7% and RAN 4.8 %.3.6 % to fluency. Both full models (accuracy and fluency) significantly improved model fit as compared to the preliteracy skills models.

For the decoding model, the lack of a PPA effect in the presence of covariates was further examined. We reasoned that if the effect of PPA on reading was modulated by any of these control factors, as evidenced by the change in the model coefficient for PPA, interaction effects were likely. Thus, we estimated three new models including interaction terms between PPA and verbal short-term memory (model 1), non-verbal short-term memory (model 2), and SES (model 3). The only significant interaction effect observed was for PPA and SES (Table S4). An examination of the pattern of reading-PPA relations by SES group showed that the PPA-reading relation was stronger for the low than the middle and high SES groups. This new model significantly improved model fit over the full model without any interaction terms (delta r squared = 3.1%, LRT $X^2(2) = 13.69$, $p < .001$). In order to check the sensitivity of our results to changes in the cut-off value separating readers from non-readers, we re-run the analysis with a threshold of 5 pseudowords (Supporting information Table S5). The results did not



Note. Prediction model coefficients for decoding (top panel, n = 243) and fluency (bottom panel, n = 239). School was included as a random intercept (not shown). Error bars represent 95% confidence intervals. Colour shows significant predictors for each model (different from zero). LK: Letter Knowledge, RAN: Rapid Automatized Naming, vSTM: verbal Short-Term Memory, nv-STM: non-verbal Short-Term Memory, Voc: Vocabulary, SES: Socio-economic Status. RAN coefficients are reversed for illustration purposes. For SES, since it is an ordinal variable, L indicates a coefficient for a linear term, and Q for a quadratic term.

Figure 3. Regression Coefficients for Prediction Model of Reading from Preliteracy Skills

change after this manipulation. Moreover, we evaluated a model of decoding (which is the dependent variable more likely to show PA effects) including K5 readers and a categorical variable representing whether they read or not at K5 (Table S6). The model does not show an independent effect of PA above and beyond all other predictors.

Discussion

In the current study we assessed the unique contribution of PA to early reading skills in a transparent orthography. By computing latent ability scores from a comprehensive PA battery, we overcame the floor effects of PA often reported for more transparent orthographies. As in many previous studies, PA measured in K5 is correlated with reading scores measured in G1. Nevertheless, in two regression models of decoding and fluency, we showed that pre-reading phonological awareness (PPA) does not uniquely contribute to the prediction of early reading acquisition above and beyond other preliteracy skills, while controlling for several relevant covariates. We have also shown that our failure to find a contribution of PPA above and beyond the other factors is not due to lack of power. Instead, we showed that LK and RAN (and vSTM in the case of decoding) are the most relevant predictors of early reading skills. Crucially, our prediction models can account for large amounts of variance (38% and 45% for decoding and fluency respectively) even in the absence of a significant unique contribution from PPA. Our findings shed light on how the dynamic interplay among preliteracy skills may reveal itself across orthographies.

Development of PA in a Transparent Orthography

As reported in studies of PPA in prereaders in more transparent orthographies, phonemic awareness showed floor effects (at chance or barely above chance levels) in our sample, as evidenced by average scores and by difficulty parameters in the item-response theory model. Floor effects have been a main explanatory reason for not finding a unique contribution of phonemic awareness to reading in more transparent orthographies (De Jong & Van der Leij, 2003; Georgiou et al., 2012; Landerl et al., 2019; Van Bergen et al., 2011). However, while this argument makes methodological sense –one would expect no significant unique contribution when the predictor does not show sufficient variance— its theoretical interpretation should not be dismissed. Why is it common to see floor effects in phonemic awareness measures in kindergarten children from languages with more transparent orthographies? In line with previous studies, our results suggest that that *phonemic awareness* develops “late but fast” (Defior et al., 2008; Mann & Wimmer, 2002). Moreover, it is possible that in opaque orthographies PA is stimulated during kindergarten and that this is not the case in transparent orthographies, as it is felt by teachers and curriculum designers that this is a skill that can be learned easily. In line with this possibility, Ziegler and Goswami (2005) review evidence that children learning to read in transparent orthographies show little phonemic awareness in K5 (e.g., Italian or Greek) but develop it quickly during first grade.

Unique Contribution of PPA to Reading Acquisition

Results from the full regression models for both decoding and fluency show that PPA does not contribute uniquely to reading acquisition above and beyond other preliteracy skills when critical covariates are included. The exclusion of K5 readers in our model could raise the question of whether we are artificially reducing the effect that PA can have in reading in G1. However, as shown above, results are robust even when changing the threshold separating readers from non-readers.

The comprehensive assessment and large sample size in our study confirm that the null unique contribution from PPA was not a result of measurement error or lack of power. In addition, the reported pattern of results showing a significant correlation between PA skills in K5 and reading scores in G1**, the** growth of PA skills from K5 to G1, the increased performance for syllabic to phonemic items, the separation across cognitive operations, and the discrimination between K5 readers and non-readers, strongly suggest that the null contribution does not stem from random behaviour. These findings add converging evidence from a Spanish speaking population to the available studies on more transparent orthographies such as Dutch, German, Finnish, and Greek (De Jong & Van der Leij, 2003; Defior et al., 2008; Georgiou et al., 2012; Landerl et al., 2019; Mann & Wimmer, 2002; Schmitterer & Schroeder, 2019; Van Bergen et al., 2011). On the other hand, these results contradict those reported by Caravolas and colleagues (2012) in their longitudinal crosslinguistic study including Spanish. A possible explanation for the discrepancy is that in their study children had some reading experience at study onset. This could have prompted the development of PA. The present results, in contrast, come from a sample of children who, at study onset, could not decode any pseudowords; therefore, no reciprocal effects were expected. The reciprocal effects of reading on the development of PA could unfortunately not be tested in the present sample since the proportion of readers at study onset was very low (13%). This did not warrant inclusion of an interaction term in the model, nor building a separate model specifically for those children. Additionally, in our study, unlike that by Caravolas et al., we see a significant unique contribution from pre-reading vSTM to reading. In their study, Caravolas et al. (2012) cite the low reliability of vSTM as an explanatory factor, noting it did not make a unique contribution to reading prediction. This suggests they may have found a pattern of results similar to ours if the vSTM measure had been more reliable in their study. Also, the decoding measures used in their study and ours differed considerably. With regard to other more transparent orthographies, results on Finnish are also pertinent to our findings, since, like Spanish, Finnish can be categorized at the extreme of orthographic consistency. In a study reported in Puolakanaho et al. (2007), preliteracy skills were compared in a sample of 200 children from 3.5 years of age, half of whom had a family history of dyslexia. Although they reported PA as a longitudinal predictor of reading skills in pre-reading children, this effect was only observed at a time point where RAN was not measured. At the other two time points, in which RAN was measured, PA did not show any effect above LK and RAN. Moreover, differences in sample composition between their study and ours likely had consequences for the findings. The Finnish sample was enriched by children with a family risk of dyslexia, while the present study was composed of an unselected sample of children.

The sum of evidence from longitudinal studies on transparent orthographies, casts doubt on a universal role for PPA as predictor of reading acquisition. Nevertheless, we should ask if PPA has any role to play in such reading acquisition. Landerl and colleagues (2019) have put forward an account based on their results from a crosslinguistic longitudinal study of preliteracy skills in English, French, German, Dutch, and Greek. Having found a complex pattern of prediction across orthographies, they propose that PA in more transparent orthographies may develop as a co-requisite rather than as a prerequisite of reading acquisition. We believe that a dynamic interplay between PPA and LK can accommodate the observed pattern. Following Mann and Wimmer's (2002) thesis, in line with the proto-literacy hypothesis (Barron, 1991), "phoneme awareness must be triggered by something above and beyond the experiences that are sufficient to support primary language development" (2002, p. 676). That "something" might come from explicit letter name/sound instruction or from explicit phonological awareness activities. In the former case, at an initial point in time, we should see LK as a main predictor of future decoding and none or only a small unique contribution from PPA. In the latter, we would see a main role for PPA. From an interactive LK-PA standpoint (Hulme et al., 2005; Kim et al., 2010; Piasta & Wagner, 2010) both skills should develop later on. This account would seem to suggest that the differences observed in prediction patterns for decoding are just a matter of differences in kindergarten instruction or home literacy environments across countries. However, a further point can be made. When both skills are present, their relative contribution differs across orthographies based on the amount of information they convey (Vousden, Ellefson, Solity, & Chater, 2011). In less transparent orthographies, where the number of phonemes tends to be larger than the number of letters, thus requiring more complex graphemes to represent them, the ability to identify and manipulate phonemes (i.e., PA) has a larger explanatory value than knowing the letters. In such orthographies, knowledge of letter sounds is not enough to correctly sound out words. Therefore, in a predictive model, both skills will contribute significant and independent amounts of variance to explaining early reading acquisition. On the contrary, in more transparent orthographies, given the almost one to one mapping between graphemes and phonemes, letter sounds are virtually equivalent to the phonemes they represent. As pointed out, in reference to Finnish, "Because the Finnish language is so transparent, letter sound knowledge and phonemic awareness are near synonymous, and consequently, once mastery of the alphabetic principle, i.e., sounds of the letters, has been achieved, reading is underway" (Lyytinen, Erskine, Hämäläinen, Torppa, & Ronimus, 2015, p. 334). In this case, LK indexes children's ability to identify phonemes, thus replacing PA as a main contributor to later reading acquisition. What we measure as LK includes in part what we measure as PA.

In sum, the unique contributions of PA and LK as longitudinal predictors of decoding abilities are the result of a combination of kindergarten instructional practices, the home literacy environment and the differential information content contributed by LK and PA across orthographies. In a transparent orthography, a child with good LK, good memory, and good lexical access (i.e., RAN), only needs to learn to synthesize to be able to decode words, whereas this is not enough in an opaque orthography, and the child still needs to refine her knowledge of the grapheme – phoneme mappings, as it is not one-to-one. That is why in more transparent orthographies reading and PA develop concurrently (and those children are excluded from our prediction), whereas there is an intermediate step of good PA and letter knowledge of non-readers in more opaque orthographies.

PA Tasks: Response Format and Procedure

An additional difference between this and previous studies is the operationalization of PA. Probably, the most critical difference stems from response formats. The PA construct is frequently measured through verbal responses, while in our tasks all responses were given in a multiple-choice format. Two points need to be considered when analysing this difference. First, despite the change in response format, we successfully replicated the developmental trajectories and the difficulty pattern reported in previous studies, both within and across testing times. Second, as stated before, Cunningham and colleagues (2015) have shown that producing a verbal response explains unique variance in the PA-reading relation, above and beyond that explained by comparison measures (the same task) with no verbal response. Clearly, this additional dimension of PA is lacking in our study. However, we see no reason, in principle, to include a verbal response as part of the core construct of PA. Also, by displaying response options on screen (and accompanying auditory stimuli with a visual representation) we have substantially decreased the memory load involved in solving the task. Thus, the response format used helps separate PA from speech production and memory. In this way, our PA tasks are tapping into the PA construct, albeit through a different measurement than usual.

PA and Verbal Short-term Memory

A surprising finding from this study was the relevant role that pre-reading vSTM plays in the prediction of decoding skills. We originally included vSTM as a covariate, in order to control for the large memory load that PA tasks place on participants. However, as stated earlier, vSTM belongs to the broader construct of phonological skills important for reading acquisition, which includes PA and RAN in addition to vSTM. Hence, vSTM it is sometimes treated as a preliteracy skill *per se* (Moll et al., 2014; Ramus & Szenkovits, 2008; Torgesen et al., 1994), sometimes treated as a covariate (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Puolakanaho et al., 2007; Vaessen et al., 2010), and sometimes treated as a single phonological construct together with PA (Wagner & Torgesen, 1989; Knoop-van Campen, Segers, & Verhoeven, 2018; Martinez Perez et al., 2012; Moll et al., 2014). The present results suggest that vSTM predicts reading skills above and beyond other preliteracy skills and other general cognitive factors. We argue that this result can be explained by the underlying cognitive operations involved in learning to read in a transparent orthography. As stated before, given the almost one to one mapping between graphemes and phonemes, and thus the strong information content of letter sounds, converting each grapheme into its corresponding phoneme is almost trivial when there is advanced knowledge of letter sounds. Once this first step has been achieved, the next most critical operation is maintaining these letter sounds in memory to blend them. Thus, in more transparent orthographies, strong letter knowledge and memory skills are paramount for successfully acquiring early reading skills.

Limitations

Languages vary not only in their orthographic consistency but also in properties of the oral language itself, such as the rhythm of their syllabic structure. It is possible that these, less explored, properties also influence the development of PA and thus the PA-reading relation. For example, rhythmic properties vary in stressed-timed languages and syllable-timed languages, such as English and Spanish, respectively. Rhythm, in turn, has recently been given more attention in defining the process of speech segmentation, which, in turn, affects the development of phonological skills (Wood & Connelly, 2009). While these linguistic properties have been much less explored, a provocative study across six alphabetic orthographies varying in consistency, syllabic structure, and rhythm found rhythm explained differences in the development of phonological awareness better than orthographic consistency (Duncan et al., 2013). The role of these other linguistic properties should be further explored in order to better understand how they interact with orthographic consistency to modulate the development of PA and reading acquisition.

A second limitation is the lack of information acquired on teaching practices. While assessment of teaching practices was beyond the scope of the present study, there is large variability in the methods used for teaching reading in Uruguay. We are aware that variations in teaching methods might impact both the development of PA skills and the PA-reading relation. As mentioned before, it might underlie the little development of PA skills in K5. Including teaching practice as an additional variable in our model might shed further light on the conditions under which PPA uniquely contributes to reading acquisition and how this is modulated by teaching practices.

Finally, an additional factor that needs to be considered is the fact that children were tested in groups, which could lead to less focused attention and, consequently, impaired understanding of the instructions. However, the reliability of the tasks, the correlation matrix, and the developmental trajectories observed, suggest that children did understand the instructions and tried to complete each task to the best of their capacity.

To summarise, we found that in a transparent orthography, among children who had not yet developed reading, PPA does not make a unique contribution to the prediction of later reading acquisition, that is, when other variables such as letter knowledge, memory and RAN are included. These results cannot be explained by measurement error in PA, as has been cautioned with respect to previous studies. Instead, we found that the strongest contributors to decoding were RAN, LK (and vSTM), while the strongest contributors to fluency were RAN and LK. We propose that a delayed developmental trajectory for PA, a strong role for vSTM, and a dynamic interplay between LK and PA —influenced by home literacy and educational practices as well as the intrinsic characteristics of the orthographic system— can accommodate these and previous results.

Acknowledgments

This project was funded by ANII FSED_2_2015_1_120741 and ANII FSED_2_2016_1_131230 grants. Camila Zugarramurdi received a PhD Scholarship from Fundación Carolina. We thank Alvaro Cabana for helpful insights on data analysis, and all the families and schools involved in the study.

Data availability statement

The data that support the findings of this study are available from Camila Zugarramurdi (czugarramurdi@psico.edu.uy) upon reasonable request. This study's design and its analysis were not pre-registered.

Disclosures

The authors declare that the research was conducted is the absence of any commercial or financial relationships that should be construed as a potential conflict of interest. Manuel Carreiras and Juan Valle-Lisboa should be both considered senior authors of this publication.

References

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38*(4), 470–487.

Anthony, J. L., Williams, J. M., Durán, L. K., Gillam, S. L., Liang, L., Aghara, R., … Landry, S. H. (2011). Spanish Phonological Awareness: Dimensionality and Sequence of Development During the Preschool and Kindergarten Years. *Journal of Educational Psychology*, *103*(4), 857–876. https://doi.org/10.1037/a0025024

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). *Severity of Emotional and Behavioral Problems Among Poor and Typical Readers*. *Journal of Abnormal Child Psychology* (Vol. 33, pp. 205–217). https://doi.org/10.1007/s10802-005-1828-9

Barron, R. W. (1991). Proto-literacy, literacy and the acquisition of phonological awareness. *Learning and Individual Differences*, *3*(3), 243–255. https://doi.org/10.1016/1041-6080(91)90010-X

Baker, F. B. (2001). *The Basics of Item Response Theory* (C. Boston & L. Rudner (eds.); Second). ERIC Clearninghouse on Assessment and Evaluation

Boets, B., Wouters, J., van Wieringen, A., & Ghesquière, P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: a longitudinal study of the auditory temporal processing theory. *Neuropsychologia*, *45*(8), 1608–1620. https://doi.org/10.1016/j.neuropsychologia.2007.01.009

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*(6), 579–586. https://doi.org/10.3758/BF03210580

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(4), 422–432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. Developmental Psychology, 26(3), 429-438. https://doi.org/10.1037/0012-1649.26.3.429

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., … Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*(6), 678–686. https://doi.org/10.1177/0956797611434536

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. https://doi.org/10.1037/a0033805

Corsi, P. (1972). Memory and the medial temporal region of the brain. *Unpublished Doctoral Dissertation), McGill University, Montreal, QB*.

Cunningham, A. J., Witton, C., Talcott, J. B., Burgess, A. P., & Shapiro, L. R. (2015). Deconstructing phonological tasks: The contribution of stimulus and response type to the prediction of early decoding skills. *Cognition*, *143*, 178–186. https://doi.org/10.1016/j.cognition.2015.06.013

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST Dataset of Language Proficiency. *Frontiers in Psychology*, *8*(April). https://doi.org/10.3389/fpsyg.2017.00522

De Jong, P. F., & Van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in Dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22–40. https://doi.org/10.1037/0022-0663.95.1.22

Defior, S., Serrano, F., & Marín-Cano, M. J. (2008). El poder predictivo de las habilidades de conciencia fonológica en la lectura y escritura en castellano. En. *Estudios de Desarrollo Del Lenguaje y Educación. Studies on Language Development and Education*, (January), 339–347.

Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H. K., Baillie, S., Leybaert, J., … Serrano, F. (2013). Phonological development in relation to native language and literacy: Variations on a theme in six alphabetic orthographies. *Cognition*, *127*(3), 398–419. https://doi.org/10.1016/j.cognition.2013.02.009

Eckert, M. A., Vaden, K. I., & Gebregziabher, M. (2018). Reading profiles in multi-site data with missingness. *Frontiers in Psychology*, *9*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2018.00644

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Furnes, B., Elwér, Å., Samuelsson, S., Olson, R. K., & Byrne, B. (2019). Investigating the Double-Deficit Hypothesis in More and Less Transparent Orthographies: A Longitudinal Study from Preschool to Grade 2. *Scientific Studies of Reading*, *23*(6), 478–493. https://doi.org/10.1080/10888438.2019.1610410

Furnes, B., & Samuelsson, S. (2010). predicting reading and spelling difficulties in transparent and opaque orthographies: *Dyslexia*, *239*, 234–239. https://doi.org/10.1002/dys

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, *25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54. https://doi.org/10.1038/nrn3836

Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, *111*(5), 751–763. https://doi.org/10.1037/edu0000321

Huettig, F., Lachmann, T., Reis, A., & Petersson, K. M. (2018). Distinguishing cause from effect–many deficits associated with developmental dyslexia may be a consequence of reduced and suboptimal reading experience. *Language, Cognition and Neuroscience*, *33*(3), 333–350. https://doi.org/10.1080/23273798.2017.1348528

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological Skills Are (Probably) One Cause of Success in Learning to Read: A Comment on Castles and Coltheart. *Scientific Studies of Reading*, *9*(4), 351–365.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The Contributions of Phonological Awareness and Letter-Name Knowledge to Letter-Sound Acquisition-A Cross-Classified Multilevel Model Approach. *Journal of Educational Psychology*, *102*(2), 313–326. https://doi.org/10.1037/a0018449

Knoop-van Campen, C. A. N., Segers, E., & Verhoeven, L. (2018). How phonological awareness mediates the relation between working memory and word reading efficiency in children with dyslexia. *Dyslexia*, (May 2016), 156–169. https://doi.org/10.1002/dys.1583

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., … Georgiou, G. K. (2019). Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency. *Scientific Studies of Reading*, *23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

Lefavrais, P. (2005). Test de l'Alouette Révisé. *Paris: Editions Du Centre de Psychologie Appliquée*.

Lyytinen, H., Erskine, J., Hämäläinen, J., Torppa, M., & Ronimus, M. (2015). Dyslexia—Early Identification and Prevention: Highlights from the Jyväskylä Longitudinal Study of Dyslexia. *Current Developmental Disorders Reports*, *2*(4), 330–338. https://doi.org/10.1007/s40474-015-0067-1

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006). Trajectories of Reading Development: A Follow-up From Birth to School Age of Children With and Without Risk for Dyslexia. *Merrill-Palmer Quarterly*, *52*(3), 514–546. https://doi.org/10.1353/mpq.2006.0031

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing*, *15*(7–8), 653–682. https://doi.org/10.1023/A:1020984704781

Martinez Perez, T., Majerus, S., & Poncelet, M. (2012). The contribution of short-term memory for serial order to early reading acquisition: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *111*(4), 708–723. https://doi.org/10.1016/j.jecp.2011.11.007

McBride-Chang, C. (1995). What Is Phonological Awareness? *Journal of Educational Psychology*, *87*(2), 179–192. https://doi.org/10.1037/0022-0663.87.2.179

McGuinness, D. (2004). *Early reading instruction*. MIT Press.

Melby-Lervåg, M., Lyster, S.-A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*(March 2019), 104092. https://doi.org/10.1016/j.jml.2020.104092

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., … Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the 'dyslexia paradox': Reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 156–176. https://doi.org/10.1002/wcs.1383

Papadopoulos, T. C., Spanoudis, G., & Kendeou, P. (2009). The Dimensionality of Phonological Abilities in Greek. *Reading Research Quarterly*, *44*(2), 127–143. https://doi.org/10.1598/rrq.44.2.2

Piasta, S. B., & Wagner, R. K. (2010). Learning letter names and sounds: Effects of instruction, letter type, and phonological processing skill. *Journal of Experimental Child Psychology*, *105*(4), 324–344. https://doi.org/10.1016/j.jecp.2009.12.008

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Poikkeus, A., Tolvanen, A., & Torppa, M. (2007). Very early phonological and language skills: estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, *9*, 923–931. https://doi.org/10.1111/j.1469-7610.2007.01763.x

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology (2006)*, *61*(1), 129–141. https://doi.org/10.1080/17470210701508822

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling. *Journal Of Statistical Software*, *17*(5), 1–25.

Runge, T. J., & Watkins, M. W. (2006). The structure of phonological awareness among kindergarten students. *School Psychology Review*, *35*(3), 370–386.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic*

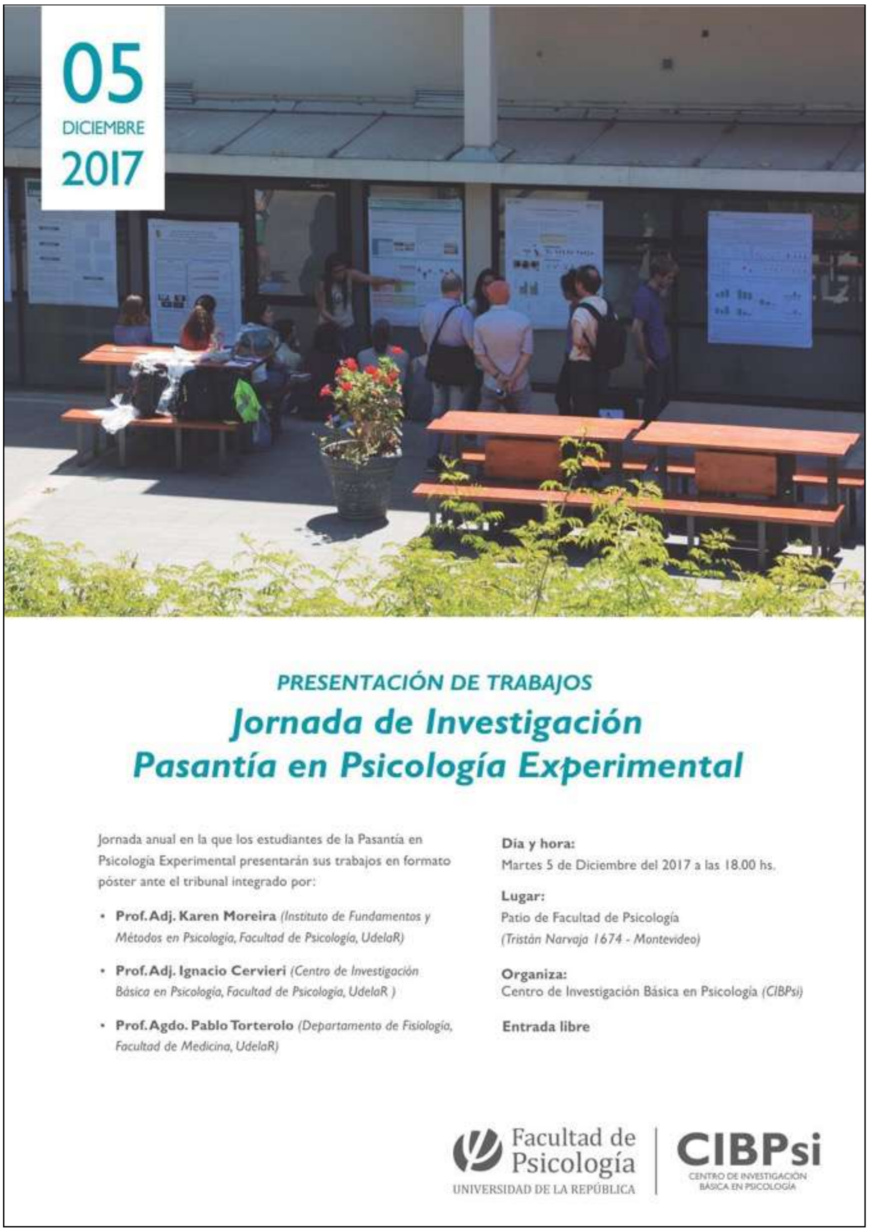
Bulletin and Review, *22*(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2
Schmitterer, A. M. A., & Schroeder, S. (2019). Effects of reading and spelling predictors before and after school entry: Evidence from a German longitudinal study. *Learning and Instruction*, *59*(2019), 46–53. https://doi.org/10.1016/j.learninstruc.2018.09.005
Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England: 1953)*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859
Share, D. L. (2008). On the Anglocentricities of Current Reading Research and Practice: The Perils of Overreliance on an "Outlier" Orthography, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584
Solheim, O. J., Torppa, M., & Uppstad, P. H. (2020). Screening for Slow Reading Acquisition in Norway and Finland – a Quest for Context Specific Predictors Screening for Slow Reading Acquisition in Norway and. *Scandinavian Journal of Educational Research*, *0*(0), 1–17. https://doi.org/10.1080/00313831.2020.1739130
Stekhoven, D. J., & Buehlmann, P. (2012). MissForest - non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.
Tang, F., & Ishwaran, H. (2017). Random Forest Missing Data Algorithms. *Statistical analysis and data mining*, *10*(6), 363–377. https://doi.org/10.1002/sam.11348
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5). https://doi.org/10.1177/002221949402700503
Treiman, R., & Kessler, B. (2004). The Role of Letter Names in the Acquisition of Literacy. *Advances in Child Development and Behavior*, *31*(C), 105–135. https://doi.org/10.1016/S0065-2407(03)31003-1
Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive Development of Fluent Word Reading Does Not Qualitatively Differ Between Transparent and Opaque Orthographies. *Journal of Educational Psychology*, *102*(4), 827–842. https://doi.org/10.1037/a0019465
Van Bergen, E., De Jong, P. F., Regtvoort, A., Oort, F., Van Otterloo, S., & Van Der Leij, A. (2011). Dutch children at family risk of dyslexia: Precursors, reading development, and parental effects. *Dyslexia*, *17*(1), 2–18. https://doi.org/10.1002/dys.423
Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40.
Verhoeven, L., & Keuning, J. (2017). The Nature of Developmental Dyslexia in a Transparent Orthography. *Scientific Studies of Reading*, 1–17. https://doi.org/10.1080/10888438.2017.1317780
Vloedgraven, J. M. T., & Verhoeven, L. (2009). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences*, *19*(2), 161–169. https://doi.org/10.1016/j.lindif.2008.09.005
Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological Bulletin, 101(2), 192-212. https://doi.org/10.1037/0033-2909.101.2.192
Wechsler, D. (2001). *WPPSI-III: Escala de Inteligencia de Wechsler para Preescolar y Primaria*. Madrid, Spain: TEA.
Wood, C., & Connelly, V. (2009). *Phonological awareness beyond phonemes*. *Contemporary Perspectives on Reading and Spelling*. https://doi.org/10.4324/9780203877838
Yopp, H. K. (1988). The Validity and Reliability of Phonemic Awareness Tests. *Reading Research Quarterly*, *23*(2), 159. https://doi.org/10.2307/747800
Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3
Zugarramurdi, C., Fernández, L., Lallier, M., Carreiras, M., & Valle-Lisboa, J. C. (in press). Lexiland: a Tablet-based Universal Screener for Reading Difficulties in the School Context. *Journal of Educational Computing Research.*

Testimonios e imágenes
Testimonials and images

De izquierda a derecha y de arriba a abajo, poster de la charla y visita del Dr. Rafael Núñez, la defensa de la tesis de Maestría en Ciencias Cognitivas de Lucía Fernández y de la disertación doctoral de Camila Zugarramurdi . En el extremo inferior, de izquierda a derecha, detalle de la exposición de póster en las Jornadas de la Pasantía de Psicología Experimental, estudiantes acompañados de la Dra. Jill Morford —quien fuese profesor visitante en 2019— y detalle de la defensa de grado de Germán Cipriani.

From left to right and from top to bottom, poster of the talk and visit by Dr. Rafael Núñez, the defense of the Master's thesis in Cognitive Sciences of Lucía Fernández, and doctoral dissertation of Camila Zugarramurdi. At the bottom, from left to right, detail of the poster exhibition at the Experimental Psychology Internship, students accompanied by Dr. Jill Morford —who was a visiting professor in 2019— and detail of Germán Cipriani's undergraduate defense.

Pasantía en Psicología experimental
Experimental Psychology Internship



La actividad investigadora del **CIBPsi** ha impactado positivamente a la enseñanza de grado y posgrado en la Facultad de la Psicología y la Universidad de la República en su conjunto. Los cursos optativos de grado, prácticas y proyectos desplegados por el centro no tienen una cobertura amplia para una facultad con alta numerosidad. Sin embargo, sí han representado una aportación cualitativa y constante importante al abrir paso a tradiciones teóricas o temáticas -algunas incluso primigenias al campo psicológico como la Psicología del tiempo- que estaban ausentes o mínimamente representadas en el currículo y el cuerpo docente, impulsar técnicas de estudio y formas de evaluación que descansan la didáctica y el aprendizaje en la cultura de investigación. En la imagen superior, el afiche de la Jornada de la **Pasantía en Psicología Experimental**. Esta práctica se ha consolidado como una valiosa oportunidad para la/el estudiante de grado de la UdelaR para tener una primera experiencia de investigación que le aporta créditos a su avance, le permite interactuar con equipos de investigación y le involucra de pleno en la actividad investigadora de inicio a fin.

CIBPsi's research activity has had a positive impact on undergraduate and graduate teaching at the School of Psychology and the University of the Republic as a whole. The undergraduate elective courses, internships and projects deployed by the center do not have a wide coverage for a faculty with a high number of students. However, they have represented an important and robust qualitative contribution by opening the way to theoretical or thematic traditions -some even primordial to the psychological field, such as the Psychology of time- that were absent or minimally represented in the curriculum and the faculty, promoting study techniques and forms of evaluation that rest didactics and learning in the research culture. In the image above, the day's poster of the **Pasantía en Psicología Experimental.** This practice has been consolidated as a valuable opportunity for the UdelaR undergraduate student to have a first research experience that contributes credits to his/her progress, allows him/her to interact with research teams and fully involves the student in all the steps of research.

Llegué al Cibpsi en 2018 casi que por equivocación. Aprovechando la flexibilidad del plan de estudios en Ciencias Biológicas, durante mi segundo semestre de facultad, cursé una gran variedad de UC (unidades curriculares/materias) con diversos acercamientos a la Biología. En una de ellas tuvimos la participación de un docente que nos contó sobre ciertos seminarios dictados por el Centro de Investigación Básica en Psicología. Meses después decidí buscar información sobre estos seminarios. Terminé leyendo sobre todas las líneas de investigación del centro. Entonces, escribí un correo al director de la línea que más me interesó y le consulté si él o su equipo dictaría algún tipo de curso corto que una estudiante de segundo año en Ciencias Biológicas pudiera tomar. Me respondió que no tenía noción respecto a qué seminarios me refería, pero me invitaba a hablar y conocer mis intereses en una reunión presencial. Casi lo primero que me dijo al vernos fue que no se contaba con dinero suficiente como para crear un cargo por proyecto o de contrato, pero que me invitaba a participar de manera honoraria en su equipo. Más que superadas mis expectativas, me sumé a recabar datos para un experimento en curso y comencé a participar de las reuniones de equipo. Con el pasar del tiempo decidimos crear un seminario semanal para presentar lecturas y trabajos en curso, todas las y los integrantes por igual. Más adelante desarrollé un proyecto de investigación propio, a través del programa PAIE. Actualmente, aquél director de línea, Roberto Aguirre, se volvió mi tutor de carrera y de tesis. Hoy puedo decir que el vínculo humano con el resto del equipo es de las mejores cosas que me llevo del CIBPsi.

I came to CIBPsi in 2018, almost by mistake. Taking advantage of the flexibility of the curriculum in Biological Sciences, during my second semester of college, I took a variety of UC (curricular units/subjects) with different approaches to Biology. In one of them, we had the participation of a professor who told us about certain seminars given by the Centro de Investigación Básica en Psicología. Months later, I decided to look for information about these seminars. I ended up reading about all the research lines of the center. Then, I wrote an email to the director of the line that most interested me. I asked him if he or his team would give some short course that a undergraduate student in Biological Sciences could take. He replied that he had no idea which seminars I was referring to, but invited me to discuss my interests in a face-to-face meeting. Almost the first thing he told me that there was not enough money to create a project or contract position, but he invited me to participate as an honorary member of his team. More than meeting my expectations, I joined in collecting data for an ongoing experiment and began participating in team meetings. As time goes by, we decided to do a weekly seminar to present readings and work in progress, all members equally. Later on, I developed my own research project through the PAIE program. Currently, that line director, Roberto Aguirre, has become my career and thesis tutor. Today I can say that the human bond with the rest of the team is one of the best things I have taken away from CIBPsi.

Agustina Echaider Lombardi. Estudiante
Biología Humana. TFG: Interacción educativa
en el aprendizaje lector de sordos señantes. Integro
la línea de investigación Mente, Acción y Lenguaje

Aspecto de la Jornada de pósters de la Práctica de Psicología experimental del centro
Aspect of the Poster session of the Pasantía en Psicología experimental

En la imagen, un detalle de las jornadas de presentación de póster de la Pasantía en Psicología experimental, celebrada en noviembre-diciembre de cada año, a fin de mostrar los resultados de un año de trabajo de los estudiantes a una terna de evaluadores externos.

At the picture, a detail of the Jornada de posters de la Práctica de Psicología experimental. This event is celebrated between November-December of each year. In these opportunity, the undergraduate students show the results of a one year research project to a panel of three external researchers and the general audience.

Corría el año 2017 y yo estaba comenzando el último año de la "Licenciatura en Ciencias Biológicas" de la Facultad de Ciencias. Hacía ya unos meses que me había propuesto acercarme a un laboratorio para poder comenzar a planear mi trabajo final de grado. Un día, no recuerdo bien en qué evento, vi un póster que presentaba los resultados de una investigación de neurociencias realizada en humanos, cuyos autores pertenecían al "Centro de Investigación Básica en Psicología". Como me interesaban las neurociencias, y en particular quería estudiar el funcionamiento del cerebro humano, decidí buscar a ese centro en internet, lo que me llevó a encontrar una página web donde se describían las diferentes líneas de trabajo del Centro. Una de ellas capturó poderosamente mi atención: la línea de "Neurociencia Cognitiva y Salud Mental", dirigida por la Dra. Victoria Gradín. Decidí escibirle a Victoria para que me contara un poco más acerca del trabajo que estaban haciendo.

Hoy en día pasaron seis años de ese momento. Durante ese tiempo participé en un proyecto PAIE, realicé mi trabajo final de grado y mi tesis de maestría bajo la tutoría de Victoria. Además, junto al resto del equipo de la línea, publicamos dos artículos científicos y realizamos varias presentaciones en congresos. El año pasado tuve la oportunidad de aportar desde un nuevo rol: el de orientador, tutoreando un proyecto PAIE y trabajos finales de grado.

Creo que todo este proceso ha sido el pilar de mi formación, todavía en curso, como investigador. Ocurrió en medio de un ambiente de contención y apoyo mutuo, tanto con el resto de los integrantes del equipo, como del Centro en general.

No me queda más que agradecer al Centro por todas las oportunidades brindadas y por su cálido ambiente de trabajo. ¡Espero poder seguir trabajando con ellos durante muchos años más!

It was 2017, I was starting the last year of the "Licenciatura en Ciencias Biológicas" of the Faculty of Science. It had been a few months since I had proposed to get close to a laboratory to I could start planning my final degree work. One day, I don't quite remember at which event, I saw a poster presenting the results of neuroscience research conducted on humans, whose authors belonged to the "Centro de Investigación Básica en Psicología". As I was interested in neurosciences, and in particular, I wanted to study the functioning of the human brain, so I decided to look for this center on the Internet, which led me to find a web page describing the different lines of work at the center. One of them caught my attention: the line "Neurociencia Cognitiva y Salud Mental", directed by Dr. Victoria Gradín. I decided to ask Victoria to tell me a little more about the work they were doing.

Today six years have passed since then. During that time, I participated in a PAIE project and did my final undergraduate work and my master's thesis under Victoria's direction. In addition, together with the members of the team, we published two scientific articles and made several presentations at conferences. Last year, I had the opportunity to contribute in a new role: as an advisor, tutoring a PAIE project and final projects.

I believe that this whole process has been the keystone of my training, still in progress, as a researcher. It took place in an atmosphere of mutual support and contention, both with the rest of the members of the team, and the center in general. I can only thank the center for all the opportunities provided and its warm working environment, and I hope to continue working with them for many more years!

Alejo Acuña. Licenciado en Ciencias
Biológicas magister en Ciencias Cognitivas.
Asistente G1, Laboratorio de
Neurociencias y CIBPsi

Primeros estudios cognitivos de resonancia magnética funcional en Uruguay
First cognitive studies with magnetic functional resonance in Uruguay

Puesta a punto de primer estudio de resonancia magnética funcional realizado en Uruguay. El mismo fue llevado a cabo por investigadores del **CIBPsi**, en conjunto con el Centro Uruguayo de Imagenología Molecular.

Setting up and preparing the instruments for the first functional magnetic resonance imaging study performed in Uruguay. The study was carried out by researchers from **CIBPsi**, together with the Uruguayan Center of Molecular Imaging.

En 2010 apliqué junto a unos colegas al llamado a proyectos del CIBPSI con una propuesta de investigación centrada en el desarrollo de hardware y software para niños con parálisis cerebral. En aquel entonces poco se hablaba de las ciencias cognitivas en Facultad. El centro se acaba de constituir y su director era el profesor A. Maiche. Cinco proyectos fueron seleccionados para poder llevarse a cabo contando con la infraestructura y equipos del centro. Gracias a esa vinculación y al apoyo material, aquellas investigaciones dieron sus frutos y los que éramos más jóvenes hicimos una carrera académica amparados por el CIBPSI. Unos años más tarde, el centro había crecido tanto que se artículo en grandes líneas de investigación, atraía a investigadores del exterior, ofertaba su propia maestría, organizaba conferencias y escuelas de verano de gran afluencia, y además había establecido colaboraciones con laboratorios prestigiosos del exterior. Este crecimiento exponencial y los logros científicos de las personas que algún día pasaron por allí, dan cuenta de la calidad del ambiente científico que se sigue cultivando hasta el día de hoy. El CIBPSI ha sido para todos nosotros una gran oportunidad formativa.

In 2010 I applied with some colleagues to the CIBPSI call for projects with a research proposal focused on developing hardware and software for children with cerebral palsy. At that time there was little talk of cognitive sciences in the Faculty. The center had just been established and its director was Professor A. Maiche. Five projects were selected to be carried out using the center's infrastructure and equipment. Thanks to this linkage and material support, that research bore fruit and those of us who were younger made an academic career under the protection of CIBPSI. A few years later, the center had grown so much that it was involved in major lines of research, attracted researchers from abroad, offered its own master's degree, organized conferences and summer schools of great affluence, and had also established collaborations with prestigious laboratories abroad. This exponential growth and the scientific achievements of the people who went by there one day, attest to the quality of the scientific environment that continues to be cultivated to this day. CIBPSI has been a great formative opportunity for all of us.

Ana Martín. Doctorado en Tecnologías de la Información y la Comunicación. Investigador postdoctoral en el Departamento de Estudios Cognitivos, École Normale Supérieure, Paris

Participación de CIBPsi en congresos internacionales y locales
Participation of CIBPsi in international and local congresses

En la imagen, a la izquierda, Stanislas Dehaene, del Collège de France de París, y a la derecha, Alejandro Maiche, en una conversación en los descansos del I Congreso Uruguayo de Ciencias Cognitivas & II Simposio de Educación, Cognición y Neurociencia de la Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento, celebrado de17 al 21 de Noviembre de 2021. Este asociación, creada en 2019, cuenta con na amplia participación de miembros del CIBPsi.

In the picture, on the left, Stanislas Dehaene, from the Collège de France in Paris, and on the right, Alejandro Maiche, in a dialogue during the breaks of the II Congreso Uruguayo de Ciencias Cognitivas & II Simposio de Educación, Cognición y Neurociencia of the Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento, held from November 17th to 21st, 2021. This association, created in 2019, has a wide participation of CIBPsi members.

Mi nombre es Germán Cipriani y soy licenciado en Psicología y magíster en Ciencias Biológicas (Neurociencias) por la Universidad de la República (Uruguay). Actualmente, me encuentro realizando mis estudios del Doctorado en Psicología en la Universidad de Granada (España) bajo la codirección de los doctores Fabiano Botta y Juan Lupiáñez, donde formo parte de la Línea de Atención del Centro de Investigación, Mente, Cerebro y Comportamiento (CIMCYC). Allí, mi investigación se centra en la caracterización a nivel comportamental y cerebral de los mecanismos atencionales que operan sobre los contenidos de la Memoria Operativa (Working Memory). Para ello utilizo tareas experimentales y analizo medidas conductuales y neurofisiológicas.

Durante el grado y la maestría formé parte de la Línea de Investigación en Atención del CIBPsi, donde he podido aprender, participar y trabajar en diversas investigaciones bajo la tutela de las doctoras Alejandra Carboni y Dominique Kessel. A su vez, he tenido la oportunidad de asistir a cursos del centro y participar de congresos nacionales e internacionales, como también he impartido docencia en diversas asignaturas de grado. En el centro he encontrado el espacio y las oportunidades para desarrollarme en mis habilidades docentes e investigadoras, y un lugar donde pensar de forma crítica y científica la producción de conocimiento universitaria. El trabajo con mis compañeres, dentro y fuera de mi línea, me ha permitido pensar desde una perspectiva más interdisciplinaria, inclusiva y humana, a la Universidad en sí misma.

Agradezco la oportunidad que ha sido central, de formarme en un centro del cual adquirí mucho conocimiento y capacidades, y del cual fundamentalmente me llevo grandes amigues.

¡Felicidades a la facultad y al centro por sus primeros 12 años!

My name is Germán Cipriani and I have a degree in Psychology and a Master in Biological Sciences (Neurosciences) from the University of the Republic (Uruguay). Currently, I am doing my PhD studies in Psychology at the University of Granada (Spain) under the co-direction of Dr. Fabiano Botta and Dr. Juan Lupiáñez, where I am part of the Attention Line of the Center for Research, Mind, Brain and Behavior (CIMCYC). There, my research focuses on the behavioral and cerebral characterization of the attentional mechanisms that operate on the contents of Working Memory. For this purpose, I use experimental tasks and analyze behavioral and neurophysiological measures.

During my undergraduate and master's degree I was part of the Attention Research Line at CIBPsi, where I have been able to learn, participate and work in several research projects under the tutelage of Dr. Alejandra Carboni and Dr. Dominique Kessel. At the same time, I have had the opportunity to attend courses at the center and to participate in national and international congresses, as well as to teach various undergraduate courses. At the center I have found the space and opportunities to develop my teaching and research skills, and a place to think critically and scientifically about the production of university knowledge. Working with my colleagues, inside and outside my line, has allowed me to think from a more interdisciplinary, inclusive and human perspective, to the University itself.

I am grateful for the opportunity, which has been central, to be formed in a center from which I acquired a lot of knowledge and skills, and from which I fundamentally take with me great friends.

Congratulations to the faculty and the center for its first 12 years!

Germán Cipriani. Licenciado en
Psicología, magister en Neurociencias
y doctorando en Psicología

Detalle de entrevista televisiva al Dr. Álvaro Cabana por el proyecto Lexicón
Capture of television interview with Álvaro Cabana PhD. for the Lexicon project

En la imagen el Dr. Álvaro Cabana en una entrevista televisiva del programa “La mañana en casa”, en el canal 10 de la televisión uruguaya el 17 de octubre de 2018. La entrevista versó sobre el proyecto Lexicón al cargo del investigador. La entrevista se puede consultar en https://youtu.be/Z3WnsxR3vDw?t=500

At the picture, Alvaro Cabana PhD., member of CIBPsi, in a television interview on the program “La mañana en casa" on Uruguayan television channel 10 on October 17th, 2018. The interview was about the Lexicon project in charge of him. The interview is available at https://youtu.be/Z3WnsxR3vDw?t=500

Mi primer contacto con el CIBPsi se remonta al año 2011. Recuerdo muy bien, que todavía transitando su período fundacional, el CIBPsi ofertaba un curso corto titulado "Cerebro y Conducta: Introducción a la Neurociencia Cognitiva", en el cual una investigadora española se proponía ahondar en temas tales como la conciencia o el libre albedrío. Las fascinantes discusiones sobre la naturaleza de la mente aterrizaban por fin, al menos para mí, en la Udelar. Es que por aquel entonces lo cognitivo no tenía un lugar claro en las currículas. A partir de ese primer acercamiento comencé a recibir el boletín de noticias "cibpsiano" y a participar ocasionalmente de actividades. Fue en el año 2015 que me integré al -todavía incipiente- grupo de Desarrollo cognitivo infantil dirigido por Alejandra Carboni, en el cual desde 2017 se inscribe mi proyecto doctoral.

Sobre la base de mucho esfuerzo y sacrificio, el CIBPsi se ha constituido como un espacio peculiar dentro de la Udelar por su interdisciplinariedad, por demás estimulante para investigadores e investigadoras en formación en el vasto, dinámico y exigente campo de la Ciencia Cognitiva. En estos 10 años de impulsos y frenos, el centro ha logrado generar las condiciones para fortalecer trayectorias en múltiples esferas ligadas al quehacer científico: diseño y redacción de proyectos, evaluación psicológica y psicofisiológica, programación, análisis estadístico, comunicación científica, entre otras. En lo personal, además, el desarrollo de mi proyecto doctoral en el CIBPsi posibilitó que encontrase áreas de estudio que hoy me apasionan, permitiendo proyectarme como investigador. Pero más allá de sus virtudes en el terreno académico, me gustaría también rescatar el sentido de comunidad basado en el compañerismo, el afecto y la solidaridad que se ha cultivado en el centro en estos primeros 10 años. No digo nada nuevo al afirmar que el camino hacia la profesionalización científico-académica en Uruguay es espinoso y engorroso. Por eso valoro esa suerte de propiedad emergente que hace a la cotidianidad "cibpsiana": la habilidad para procesar colectivamente las frustraciones. Considero que ésta ha abonado el desarrollo individual de la tenacidad de quienes han participado en este período, desde distintos lugares y en diversas formas, en la construcción de este espacio y su acervo.

Me siento agradecido por haber encontrado en la Facultad de Psicología un sitio que habilita la emergencia de algo nuevo dentro de la Udelar, así como una comunidad comprometida con diversas problemáticas de interés social. El CIBPsi conmemora sus primeros 12 años, y es motivo, ante todo, de festejo y orgullo.

My first contact with CIBPsi dates back to 2011. I remember very well that, still in its foundational period, the CIBPsi offered a short course entitled "Cerebro y Conducta: Introducción a la Neurociencia Cognitiva", in which a Spanish researcher proposed to delve into topics such as consciousness or free will. The fascinating discussions on the nature of the mind finally landed, at least for me, at Udelar. At that time, cognitive subjects did not have a clear place in the curricula. From that first approach I began to receive the "cibpsiano" newsletter and to participate occasionally in activities. It was in 2015 that I joined the -still incipient- group of Desarrollo cognitivo infantilt headed by Alejandra Carboni, in which since 2017 my doctoral project is inscribed.

Based on much effort and sacrifice, the CIBPsi has become a peculiar space within the Udelar for its interdisciplinarity, stimulating for researchers and researchers in training in the vast, dynamic and demanding field of Cognitive Science. In these 10 years of impulses and brakes, the center has managed to generate the conditions to strengthen trajectories in multiple spheres linked to scientific work: project design and writing, psychological and psychophysiological evaluation, programming, statistical analysis, scientific communication, among others. Personally, moreover, the development of my doctoral project at CIBPsi did it possible for me to find areas of study that I am passionate about today, allowing me to project myself as a researcher. But beyond its virtues in the academic field, I would also like to highlight the sense of community based on companionship, affection and solidarity that has been cultivated in the center during these first 10 years. I am not saying anything new when I say that the road to scientific-academic professionalization in Uruguay is thorny and cumbersome. That is why I value that sort of emergent property that makes the "cibpsian" cotidinanity: the ability to collectively process frustrations. I believe that this has contributed to the individual development of the tenacity of those who have participated in this period, from different places and in different ways, in the construction of this space and its heritage.

I am grateful for having found in the Facultad de Psicología a space that enables the emergence of something new within Udelar, as well as a community committed to various issues of social interest. CIBPsi is celebrating its early 12 years, and this is, above all, a reason for celebration and pride.

Hernán Delgado Vivas. Licenciado en Ciencias Biológicas, magister en Neurociencias y doctorando en Ciencias Biológicas.
Asistente G2, Instituto de Fundamentos y Métodos en Psicología y CIBPsi

Participación en la Semana del Conocimiento del Cerebro
Participating at the Semana del Conocimiento del Cerebro

En la foto, investigadores jóvenes, algunos del **CIBPsi,** durante su participación en la Feria de stands y juegos para niños de la Semana del Conocimiento del Cerebro en su edición de 2015 (16 al 21 de marzo, Centro Cultural Goes, Montevideo, Uruguay). Este evento, que se realiza cada año, convoca investigadores de distintas disciplinas para difundir en la población uruguaya el conocimiento sobre la actividad cognitiva y su sustrato neural.

In the photo, young researchers, some of them from **CIBPsi**, during their appearance at the stands and games fair for children during the 2015 edition of the Brain Knowledge Week (March 16th-21st, Goes Cultural Center, Montevideo, Uruguay). This event, held annually, brings together researchers from different disciplines to disseminate knowledge about cognitive activity and its neural substrate among the Uruguayan population.

Mi primer contacto con el CIBPsi fue en 2013, fui invitada para llevar a mi hijo (que en ese entonces tenía 8 años) a participar en una investigación que se desarrollaba en el hospital Policial, en donde participaban varios investigadores del centro. Como estudiante de grado de la licenciatura en Psicología, esto despertó cierta curiosidad en mi lo que generó que me involucrará de forma gradual en cursos y actividades del centro. En 2015, formalicé mi relación con la investigación y las Ciencias Cognitivas, a través del curso Pasantía en Psicología Experimental, que se brinda desde el centro, y ya nunca me fui. Luego de esa experiencia participé de un proyecto PAIE y de algunos proyectos ya en curso en la línea de investigación Estudios experimentales de la metáfora cognitiva y LSU. También, mi trabajo de grado estuvo enmarcado en los temas de ésta línea. Y actualmente estoy finalizando la Maestría en Ciencias Cognitivas y continúo colaborando en los proyectos de la línea.

El pasaje por el Centro ha sido el pilar de mi formación como investigadora, además de la contención del equipo de investigación, en cuanto a convocatorias, el cómo-hacer de las distintas actividades de investigación, como las charlas y seminarios regulares que se ofrecen en el Centro. Ellas han aportado a mi formación teórica y metodológica. Agradezco sobre todo el apoyo del Dr. Roberto Aguirre, que ha sido mi tutor de grado y que hoy es mi tutor de posgrado por su inversión de energía en mi formación, así como por la buena onda y contención siempre. ¡Espero logremos muchos avances más en este equipo!

My first contact with CIBPsi was in 2013, I was invited to take my son (who was 8 years old at that time) to participate in a research that was being developed at the Police Hospital, where several researchers from the center were participating. As an undergraduate student of the degree in Psychology, this aroused some curiosity in me which generated that I gradually became involved in courses and activities of the center. In 2015, I formalized my relationship with research and Cognitive Sciences, through the Internship in Experimental Psychology course, and I never left. After that experience I participated in a PAIE project and in some ongoing projects in the research line Experimental Studies of Cognitive Metaphor and LSU. Also, my graduate work was framed in the topics of this line. I am currently finishing my Master's degree in Cognitive Sciences and I continue to collaborate in the projects of this line.

The passage through the Center has been the pillar of my training as a researcher, in addition to the support of the research team, in terms of calls, the how-to of the various research activities, such as regular lectures and seminars offered at the Center. They have contributed to my theoretical and methodological training. I am especially grateful for the support of Dr. Roberto Aguirre, who has been my undergraduate tutor and who is now my graduate tutor, for his investment of energy in my training, as well as for his constant good vibes and support. I hope we will achieve many more advances in this team!

Maria Noel Macedo. Licenciada en Psicología,
maestranda en Ciencias Cognitivas.
Ayudante G1, CIBPsi. Asistente G2, CICEA

Participación de colegas del CIBPsi en otros colectivos y estructuras académicas
Participation of CIBPsi's members in other academic groups and structures

Primer cuadro del video que recoge la actividad del Simposio Regional Aportes de las Ciencias Cognitivas a la Educación, celebrado el 17 de octubre de 2011. El evento contó con una importante del Núcleo Interdisciplinario en Ciencias Cognitivas (NCC). El NCC, antecedente del Centro de Investigación en Cognición para la Enseñanza y el Aprendizaje (CICEA). Al igual que este último, el NCC fue un grupo que contó con una presencia relevante de investigadores del **CIBPsi** que vieron en el aprendizaje una ámbito importante de desarrollo y aplicación de los hallazgos de las Ciencias Cognitivas.

Frame 1 of the video that shows the activity of the Simposio Regional Aportes de las Ciencias Cognitivas a la Educación, held on October 17, 2011. The event counted with important participation of the Núcleo Interdisciplinario en Ciencias Cognitiva (NCC). The NCC, the predecessor of the Centro de Investigación en Cognición para la Enseñanza y el Aprendizaje (CICEA). ike the latter, the NCC was a group that had a relevant presence of **CIBPsi** researchers who saw in learning an important area of development and application of the findings of the Cognitive Sciences.

Conocí al CIBPsi en el año 2013 a través del curso de Procesos cognitivos I, luego del interés generado por la temática, decidí ser voluntario como sujeto experimental al año siguiente. A partir de esa experiencia, coordiné una reunión con Alejandra Carboni y acto siguiente me recibió en la línea de investigación en Atención como estudiante de grado.

En dicha línea tuve mis primeras experiencias como investigador. Pude participar en diversos proyectos propuestos por el grupo, fui orientado en la realización de un proyecto PAIE y fui colaborador en la organización de la Semana del conocimiento del cerebro, es decir compartí experiencias de trabajo con investigadores/as de otras facultades, fortaleciendo así mis vínculos académicos.

Posteriormente, comencé a trabajar junto con Roberto Aguirre en la línea de Lenguaje en temas relativos a los estudios experimentales con señantes. Allí me apoyaron a solidificar aún más mi formación como investigador. Continué siendo colaborador y autor de investigaciones, fui impulsado a realizar una pasantia en la Universidad de Nuevo México, a presentar trabajos en congresos, entre otras actividades.

En todo ese trayecto de aprendizaje también tuve acceso a cursos a fines a las ciencias cognitivas, seminarios de investigación, recibí asistencia en mis investigaciones por parte de compañeros e investigadores expertos, me prestaron y enseñaron a utilizar herramientas de registros experimentales como el Electroencefalograma, Eyetracker, entre otras herramientas prestadas.

Todas las experiencias detalladas en este relato me han aportado significativamente en mi formación como investigador tanto en lo técnico como en los vínculos establecidos con otros actores pertenecientes al ámbito académico. Gran parte de mi formación se ha dado en el CIBPsi, gracias al trabajo, el compañerismo, las enseñanzas de sus integrantes y el espacio prestado, hoy cuento con conocimientos y experiencias que no podría conseguir en otro lugar.

Mauricio Castillo Fernandez. Licenciado en Psicología
y maestrando en Ciencias Cognitivas.
Investigador de posgrado en el CIBPsi.
Asistente académico (G2) en la Escuela
de Posgrado de Facultad de Derecho (Udelar)

Detalle de entrevista televisiva a la Dra. Victoria Gradín y el magister Alejo Acuña acerca de su investigación en materia de los componentes neurales asociados a depresión y ansiedad
Detail of television interview with Victoria Gradín PhD. and MagísterAlejo Acuña about their research on the neural components associated with depression and anxiety.

En la imagen el Mag. Aleo Acuña en la entrevista televisiva del programa “Sobre Ciencia”, en el canal TVCiudad de la televisión uruguaya el 3 de septiembre de 2022. La entrevista con él y la Dra. Victoria Gradín versó sobre sus estudios relativos a los componentes neurales cerebrales de la depresión y la ansiedad. La entrevista se puede consultar en https://www.youtube.com/watch?v=0oBuwniCdko

In the image, Mag. Aleo Acuña in the television interview of the program “Sobre Ciencia”, in the TVCiudad channel of the Uruguayan television on September 3rd, 2022. The interview with him and Victoria Gradín PhD. was about their studies on the neural brain components of depression and anxiety. The interview is available at https://www.youtube.com/watch?v=0oBuwniCdko

En la actualidad soy psicólogo clínico, me especializo en trastornos mentales severos y persistentes y en duelo. Ejerzo la docencia en la Cátedra de Psicopatología Fundamental de la Licenciatura en Psicología de la UCU, además participé en dos oportunidades como docente interino de la UCO Metodología General de la Investigación en la Licenciatura en Psicología de la Udelar. Atiendo pacientes en mutualistas, en un centro de rehabilitación y en el ámbito privado. El doctorado que estoy comenzando pretende asociar efectos del cannabis en el pensamiento de fumadores, comparando grupos no clínicos de personas que consumen la droga frente a quienes no. Concreté la licenciatura luego de varios años habiendo abandonado los estudios formales y habiendo estado dedicado como empleado en trabajos de los más variados hasta pasados mis 20 y largos. Soy el primer universitario de mi familia y fue todo un desafío reinsertarme en la vida académica y el comenzar una carrera universitaria.

Mi pasaje por el CIBPsi, en el marco de la pasantía y colaborando posteriormente en algún proyecto relacionado a lenguaje con Roberto Aguirre. Este pasaje me permitió conocer y aprehender sobre metodología y técnica, sobre la labor diaria persistente y continua que requiere e implica el trabajo científico de investigación. Conocí grupos de diferentes líneas, pude ver diferentes formas de trabajo en equipo, y participé como sujeto experimental de otros proyectos con el fin de aprehender y conocer otras temáticas. Por otro lado, me dejó grandes aportes en cuanto a rigurosidad y exigencia al momento de informarme sobre situaciones clínicas, teorías y técnicas, con una mirada exigente y crítica ante nueva información, divulgación e intercambio con colegas. Por otro lado, me permitió conocer mi curiosidad y gusto por la investigación, al punto que creo que fue una etapa de mi vida personal y académica que hoy tiene mucha influencia en continuar la formación de posgrado con el fin de generar conocimiento nuevo que aporte a la sociedad, además de satisfacer mi gusto y curiosidad personal. El pasaje por el CIBPsi fue muy importante para la vida de este primer universitario de mi familia.

I am currently a clinical psychologist, specializing in severe and persistent mental disorders and bereavement. I teach in the Cátedra de Psicopatología Fundamental at the Licenciatura en Psicología at UCU, and I have also participated in two opportunities as an interim professor of the UCO Metodología General de la Investigación at the Udelar. I attend patients in mutual insurance companies, in a rehabilitation center and in private practice. The doctorate I am starting aims to associate the effects of cannabis on the thinking of smokers, comparing non-clinical groups of people who use the drug versus those who do not. I completed the degree after several years having abandoned formal studies and having been employed in a wide variety of jobs until I was in my late 20's and early 20's. I am the first university graduate in my family. I am the first university student in my family and it was quite a challenge to reinsert myself into academic life and to start a university career.

My time at CIBPsi, in the framework of the internship and later collaborating in a project related to language with Roberto Aguirre. The time at CIBPsi allowed me to know and learn about methodology and technique, about the persistent and continuous daily work that requires and implies the scientific work of research. I met groups of different lines, I could see different forms of teamwork, and I participated as an experimental subject of other projects in order to learn and learn about other topics. On the other hand, it gave me great contributions in terms of rigor and exigency when informing myself about clinical situations, theories and techniques, with a demanding and critical look at new information, dissemination and exchange with colleagues. On the other hand, it allowed me to learn about my curiosity and taste for research, to the point that I believe it was a stage in my personal and academic life that today has a great influence in continuing postgraduate training in order to generate new knowledge that contributes to society, in addition to satisfying my personal taste and curiosity. The passage through the CIBPsi was very important for the life of this first university student in my family.

Richard Rodriguez. Licenciado en
Psicología, magister en Psicoterapia
y doctorando en Psicología

Participación de CIBPsi en congresos internacionales y locales
Participation of CIBPsi in international and local congresses

En la foto, miembros de la línea de Cognición Numérica en la sesión de pósters del I Congreso Uruguayo de Ciencias Cognitivas & II Simposio de Educación, Cognición y Neurociencia de la Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento, celebrado de17 al 21 de Noviembre de 2021. Este asociación, creada en 2019, cuenta con na amplia participación de miembros del CIBPsi.

In the picture, members of the Numerical Congition research line at the poster’s session of the II Congreso Uruguayo de Ciencias Cognitivas & II Simposio de Educación, Cognición y Neurociencia of the Sociedad Uruguaya de Ciencias Cognitivas y del Comportamiento, held from November 17th to 21st, 2021. This association, created in 2019, has a wide participation of CIBPsi members.

En septiembre de 2012 me encontraba a la espera de la fecha para mi defensa de tesis doctoral en la Universitat Autònoma de Barcelona. Como cualquier nuevo doctor, buscaba dónde continuar y proyectar mi actividad académica. Ana Pires, quien fue mi compañera de doctorado, me contó de las becas del fondo Roberto Caldeyro y Barcia, en Uruguay. Esta opción era una alternativa que se distinguía por ser un espacio nuevo con caras conocidas (Alejandro Maiche y Fernando González). Me postulé y obtuve la beca, al igual que Victoria Gradin. En todo el proceso de postulación tuve el apoyo de Juan Valle-Lisboa y Camila Zugarramurdi. Al pasar dos años de la beca tenía un cargo interino en Salto y uno efectivo en Montevideo. Debía decidir si volver a mi tierra natal -México- o quedarme en Montevideo. Opté por quedarme, por razones académicas y personales. Desde 2014 he venido impulsando una línea de investigación sobre la producción de significado. En ese esfuerzo he tenido el gusto del acompañamiento de estudiantes y colegas nacionales e internacionales. Maria Noel, Mauricio, Agustina, Martín, Carlos, Mauro son, en los estudiantes, quienes más claramente han apostado por el proyecto propuesto. Julio Santiago, Jorge Vivas, Alejandro Fojo, Yliana Rodríguez, Manuel Garcia-Ruiz y Jill Morford han sido colegas que, con distintos alcances, han puesto el hombro a la propuesta. Como cualquier organización con personas, el CIBPsi ha tenido buenos y malos momentos y hábitos, pero cuenta con virtudes esenciales para investigar con pasión, compañerismo (p.e., Álvaro Cabana siempre nos saca adelante cuando nos atoramos con el R Studio), talento y seriedad. Todo eso, ingredientes esenciales. Tenemos retos grandes, entre ellos, una facultad que dista de conocernos y personajes con algunos recelos, aunque eso ha tenido progresos notables. También, el reto de empujar unas Ciencias Cognitivas uruguayas equilibradas, donde las tradiciones psicológicas, antropológicas, lingüísticas, biológicas y filosóficas, entre otras, puedan aportar en versiones actualizadas de su propio campo, sintonizadas con el progreso de sus disciplinas y de la epistemología. En sus mejores versiones, estas colaboraciones podrían minimizar el lastre que en ocasiones representa la estructura federal de la Udelar y las distancias con los que aún en el imaginario académico de muchos se siguen situando los campos de conocimiento.

In September 2012 I was waiting for the date for my doctoral thesis defense at the Universitat Autònoma de Barcelona. Like any new doctor, I was looking for a place to continue and project my academic activity. Ana Pires, who was my doctoral partner, told me about the Roberto Caldeyro y Barcia fund scholarships in Uruguay. This option was an alternative that stood out for being a new space with familiar faces (Alejandro Maiche and Fernando González). I submitted and got the scholarship, as did Victoria Gradin. Throughout the submission process I had the support of Juan Valle-Lisboa and Camila Zugarramurdi. After two years of the scholarship I had a position in Salto and another one in Montevideo. I had to decide whether to return to my homeland -Mexico- or stay in Montevideo. I chose to stay, for academic and personal reasons. Since 2014 I have been promoting a line of research on the production of meaning. In this effort I have had the pleasure of the support of national and international students and colleagues. Maria Noel, Mauricio, Agustina, Martín, Carlos, Mauro are, among the students, those who have most clearly bet on the proposed project. Julio Santiago, Jorge Vivas, Alejandro Fojo, Yliana Rodríguez, Manuel Garcia-Ruiz, and Jill Morford have been colleagues who, with different scopes, have supported the proposal. As a human organization, CIBPsi has had good and bad moments and habits, but it has essential virtues to research with passion, fellowship (e.g., Álvaro Cabana always helps us when we get stuck with the R Studio), talent and seriousness. All these are essential ingredients. We have great challenges, among them, a faculty that is far from knowing what we do and leaving some misgivings, although it has made remarkable progress. Also, the challenge of pushing for a balanced Uruguayan Cognitive Sciences, where the psychological, anthropological, linguistic, biological, and philosophical traditions, among others, can contribute in updated versions of their own field, aligned the progress of their disciplinary and epistemological changes. In their best versions, these collaborations could minimize the burden that sometimes represents the federal structure of the Udelar and the distances with which the fields of knowledge are still located in the academic imaginary of some colleagues.

Roberto Aguirre. Profesor Adjunto G3. Facultad
de Psicología. Titular de la línea Mente, Acción
y Lenguaje (MAL) en el CIBPsi.
Labpex CENUR-Salto

Preparación de materiales del proyecto Mate Marote
Preparing materials of Mate Marote project

Preparación de materiales del proyecto **Mate Marote**, que contó con la participación de investigadores del **CIBPsi**. Este proyecto incluyó un software educativo libre y de código abierto formado por un conjunto de juegos para niños en edad escolar y preescolar que buscan estimular distintos aspectos esenciales del aprendizaje, ayudando a sentar las bases para un desarrollo cognitivo adecuado.

Preparation of materials for the Mate Marote project, with the participation of CIBPsi researchers. This project included free and open-source educational software consisting of a series of games for school-age and preschool children that seek to stimulate different essential aspects of learning, helping to lay the foundations for proper cognitive development.

Me acerqué al Centro de Investigación Básica en Psicología (CIBPsi) a mediados de 2017, cuando Alejo Acuña y Laura Uriarte me invitaron a formar parte del equipo de un proyecto PAIE que estaban escribiendo. Participé del mismo como estudiante de grado de Licenciatura en Ciencias Biológicas. El proyecto fue financiado por CSIC y lo estuvimos ejecutando durante todo el año 2018. Posterior a eso nos pareció bueno poder continuar con la investigación y continuamos reclutando participantes y analizando los datos durante el siguiente año. En el 2019 gané un cargo vinculado al CIBPsi y continúe participando de proyectos de investigación relacionados a la actual línea de investigación “Neurociencias Cognitivas y Salud Mental”, particularmente, en el desarrollo y la implementación de tareas interactivas. Encantado por la temática, realice mi trabajo de grado enmarcado en los temas de ésta línea. Actualmente estoy cursando la Maestría en Ciencias Cognitivas, estudiando las activaciones cerebrales relacionadas con los comportamientos de acercamiento y evitación social. Continúo colaborando en los proyectos de la línea. Considero que el pasaje por el centro ha sido fundamental para mi formación como investigador, ayudándome a crecer tanto personalmente como académicamente, capacitandome en cuanto a temas relacionados a la salud mental, programación, diseño de tareas experimentales, adquisición y análisis de datos, y en cómo pensar un problema de investigación y como resolverlo. Además de contar con un excelente grupo humano siempre dispuesto a colaborar, que conforma el centro, también es de destacar que el mismo organiza actividades como cursos, charlas y seminarios que nos permite compartir conocimientos entre las distintas líneas de investigación y con investigadores invitados. Agradezco sobre todo el apoyo de la Dra. Victoria Gradin y el Dr. Alvaro Cabana, que ha sido mi tutora de grado y mi co-tutor -correspondientemente- por su apoyo durante mi formación, su orientación y contención siempre. Hoy son mis tutores de posgrado. Espero que el centro pueda seguir creciendo, incorporando nuevas líneas y dando la oportunidad a nuevos estudiantes para formarse como investigadores.

I approached the Centro de Investigación Básica en Psicología (CIBPsi) in mid-2017, when Alejo Acuña and Laura Uriarte invited me to be part of the team of a PAIE project they were writing. I participated in it as an undergraduate student of Biological Sciences. The project was funded by CSIC and we were running it throughout 2018. After that we thought it was good to continue with the research. Then we kept on recruiting participants and analyzing the data during the following year. In 2019 I won a position at the CIBPsi and kept participating in research projects related to the current line of research “Neurociencias Cognitivas y Salud Mental”, particularly in the development and implementation of interactive tasks. Delighted by the subject, I did my undergraduate work framed in the topics of this line. I am currently pursuing my Master's degree in Cognitive Sciences, studying brain activations related to social approach and avoidance behaviors. I keep collaborating on the projects of this line. I consider that my time at the center has been essential for my training as a researcher, helping me to grow both personally and academically, training me in topics related to mental health, programming, experimental task design, data acquisition and analysis, how to think about a research problem, and how to solve it. Besides having an excellent group of people always ready to collaborate, who make up the center, it is also worth mentioning that the center organizes activities such as courses, lectures, and seminars that allow us to share knowledge among the different lines of research and with invited researchers. I am especially grateful for the support of Dr. Victoria Gradin and Dr. Alvaro Cabana, who have been my undergraduate tutor and my co-tutor -correspondingly- for their support during my training, their guidance and support. Today they are my graduate tutors. I hope that the center can continue to grow, incorporating new lines and giving the opportunity to new students to train as researchers.

Sebastian Morales Duran. Licenciado en Ciencias Biológicas. Maestrando en Ciencias Cognitivas. Ayudante G1, CIBPsi

Presentes para participantes en estudio de procesamiento de Lengua deSeñas
Gifts for Deaf participantes in a project devoted to Sign Language processing

Relojes de arena como los mostrados en estas imágenes fueron otorgados como presentes a informantes sordos señantes de la Lengua de Señas Uruguaya (LSU) en el proyecto **Frecuencia subjetiva en LSU**. El proyecto, cuyo responsable fue el maestrando **Martín Dutra**, fue financiado por el Programa de Iniciación a la investigación de la Comisión Sectorial de Investigación Científica de la Universidad de la República.

Hourglasses like the ones shown in these images were given as gifts to deaf informants who sign Uruguayan Sign Language (LSU) in the project **Frecuencia subjetiva en LSU**. The project, whose head was the Master's student **Martín Dutra**, was funded by the Programa de Iniciación a la investigación de la Comisión Sectorial de Investigación Científica at the Universidad de la República.

Mi experiencia se remonta al año 2014, siendo una estudiante de la Licenciatura en Psicología, un tanto desorientada con respecto al lugar que desearía desarrollar en su rol de futura psicóloga. Fue así que participé de la Pasantía en Psicología Experimental y conocí el CIBPsi. Con mucha curiosidad y asombro, al saber sobre las investigaciones que se estaban desarrollando en el centro y que podía ser parte de un equipo de investigación, decidí integrarme a la línea de Atención. Bajo la mentoría de Alejandra Carboni y Francisco Cervantes, desarrolle en el CIBPsi mi trabajo de grado y actualmente la tesis de posgrado. Acompañada de un gran equipo, que estimulo en mí las ganas de responder preguntas de investigación, aprender a desarrollar un proyecto y sobre la rigurosidad científica, me deje vislumbrar por el vasto campo de las Neurociencias y decidí continuar mi formación en esa área. Indudablemente, la oportunidad de transitar por la investigación en Psicología experimental y formación en Ciencias Cognitivas me ha integrado a la comunidad científica y académica regional. Esta experiencia ha generado en mí el interés en continuar participando, explorando y compartiendo en el centro e interinstitucionalmente.

My experience dates back to 2014, when I was somewhat disoriented as an undergraduate psychology student regarding the place I would like to develop as a future psychologist. Thus, I participated in the Práctica de Psicología Experimental and got to know CIBPsi. With much curiosity and amazement, I learned about the research being developed at the center and that I could be part of a research team. So, I decided to join the research line in Attention. Under Alejandra Carboni's and Francisco Cervantes's mentorship, I developed my undergraduate work and my graduate thesis at CIBPsi. Supported by a great team, who stimulated my desire to answer research questions and learn how to create a project and about the scientific rigor, I let myself glimpse the vast neurosciences field and decided to continue my training in this area. Undoubtedly, the opportunity to go through Experimental Psychology research and Cognitive Sciences training has integrated me into the regional scientific and academic communities. This experience has generated my interest to continue participating, exploring, and sharing at the center and with other institutions.

Thaiz Priscilla Sánchez Costa. Licenciada en
Psicología, maestranda en Ciencias Cognitivas.
Ayudante G1, CIBPsi

Uso de eye-tracker portátil en investigación sobre comportamiento alimentario
Using mobil eye-tracker for Food Behavior

El el contexto de la línea de **Comportamiento Alimentario** se han desarrollado diversos proyectos vinculados a la influencia de la información de las etiquetas en la selección de alimentos, los cuales han contado con financiación de la Comisión Sectorial de Investigación Científica (CSIC) de la Universidad de la República, el Instituto Nacional de Alimentación del Ministerio de Desarrollo Social y el Fondo de las Naciones Unidas para la Infancia (UNICEF, por sus siglas en inglés), Uruguay. En la imagen se presenta un ejemplo de uno de los estudios realizados, en el cual se utilizó un detector de movimientos oculares móviles para estudiar la toma de decisiones en supermercados. Los resultados permitieron confirmar que los consumidores no realizan un procesamiento detallado de la información disponible y toman las decisiones en tiempos sumamente cortos, reforzando la necesidad de incluir información simplificada y regular la presencia de elementos persuasivos en las etiquetas de alimentos no saludables.

In the context of the **Food Behavior** research line, several projects have been developed related to the influence of label information on food selection, which have been funded by the Comisión Sectorial de Investigación Científica (CSIC) of the Universidad de la República, the Instituto Nacional de Alimentación del Ministerio de Desarrollo Social and el Fondo de las Naciones Unidas para la Infancia (UNICEF, English abbreviation), Uruguay. The image shows an example of one of the studies carried out, in which a mobile eye movement detector was used to study decision-making in supermarkets. The results confirmed that consumers do not carry out a detailed processing of the available information and make decisions in extremely short times, reinforcing the need to include simplified information and regulate the presence of persuasive elements on unhealthy food labels.

Comencé mi trabajo en el Cibpsi colaborando en proyectos de investigación vinculados a la Linea de Procesamiento de Lenguaje. Meses después iniciaba la Maestría en Ciencias Cognitivas bajo la tutoría de dos miembros de esa misma línea, el Dr. Juan Valle Lisboa y la Dra. Camila Zugarramurdi. Actualmente, continúo mi formación académica en el CIBPsi en donde también me desempeño como ayudante de investigación.

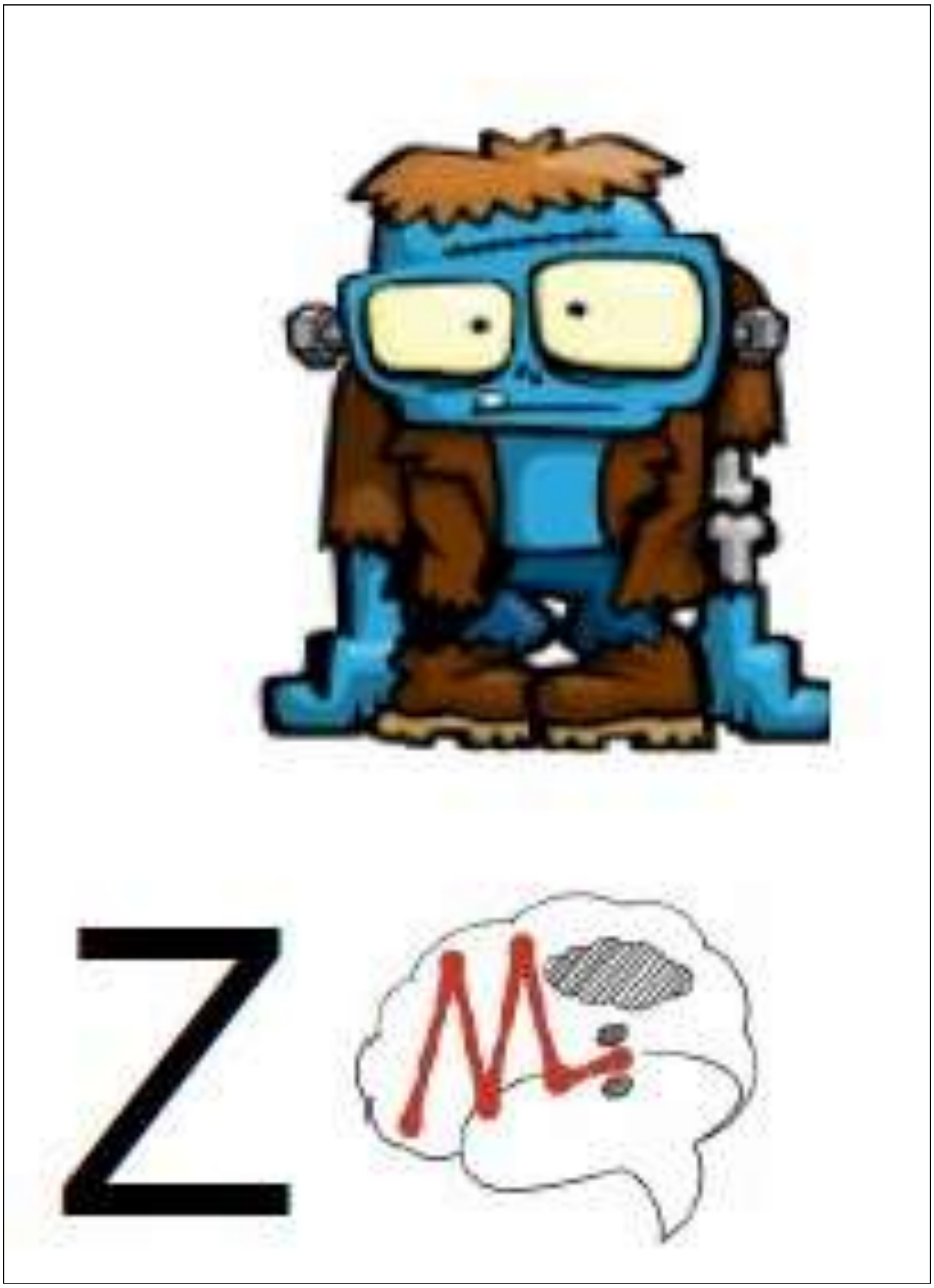
Ser parte de este equipo me dio la posibilidad de dar mis primeros pasos en el campo de la investigación científica y estar en contacto permanente con investigadores formados en diferentes disciplinas, siendo estos aportes fundamentales tanto para mi formación académica como profesional.

I began my work at CIBPsi, collaborating on research projects linked to language processing. Months later, I started my master's degree in Cognitive Sciences under the mentorship of two members of the same line, Dr. Juan Valle Lisboa and Dr. Camila Zugarramurdi. Currently, I continue my academic training at Cibpsi where I also work as a research assistant.

Being part of this team gave me the opportunity to take my first steps in the field of scientific research and be in permanent contact with researchers trained in different disciplines, which are fundamental for my academic and professional training.

Soledad Assis. Psicopedagoga. Maestranda en Ciencias Cognitivas. Ayudante G1, CICEA

Reuniones académicas de grupos y líneas
Academic meetings of research lies and groups



En marzo de 2019 el grupo Estudios Experimentales de la metáfora cognitiva en Español y LSU (hoy línea Mente, Acción y Lenguaje) empezó sus reuniones académicas quincenales o semanales para l a revisión de artículos, capítulos de libro y avances de investigación propia como herramienta para dar coherencia, densidad conceptual y teórica a sus trabajos , ajustándose a los momentos formativos de sus miembros. El grupo se llamó Zombies debido a una broma de algunos de los miembros no muy contento con realizar las sesiones a las 8 am una vez por semana.

In March 2019 the group Estudios Experimentales de la metáfora cognitiva en Español y LSU (today Mente, Acción y Lenguaje) began its biweekly or weekly academic meetings to review articles, book chapters and research advances as a tool to give coherence, conceptual and theoretical density to their work, adjusting to the formative moments of its members. The group was called Zombies due to a joke of some of the members who were not very happy with holding the sessions at 8 am once a week.

Conocí el Centro de Investigación Básica en Psicología como estudiante avanzada de la Licenciatura en Psicología cuando cursé la Pasantía en Psicología Experimental en el año 2013. La pasantía la realicé en el marco de la línea de investigación "Neurociencia Cognitiva y Salud Mental" dirigida por la Dra. Victoria Gradin. Al año siguiente participé en un Proyecto de Apoyo a la Investigación Estudiantil (PAIE) y realicé mi Trabajo Final de Grado en esta línea de investigación. Desde ese entonces tuve la oportunidad de participar de diversos proyectos de la línea y realizar mi maestría allí. Gracias a mi pasaje por el CIBPsi aprendí lo que implica investigar en el campo de las neurociencias cognitivas y tuve la oportunidad de participar de diversas actividades académicas que me formaron como investigadora. Agradezco a los integrantes de la línea y a los distintos investigadores del centro por contagiarme su pasión por esta área y permitirme conocer colegas y amigos que me acompañan hasta el día de hoy.

I got to know the Center for Basic Research in Psychology as an advanced student of the Bachelor's Degree in Psychology when I completed the Pasantía en Psicología Experimental in 2013. I did the internship as part of the research line "Neurociencia Cognitiva y Salud Mental", directed by Dr Victoria Gradin. The following year I participated in a student research project (PAIE) and did my undergraduate thesis project in this line of research. Since then, I have had the opportunity to participate in various projects at this research line and to do my master's degree there. Thanks to my time at CIBPsi, I learned what research in the field of cognitive neurosciences entails, and I had the opportunity to participate in various academic activities that shaped me as a researcher. I would like to thank the members of the research line and the different researchers at the centre for passing on their passion for this area and allowing me to meet colleagues and friends who are still with me today.

Valentina Paz. Licenciada en Psicología, magíster
y doctoranda en Ciencias Biológicas
(Neurociencias). Asistente G2 del Instituto
de Psicología Clínica.
Investigadora asociada del CIBPsi

Publicaciones colectivas para público estudiantil y local
Group publications for local academic audiences

Diversos investigadores relacionados con el **CIBPsi** han participado en publicaciones, sometidas a concurso de financiamiento en el Programa de Publicaciones de la Comisión Sectorial de Investigación Científica (CSIC), con alcance para el público académico local. Algunas obras han ido dirigidas a estudiantes de grado en Psicología y de disciplinas interesadas en temas cognitivos y otras a públicos específicos del ámbito de la educación primaria y secundaria. En la foto, investigadores del **CIBPsi** muestran el libro “Psicología del Tiempo: una introducción a la temporalidad en las Ciencias del Comportamiento. Otros títulos corresponden al “Manual de Introducción a la Psicología Cognitiva” y “Aportes de la Ciencias Cognitivas a la Educación”.

Several researchers related to the **CIBPsi** participated in publications, submitted to the Programa de Publicaciones de la Comisión Sectorial de Investigación Científica (CSIC), written for local academic public. Some of these publications were addressed to undergraduate students in Psychology and other disciplines interested in cognitive issues, and others to specific audiences in the field of primary and secondary education. In the photo, **CIBPsi** researchers show the book “Psicología del Tiempo: una introducción a la temporalidad en las Ciencias del Comportamiento. Other titles correspond to "Manual de Introducción a la Psicología Cognitiva" and "Aportes de la Ciencias Cognitivas a la Educación".