UNIVERSIDAD
NACIONAL
DE COLOMBIA

Entre el saber y el deber saber:

Mecanismos inferenciales de la Ilusión de Profundidad Explicativa

Christian David Gaviria Martínez

Universidad Nacional de Colombia
Facultad de Ciencias Humanas
Departamento de Psicología
Bogotá, Colombia

2019.

Entre el saber y el deber saber:

Mecanismos inferenciales de la Ilusión de Profundidad Explicativa

Christian David Gaviria Martínez

Tesis presentada como requisito parcial para optar al título de:
Doctor en Psicología

Director:
Javier Alejandro Corredor Aristizábal, Ph.D.

Línea de Investigación:
Cognición, medios y educación

Grupo de Investigación:
Cognición, prácticas y aprendizaje

Universidad Nacional de Colombia
Facultad de Ciencias Humanas
Departamento de Psicología
Bogotá, Colombia

2019.

A los que antes de partir

sembraron las semillas que germinaron en estas páginas.

Agradecimientos

Quizás las deudas de amistad y gratitud sean las únicas que uno disfruta pagar aunque jamás pueda saldarlas. Esta tesis no hubiese sido posible sin la ayuda generosa de muchas personas a lo largo de los últimos cuatro años. Agradezco de manera especial a Zadkiel Zuluaga, María Camila Garzón y Daniela Villareal por hacer parte del equipo de trabajo en el cual se gestó este proyecto de investigación. Mi gratitud también se extiende a todos los miembros del Semillero del Laboratorio de Cognición, Medios y Educación del Departamento de Psicología de la Universidad Nacional entre los años 2015 y 2019, por su colaboración durante las largas jornadas de recolección de datos, y por brindarme siempre un ambiente acogedor y estimulante para el avance de mi trabajo. Mis compañeros del programa de Doctorado en Psicología de la Universidad Nacional de Colombia, Costanza Beltrán, Reynel Chaparro, Jazmine Escobar, Héctor Gamba, Mónica Sánchez y Cristina Santana proporcionaron valiosas observaciones que contribuyeron a reforzar los cimientos conceptuales de este proyecto.

La ayuda de mis colegas y amigos Angélica Álvarez, Camilo Arias, Susana Barradas, Carolina Castro, Rafael Cendales, Diana Cortés, Paulo Dillon Soares, Juliana Fonseca, Jazmine Escobar, Mariantonia Lemos y Jonathan Tovar fue crucial durante las fases de recolección de datos. Mis amigos María Camila Alviar, Camilo Arias, Jonathan Buriticá, Carolina Castro, María Camila Contreras, Óscar Córdoba, Diana Cortés, Sandra Cortés, Martín Cuervo, Paulo Dillon Soares, Felipe González, Tiffany Jiménez, María Angélica López, Alejandra Martínez, Natalia Piar, Tatiana Plata, Miguel Puentes, Adriana Saavedra, Luis Sánchez, John Solórzano y Zadkiel Zuluaga me ayudaron a formular las preguntas correctas en el momento justo y a hacer más amables los momentos críticos del desarrollo de este trabajo.

Estoy muy agradecido con los profesores Mikel Asensio Brouard, Elena Pol, Nieves Pérez, Mario Carretero, Margarita Limón, María Sotillo y Everardo Pérez de la Universidad Autónoma de Madrid por su hospitalidad durante mi estancia doctoral. Sus agudas observaciones sobre mi trabajo ampliaron mi perspectiva sobre sus posibles aplicaciones en contextos educativos informales. Mis amigos Aracely Lugo y Diego Valdés también contribuyeron en gran medida a que mi estancia en Madrid fuese tan enriquecedora como placentera. Estaré en deuda por siempre con Esther Romanos y su familia por su apoyo en los días difíciles y por el privilegio de su amistad.

Agradezco a los profesores Germán Gutiérrez, William Jiménez Leal, Marisol Lamprea y Roberto Posada por su voto de confianza y su orientación sensata en diferentes momentos de mi formación de posgrado. Me siento honrado de que mis jurados Santiago Arango, Héctor Carvacho y Ricardo Tamayo hayan aceptado realizar una revisión rigurosa del proyecto de tesis y del presente documento. Sus aportes durante el proceso de evaluación contribuyeron a expandir y consolidar mi propuesta original, al tiempo que propiciaron en mí la reflexión sobre sus alcances y limitaciones.

Mi director Javier Corredor supo encontrar un delicado balance entre su deber de supervisar el desarrollo de este trabajo y su determinación a concederme el peligroso privilegio de la autonomía creativa. Su estilo proactivo y su confianza permanente en el éxito de este trabajo fueron decisivos para su culminación.

Por último, no encuentro palabras suficientes para expresar la gratitud que me une a mi madre, Ana Martínez Ruge, a mi hermano, Germán Andrés Gaviria, a mi pareja, Natalia Castro Vega, a mi gata Rubí y a toda mi familia por estar a mi lado durante los ascensos, descensos y vientos cruzados de mi proceso de formación doctoral.

Gracias por mantener viva la llama que ha permitido que esta tesis finalmente vea la luz.

Christian David Gaviria Martínez
Septiembre de 2019
Bogotá, Colombia.

Resumen

La ilusión de profundidad explicativa (IPE) es la tendencia de las personas a sobrestimar su capacidad para explicar los mecanismos causales de procesos naturales o sociales. La investigación previa sobre este sesgo metacognitivo ha sugerido que las personas exhiben la IPE cuando confunden el reconocimiento de patrones causales abstractos de los dominios físicos, biológicos, psicológicos y sociales con el conocimiento de los mecanismos de funcionamiento de objetos y procesos específicos. Sin embargo, esta explicación resulta limitada para dar cuenta de las diferencias en la sobrestimación del conocimiento sobre temas con propiedades causales similares y que pertenecen al mismo dominio explicativo. En una línea de investigación relacionada, algunos estudios empíricos sobre juicios metacognitivos en tareas de memoria y aprendizaje han mostrado que las personas utilizan información social y contextual como claves inferenciales para estimar su conocimiento sobre contenidos específicos. Conectando estas dos líneas de investigación para esclarecer los mecanismos metacognitivos y motivacionales involucrados en la IPE, se llevaron a cabo cinco experimentos para indagar si la deseabilidad social del conocimiento sobre un tema influye en la estimación de su comprensión causal y si este efecto está relacionado con el tipo de procesamiento de la información (tipo 1 y tipo 2). En particular, se examinó si la deseabilidad social percibida del conocimiento se asocia con la magnitud de la IPE (Experimentos 1, 2 y 2a) y si esta relación es moderada por la habilidad para realizar procesamientos analíticos/controlados (Experimentos 3 y 4). Los resultados mostraron que la IPE fue mayor en los temas cuyo conocimiento fue calificado como más deseable socialmente en tres dominios: histórico, económico y dispositivos (Experimento 1). Este efecto no se debió a la ausencia total de familiaridad con los temas o a un sesgo de auto-promoción (Experimentos 2 y 2a). Adicionalmente, la habilidad para llevar a cabo procesamientos tipo 2 se asoció con estimaciones más precisas de la comprensión causal de los temas más deseables socialmente (Experimento 3), mientras que la sobrecarga ejecutiva correlacionó con mayores magnitudes de la IPE (Experimento 4). Estos hallazgos son discutidos a la luz de los modelos que integran las teorías de procesamiento dual y la arquitectura de los procesos metacognitivos.

Palabras clave: metacognición, comprensión, explicación, deseabilidad social, procesos duales, ilusión de profundidad explicativa.

Abstract

The illusion of explanatory depth (IOED) is the people's tendency to overestimate their ability to explain the causal mechanisms of natural or social processes. Previous research on this metacognitive bias has suggested that people display the IOED when they confuse the recognition of abstract causal patterns in physical, biological, psychological, and social domains with the knowledge about the mechanisms underlying how specific objects and processes work. However, this explanation is limited to explain differences in the overestimation of knowledge about topics with similar causal properties and belonging to the same explanatory domain. In a related line of research, some empirical studies about metacognitive judgments in memory and learning tasks have shown that people use social and contextual information as inferential cues to estimate their knowledge about specific contents. This dissertation connects these two lines of research to shed light on the metacognitive and motivational mechanisms involved in the IOED. In particular, five experiments were carried out to investigate whether the social desirability of knowledge about a topic influences the estimation of their causal understanding, and whether this effect is related to the information processing type (type 1 and type 2). More specifically, it was examined whether the perceived social desirability of knowledge is associated with the IOED magnitude (Experiments 1, 2, and 2a), and whether this relationship is moderated by the ability to perform analytic/controlled processing (Experiments 3 and 4). The results showed the IOED was greater in topics whose knowledge was rated as more socially desirable in three domains: historic, economic and devices (Experiment 1). This effect was not due to the total lack of familiarity with the topics or a self-enhancement bias (Experiments 2 and 2a). In addition, the ability to carry out type 2 processing was associated with more accurate estimates of the causal understanding about the more socially desirable topics (Experiment 3), and the executive overload correlated with greater magnitudes of the IOED (Experiment 4). These findings are discussed in light of models that integrate the dual processes theories and the architecture of metacognitive processes.

Keywords: metacognition; understanding; explanation; social desirability; dual processes; illusion of explanatory depth.

Tabla de Contenidos

Resumen 8

Abstract 9

Lista de Tablas 16

Lista de Figuras 17

Introducción 19

Capítulo 1: Revisión Teórica 27

¿Por Qué Ocurre la Ilusión de Profundidad Explicativa? 27

Deseabilidad social del conocimiento e ilusión de profundidad explicativa 34

Deseabilidad Social del Conocimiento y *Feeling of Knowing* (FOK) 38

Memoria y juicios metacognitivos: ¿Cómo sabemos que sabemos? 38

¿Acceso directo o claves heurísticas?: Mecanismos metacognitivos del FOK 43

La influencia de claves sociales en el FOK 48

La hipótesis de mecanismo único de los juicios metacognitivos 53

Teorías de Procesamiento Dual y Metacognición 55

Encrucijadas inferenciales 55

Modelos de procesamiento dual 58

La dualidad de los procesos metacognitivos 64

Formulación del Problema 70

Preguntas de investigación e hipótesis 70

Capítulo 2: Método 73

Resumen de los Experimentos 73

Tamaños de Muestra 74

Experimento 1: Deseabilidad Social e Ilusión de Profundidad Explicativa 75

Participantes 77

Diseño 77

Materiales y procedimiento 78

Test de verificación de la manipulación de las instrucciones (*Instructional manipulation check*) 78

Procedimiento 78

Experimento 2: Controlando el Sesgo de Auto-promoción (*Self-enhancement*) y el Efecto de la Familiaridad 80

Participantes 83

Diseño 83

Materiales y Procedimiento 83

Escala de deseabilidad social del conocimiento 83

Escala de familiaridad 84

Prueba de *overclaiming* sobre el conflicto armado colombiano 84

Procedimiento 84

Experimento 2a: Controlando el Efecto de Exposición Previa 86

Participantes 86

Diseño 86

Materiales y Procedimiento 87

Experimento 3: Deseabilidad Social, Ilusión de Profundidad Explicativa y Tipo de Procesamiento 88

Participantes 89

Diseño 89

Materiales y Procedimiento 90

Cognitive Reflection Test (CRT) 90

Procedimiento 90

Experimento 4: Manipulación Experimental de la Sobrecarga Ejecutiva 92

Participantes 93

Diseño 93

Materiales y Procedimiento 94

Tarea de memoria visoespacial 94

Tarea de memorización de sílabas 95

Procedimiento 95

Capítulo 3: Resultados 100

Experimento 1 100

Plan de análisis 100

IPE y deseabilidad social del conocimiento 102

Interacción entre deseabilidad social y dominio de conocimiento 104

Interacción entre orden de presentación y deseabilidad social 106

Interacción entre orden de presentación, dominio de conocimiento y deseabilidad social 107

Experimento 2 109

Plan de análisis 109

IPE y deseabilidad social 109

IPE, deseabilidad social y *overclaming* 111

Experimento 2a 114

Plan de análisis 114

Experimento 3 117

Plan de análisis 117

IPE y deseabilidad social 118

Interacción entre deseabilidad social y CRT como predictores de la IPE 120

Experimento 4 121

Plan de análisis 121

IPE y deseabilidad social 122

IPE, sobrecarga cognitiva y CRT 126

Discusión 129

Hallazgos principales 129

Limitaciones y futuros estudios 135

Conclusión 140

Referencias 145

Anexo A. Escala de deseabilidad social y temas del estudio preliminar 182

(Experimento 1) 182

Anexo B. Deseabilidad social de los temas seleccionados para el estudio principal

(Experimento 1) 184

Anexo C. Test de verificación de lectura de las instrucciones de la tarea 186

Anexo D. Instrucciones de la escala de comprensión (Experimento 1) 187

Anexo E. Escala de deseabilidad social del conocimiento (Experimento 2 y 2a) 189

Anexo F. Deseabilidad social de los temas relacionados con la historia del conflicto armado

colombiano (Experimento 2) 189

Anexo G. Familiaridad con los temas presentados en el Experimento 2 192

Anexo H. Medias y confiabilidad de la escala de deseabilidad social de los temas

seleccionados en el Experimento 2 193

Anexo I. Medias y confiabilidad de la escala de deseabilidad social de los temas

seleccionados en el Experimento 2a 194

Anexo J. Prueba de *overclaiming* sobre conocimiento de la historia del conflicto armado

colombiano 195

Anexo K. Confiabilidad de la escala de deseabilidad social de los temas del estudio preliminar (Experimento 3) 197

Anexo L. Deseabilidad social de los ítems seleccionados en el Experimento 3 199

Anexo M. Instrucciones de la escala de comprensión (Experimento 3) 202

Anexo N. *Cognitive Reflection Test* (CRT)..... 203

Anexo O. Información enviada a los participantes al concluir el Experimento 3..... 203

Anexo P. Materiales utilizados en las tareas de memoria del Experimento 4..... 205

Anexo Q. Medias y confiabilidad de la escala de deseabilidad social de los temas utilizados en el Experimento 4..... 206

Anexo R. Implicaciones Éticas..... 207

Lista de Tablas

Tabla 1. Resumen de los experimentos realizados 98

Tabla 2. Magnitud de la IPE para cada tema del Experimento 1 102

Tabla 3. Modelos lineales jerárquicos de las relaciones entre deseabilidad social, dominio de conocimiento, orden de presentación e IPE (Experimento 1) 104

Tabla 4. Diferencias en la magnitud de la IPE según el dominio de conocimiento, orden de presentación y deseabilidad social de los temas 108

Tabla 5. Magnitud de la IPE en cada uno de los temas del Experimento 2 110

Tabla 6. Modelos lineales jerárquicos de las relaciones entre deseabilidad social, overclaiming e IPE (Experimento 2) 112

Tabla 7. Magnitud de la IPE en cada uno de los temas del Experimento 2a 115

Tabla 8. Modelos lineales jerárquicos de las relaciones entre deseabilidad social e IPE (Experimento 2a) 117

Tabla 9. Medias de las diferencias entre las estimaciones previas y posteriores para cada uno de los temas del Experimento 2 118

Tabla 10. Modelos lineales jerárquicos de las relaciones entre deseabilidad social, puntuaciones en el CRT e IPE (Experimento 3) 119

Tabla 11. Magnitud de la IPE en cada uno de los temas del Experimento 4 122

Tabla 12. Modelos lineales jerárquicos de las relaciones entre deseabilidad social e IPE (Experimento 4) 124

Lista de Figuras

Figura 1. Clasificación de las facetas metacognitivas 41

Figura 2. Secuencias de activación de los sistemas 1 y 2 en los modelos de procesamiento serial, paralelo e híbrido 60

Figura 3. Modelo híbrido de tres etapas 63

Figura 4. Modelos de procesamiento metacognitivo del efecto de la deseabilidad social del conocimiento en la IPE 68

Figura 5. Modelo de las relaciones entre las variables de interés 73

Figura 6. Ejemplos de matrices utilizadas en la tarea de memoria visoespacial en las condiciones de sobrecarga baja y sobrecarga alta del Experimento 4 95

Figura 7. Diferencias en la magnitud de la IPE entre los temas de deseabilidad alta y baja (Experimento 1) 103

Figura 8. Diferencias en la magnitud de la IPE en función de la deseabilidad social y el orden de presentación de los temas 107

Figura 9. Interacción entre la deseabilidad social y TFA como predictores de la IPE 114

Figura 10. Puntajes en el CRT como moderadores de la asociación entre deseabilidad social del conocimiento y la magnitud de la IPE 121

Figura 11. Modelos de ecuaciones estructurales de efectos directos y efecto mediado de la sobrecarga ejecutiva y el desempeño en las tareas de memoria del Experimento 4 128

“Tiger got to hunt, bird got to fly
Man got to sit and wonder ‘why, why, why’
Tiger got to sleep, bird got to land
Man got to tell himself he understand”
Kurt Vonnegut

“*If you're feeling cocky, it's because there's something you don't know.”*
Eoin Colfer

Introducción

En diálogos platónicos como *Laques*, *Menón*, *Eutifrón* o *Teeteto*, Sócrates intenta poner a prueba la comprensión que tienen sus interlocutores sobre conceptos morales y filosóficos como el coraje, la virtud, la piedad y el conocimiento, respectivamente. Estos interlocutores suelen ser personajes destacados que presuntamente conocen en detalle el tema sobre el cual Sócrates está indagando. Así, Laques es un general ateniense, veterano de las guerras del Peloponeso; Menón, joven militar aristócrata, asegura que ha pronunciado muchos discursos sobre la virtud; Eutifrón está a punto de acusar a su propio padre de cometer una impiedad y Teeteto es un aventajado estudiante de matemáticas. En general, de los 34 interlocutores de Sócrates mencionados en los *Diálogos*, 21 de ellos tienen la reputación de ser conocedores del tema principal de la discusión. Sin embargo, en opinión de Sócrates, ninguno de ellos logra demostrar su sabiduría sobre el tema respectivo y solo siete aceptan finalmente su ignorancia (Benson, 2011). De manera reiterada, Sócrates encuentra que sus interlocutores sobrestiman lo que saben sobre los temas en los cuales se consideran expertos. Subrayando este hallazgo, el filósofo declara: "me pareció que los de mayor reputación estaban casi carentes de lo más importante para el que investiga según el dios" (*Apología* 22a, trad. en 1985). Refiriéndose a los artesanos, añade: "por el hecho de que realizaban adecuadamente su arte, cada uno de ellos estimaba que era muy sabio también respecto a las demás cosas, incluso las más importantes, y ese error velaba su sabiduría." (*Apología* 22d-e, trad. en 1985).

Esta observación socrática es tan válida para las democracias del siglo V A.C. como para las del siglo XXI. En una encuesta realizada en 2012 después de que la Suprema Corte de Estados Unidos fallara a favor de la inclusión de ciertos cambios en la reforma del

sistema de aseguramiento en salud de la administración de Barack Obama (el denominado "*Obamacare*"), 36% de los encuestados estuvieron a favor del fallo, 40% en contra y 24% se abstuvieron de opinar. No obstante, aunque el 76% de las personas expresaron una opinión sobre el concepto de la Corte, solo 55% conocían su contenido (Pew Research Center, 2012). En la misma línea, en una encuesta realizada en 2014 durante la ocupación de tropas rusas en la península de Crimea en Ucrania, se encontró que los ciudadanos estadounidenses que se declaraban más a favor de una intervención militar de su país en dicha región cometieron más errores al tratar de ubicar a Ucrania en un mapa (Dropp, Kertzer, & Zeltzoff, 2014). Más recientemente, Motta, Callaghan y Sylvester (2018) reportaron que las personas con desempeños bajos en una prueba de conocimiento sobre las causas del autismo creen saber más sobre este trastorno que los expertos. Así mismo, tienden a rechazar los programas de vacunación obligatoria y a aprobar que los no expertos tengan un papel determinante en la toma de decisiones sobre dichas políticas. De manera similar, en encuestas realizadas en Estados Unidos, Francia y Alemania se ha encontrado que las personas con actitudes negativas más extremas frente a los alimentos transgénicos creen tener una comprensión robusta del tema, si bien sus desempeños en una prueba de conocimientos sobre genética tienden a ser bajos (Fernbach, Light, Scott, Inbar, & Rozin, 2019).

Los ejemplos anteriores ilustran una paradoja común a las democracias contemporáneas: los temas de la agenda pública sobre los cuales los ciudadanos tienen posiciones más definidas o incluso extremas son también los menos comprendidos por ellos (Fernbach, Rogers, Fox, & Sloman, 2013; Hall & Raimi, 2018; Ortoleva & Snowberg, 2015; Raimi & Leary, 2014; Somin, 2016). Este hecho adquiere especial relevancia ante el

aumento de los referendos en los estados democráticos durante las últimas décadas (Qvortrup, 2018), ya que a través de este mecanismo de participación directa los votantes tienen el poder de aprobar o rechazar reformas políticas complejas, controversiales y con efectos de largo plazo. Si las personas creen tener un conocimiento suficiente sobre propuestas legislativas que solo entienden de manera superficial, es probable que sus posiciones al respecto estén basadas en información falsa o incompleta. Así, por ejemplo, en una encuesta realizada en el Reino Unido semanas antes del referendo que decidió la permanencia de este país en la Unión Europea (UE), se encontró que cerca del 60% de los ciudadanos encuestados reportaron conocer en buena medida las ventajas y desventajas de dicha permanencia. No obstante, los participantes en la encuesta sobrestimaron el número de inmigrantes provenientes de países de la UE, el gasto anual del gobierno británico por pertenecer a la UE en comparación con otros países, los gastos administrativos de la UE y la cantidad de recursos destinados por el Reino Unido a financiar programas de asistencia a la niñez en otros países miembros. De manera inversa, los encuestados subestimaron la proporción de inversión extranjera proveniente de los países de la UE y el número de funcionarios británicos que trabajan en esta organización (Ipsos MORI, 2016). Estos resultados sugieren que no todos los ciudadanos que votaron a favor o en contra de la salida del Reino Unido de la comunidad europea (*Brexit*) estaban tan bien informados como creían sobre las implicaciones de su decisión en las urnas.

Haciendo eco de esta antigua y vigente intuición sobre la ignorancia humana, una gran cantidad de estudios empíricos han mostrado en las últimas décadas que las personas exhiben un exceso de confianza en la calidad y precisión de sus conocimientos y habilidades (Kardas & O'Brien, 2018; Krosnick, Boninger, Chuang, Berent, & Carnot,

1993; Kruger & Dunning, 1999; Moore & Healy, 2008; Radecki & Jaccard, 1995; Sloman & Fernbach, 2017; Zell & Krizan, 2014). La evidencia de este fenómeno es robusta y se ha encontrado en un amplio espectro de tareas, tales como recordar información previamente memorizada (Bjork, 1994; Eakin, 2005; Koriat & Bjork, 2005; Koriat, Lichtenstein, & Fischhoff, 1980; Winne & Jamieson-Noel, 2002), responder preguntas sobre conocimientos generales (Atir, Rosenzweig, & Dunning, 2015; Bradley, 1981; Fischhoff, Slovic, & Lichtenstein, 1977; Kelley & Lindsay, 1993; Lichtenstein & Fischoff, 1977; Soll & Klayman, 2004), evaluar la comprensión de textos (Glenberg & Epstein, 1985, 1987; Glenberg, Wilkinson & Epstein, 1982; Evans & Stanwyck, 1997; Jaeger & Wiley, 2015; Lin-Miao & Zabrucky, 1998; Maki, Shields, Wheeler, & Zacchilli, 2005) o tomar decisiones (Alba & Hutchinson, 2000; Fabricius & Büttgen, 2013; LaBorde, Mottner, & Whalley , 2015; Park, Gardner, & Thukral, 1988).

La discrepancia entre lo que las personas creen saber y lo que realmente saben está relacionada con las características específicas de algunos tipos de conocimiento. En este sentido, Rozenblit y Keil (2002) encontraron que estudiantes universitarios sobrestimaron más su capacidad para explicar el funcionamiento de dispositivos o fenómenos naturales que su comprensión de hechos (e.g., capitales de países), narrativas (e.g., argumentos de películas) o procedimientos (e.g., cómo hacer un nudo de corbata). Este efecto, denominado "ilusión de profundidad explicativa" (IPE en adelante), ha sido reportado tanto en adultos novatos y expertos (Fisher & Keil, 2016; Lawson, 2006) como en niños (Mills & Keil, 2004). Además, la IPE se ha evidenciado en las estimaciones de comprensión de diversos mecanismos físicos y sociales, tales como dispositivos de uso cotidiano (Fernbach, Sloman, St. Louis, & Shube, 2013; Lawson, 2006; Roeder & Nelson, 2015), fenómenos naturales

(Fisher & Keil, 2016), efectos de comportamientos saludables o responsables con el medio ambiente (Bromme, Thomm, & Ratermann, 2015), consecuencias de la implementación de propuestas políticas (Alter, Oppenheimer, & Zemla, 2010; Fernbach, Rogers, et al., 2013; Roeder & Nelson, 2015; Voelkel, Brandt, & Colombo, 2018) y tratamiento de las enfermedades mentales (Zeveney & Marsh, 2016).

En la psicología cognitiva, la existencia de la IPE ha sido explicada a partir de las propiedades representacionales del conocimiento causal. Específicamente, la explicación predominante postula que la IPE se produce porque las teorías intuitivas que las personas construyen para comprender los procesos naturales y sociales se enfocan en la detección de patrones y estructuras causales generales, prescindiendo de la información específica sobre cómo tales patrones y estructuras son implementados en mecanismos concretos (Keil, 2003, 2012; Keil, Rozenblit & Mills, 2004). Según esta perspectiva, las personas confunden su habilidad para identificar si un mecanismo tiene propiedades causales típicas de los dominios físicos, biológicos, psicológicos o sociales, con la habilidad para explicar detalladamente su funcionamiento. Por ejemplo, cuando las personas logran reconocer los atributos globales de un dispositivo (e.g., para qué sirve, cómo se usa y cuáles son los componentes principales del sistema de transmisión de un automóvil), pueden asumir erróneamente que comprenden cómo funciona (e.g., las interacciones entre el motor, el embrague, la caja de cambios y los rodamientos internos; Rozenblit & Keil, 2002).

Aunque esta aproximación teórica logra conectar el fenómeno de la IPE con las teorías contemporáneas sobre adquisición de conocimiento y formación de conceptos (e.g., Gerstenberg & Tenenbaum, 2017), no explica por qué las personas sobrestiman su comprensión con respecto a ciertos mecanismos pero no a otros. Por ejemplo, se ha

encontrado que quienes han culminado estudios de pregrado sobrestiman su habilidad para explicar temas que pertenecen a su área de formación (e.g., el ciclo de Krebs o la programación orientada a objetos), pero no temas cotidianos (e.g., el funcionamiento de los automóviles o los teléfonos inteligentes). Esto ocurre porque las personas subestiman la cantidad de información explicativa que aprendieron durante su educación formal y que han olvidado con el paso del tiempo (Fisher & Keil, 2016). Este hallazgo muestra que el contenido poco específico de las teorías intuitivas no es la única causa de la IPE, sino que ciertos factores asociados al proceso de estimación de la comprensión causal pueden contribuir a exacerbar o disminuir la magnitud de este sesgo.

¿Por qué las personas creen saber más sobre algunos mecanismos causales que sobre otros? En esta disertación se propone que la influencia de claves inferenciales de origen social en los procesos metacognitivos puede contribuir a explicar estas diferencias. En particular, cuando las personas estiman su comprensión sobre un tema específico, es posible que no solo tengan en cuenta lo que realmente saben, sino también su representación de lo que *deberían* saber. Esta representación puede ser inferida de manera implícita o explícita a partir de claves informativas como la deseabilidad social del conocimiento. Retomando el escenario de la Grecia clásica, es posible que los interlocutores de Sócrates no solo sobrestimaran su comprensión de conceptos filosóficos por su complejidad inherente, sino porque *sentían* que debían conocer en profundidad dichos temas. Si esta interpretación es correcta, es probable que en el contexto de las democracias deliberativas los ciudadanos tiendan a sobrestimar más su comprensión sobre los temas históricos, sociales, económicos o políticos más relevantes, populares, discutidos y con mayor visibilidad mediática. Esta idea es consistente con investigaciones previas que

muestran que las personas creen tener cualidades socialmente deseables por encima del promedio de su grupo de referencia (Alicke, 1985) y evalúan su desempeño en diversas tareas (e.g., resolver una prueba de razonamiento abstracto o de programación de computadores) en función de las creencias que tienen sobre sus propias habilidades (Critcher & Dunning, 2009; Ehrlinger & Dunning, 2003). En la misma línea, se ha encontrado que los participantes en encuestas de opinión pública tienden a expresar opiniones sobre hechos ficticios cuando estos se asocian a temas, dominios o nombres familiares (Bishop, Tuchfarber, & Oldendick, 1986; Schuman & Presser, 1980).

Por otra parte, algunos hallazgos empíricos provenientes de la literatura sobre metacognición y metamemoria, así como de la investigación sobre la IPE, parecen ser consistentes con esta interpretación. Concretamente, se ha encontrado que las personas no juzgan su habilidad para recordar información a partir de una inspección completa de los contenidos memorizados, sino con base en dos tipos de información: las experiencias asociadas a los procesos de aprendizaje y evocación (*recall*) de dicha información (Koriat & Levy-Sadot, 1999; Koriat, Ma'ayan, & Nussinson, 2006; Schwartz, Benjamin, & Bjork, 1997), y las creencias sobre lo que ellos mismos y otros deberían saber sobre un tema o dominio particular (Costermans, Lories, & Ansay, 1992; De Carvalho Filho & Yuzawa, 2001; Fisher & Keil, 2016; Zhao & Linderholm, 2010). Si los juicios metacognitivos generados en tareas de memoria y aprendizaje dependen de mecanismos similares a los que originan los juicios de comprensión causal, es probable que la magnitud de la IPE varíe en función del valor social que las personas le atribuyen al conocimiento de procesos causales específicos.

Examinar la influencia de claves inferenciales de orden social y motivacional sobre los juicios de comprensión causal permitiría identificar las condiciones bajo las cuales este conocimiento es más susceptible de ser sobrestimado. A su vez, esta información resultaría útil en el diseño de estrategias que permitan a las personas reevaluar su comprensión con mayor precisión y monitorear de manera más efectiva sus procesos de aprendizaje de contenidos específicos (Bjork, Dunlosky, & Kornell, 2013; Bol & Hacker, 2012; Dunlosky & Thiede, 2013; Robey, Dougherty, & Butaccio, 2017; Serra & Metcalfe, 2009; Wiley, Griffin, & Thiede, 2005; Zabrucky, 2010), especialmente en los dominios asociados al entendimiento y la toma de decisiones sobre temas de interés público (Leiser & Shemesh, 2018; Lupia, 2016; Lusk, 2016; Poitras & Lajoie, 2013). En esta línea, Fernbach, Sloman et al. (2013) han encontrado que la IPE tiende ser menor cuando las personas son capaces de inhibir inferencias intuitivas incorrectas de primer orden. Sin embargo, el proceso metacognitivo que subyace a este efecto protector aún no ha sido investigado en detalle. Para esclarecer este mecanismo, es necesario analizar las relaciones entre los tipos de procesamiento (heurístico y analítico) y los juicios metacognitivos. Si la IPE aumenta cuando los juicios de comprensión causal se infieren de manera intuitiva a partir de claves heurísticas como la deseabilidad social del conocimiento, es posible que este sesgo pueda ser contrarrestado al crear condiciones que faciliten los procesamientos analíticos (e.g., disminuyendo la carga cognitiva o incentivando el razonamiento deliberado).

En este contexto, la presente investigación tiene un doble propósito: en primer lugar, determinar si las personas utilizan la deseabilidad social percibida del conocimiento sobre objetos y procesos específicos como clave inferencial para estimar su comprensión causal de los mismos. En segundo lugar, establecer si este efecto se produce en un nivel de

procesamiento analítico/controlado o heurístico/automático. En el primer capítulo de esta disertación, se revisan antecedentes teóricos y empíricos que sugieren que la percepción de deseabilidad social del conocimiento puede influir en los juicios metacognitivos relacionados con la IPE y que la capacidad para inhibir inferencias intuitivas incorrectas, en condiciones de procesamiento controlado, puede modular este efecto. En la primera parte, se introduce la explicación de la IPE basada en la formación de conceptos y teorías intuitivas, junto con hallazgos empíricos que demuestran la influencia de factores sociales y motivacionales sobre este sesgo metacognitivo. En la segunda sección, se examinan las investigaciones que evidencian el efecto de claves inferenciales sociales en juicios metacognitivos asociados a tareas de aprendizaje y memoria. Para explicar las similitudes entre estos efectos, se plantea la hipótesis de mecanismo único de los juicios metacognitivos, incluyendo la IPE, y se analizan algunas de sus implicaciones teóricas y empíricas. Con el propósito de dilucidar la arquitectura de este mecanismo, en la tercera sección se exponen las principales teorías de procesamiento dual y se examinan los modelos clave que describen sus posibles conexiones con los juicios metacognitivos. El capítulo concluye con la formulación de las preguntas de investigación, junto con las hipótesis que guiarán el diseño y análisis de resultados de los experimentos realizados.

Capítulo 1: Revisión Teórica

¿Por Qué Ocurre la Ilusión de Profundidad Explicativa?

En el paradigma experimental original de la IPE, Rozenblit y Keil (2002) pidieron a estudiantes universitarios estimar su nivel de comprensión del funcionamiento de una serie de dispositivos o fenómenos dispuestos en una lista (e.g., máquina de coser, cremallera, reloj de pulso, helicóptero, etc.). A continuación, se les pidió explicar de la manera más

detallada posible el funcionamiento de algunos de estos elementos, seleccionados al azar. Al concluir esta tarea, se les pidió estimar nuevamente qué tan bien creían comprender el mecanismo causal de cada artefacto o fenómeno. En la siguiente fase, los participantes respondieron una pregunta que indagaba sobre algún aspecto específico del funcionamiento del objeto (por ejemplo, cómo un helicóptero puede cambiar la dirección horizontal de su movimiento o por qué un reloj de pulso avanza más despacio cuando sus baterías empiezan a agotarse). Una vez más, los participantes calificaron su comprensión del objeto a la luz de su respuesta a la pregunta anterior. En la última fase, se les presentó a los participantes una descripción detallada del funcionamiento del dispositivo o fenómeno. Para verificar el efecto de esta información, se les pidió estimar su comprensión antes y después de leerla.

Utilizando distintas variaciones de este procedimiento en una serie de 12 estudios, Rozenblit y Keil encontraron que las estimaciones subjetivas de conocimiento sobre el funcionamiento de fenómenos o artefactos disminuyeron después de que las personas trataron de explicarlos. Esta sobrestimación inicial no fue producto de un exceso de confianza generalizado, puesto que su magnitud fue mayor en las explicaciones del mecanismo de funcionamiento de objetos y fenómenos que en otros tipos de conocimiento declarativo (e.g., hechos, narrativas, procedimientos). Para descartar que la disminución en la comprensión estimada fuese un artefacto producido por solicitar dos veces el mismo juicio, Rozenblit y Keil (estudio 5) pidieron a un grupo distinto de participantes evaluar en la misma escala las explicaciones escritas por los participantes en el paradigma de IPE. Estas evaluaciones independientes fueron más cercanas a las estimaciones que los participantes originales realizaron *después* de elaborar las explicaciones que a las reportadas antes de hacerlo. Adicionalmente, la magnitud de IPE se redujo, aunque no

desapareció, cuando se advirtió a los participantes que en algún punto de la tarea se les pediría escribir sus explicaciones y responder preguntas sobre algunos de los temas presentados. Esta advertencia fue comunicada a los participantes antes de que realizaran sus estimaciones iniciales de comprensión (Rozenblit & Keil, estudio 6). Con base en estos resultados, Rozenblit y Keil concluyeron que la IPE es un fenómeno robusto, en tanto que no se origina en una desconfianza generalizada en el propio conocimiento inducida por la tarea, sino en una reevaluación genuina del conocimiento causal. A partir del trabajo seminal de Rozenblit y Keil, la IPE ha sido demostrada en dominios distintos a los del funcionamiento de artefactos y fenómenos naturales, tales como el significado de palabras (Kominsky & Keil, 2014), propuestas de candidatos presidenciales (Alter et al., 2010), efectos potenciales de la implementación de políticas controversiales (Fernbach, Sloman, et al., 2013; Vitriol & Marsh, 2018; Voelkel et al., 2018), enfermedades mentales (Zeveney & Marsh, 2016) y recomendaciones comportamentales basadas en evidencia científica (Bromme et al., 2015).

Para explicar la ubicuidad de la IPE, Keil y sus colaboradores (Keil, 2003; Keil et al., 2004; Mills & Keil, 2004) plantean que este sesgo es un efecto colateral de la formación de teorías intuitivas o legas (*folk theories*). Estas teorías son representaciones articuladas de conceptos y leyes causales que prescriben cómo dichos conceptos se relacionan entre sí (Gerstenberg & Tenenbaum, 2017). Al igual que las teorías científicas, las teorías intuitivas permiten organizar el conocimiento en dominios específicos de fenómenos (e.g., el comportamiento de los sólidos y los fluidos, las funciones vitales de plantas y animales, los estados mentales de las personas, los cambios en las instituciones sociales, etc.), aunque surgen de manera informal y no como producto de la instrucción directa sobre conceptos y

principios científicos (Gelman & Legare, 2011; Gerstenberg & Tenenbaum, 2017; Keil, 2010; Ojalehto & Medin, 2015). Por lo general, las teorías intuitivas reúnen tres características: (a) forman un conjunto estable de creencias que son consistentes entre sí, (b) surgen espontáneamente en las personas, con relativa independencia del contexto cultural e histórico, y (c) no pueden modificarse fácilmente a partir de la evidencia (Shtulman, 2017). De acuerdo con Keil y sus colaboradores, una función de dichas teorías es la de "rastrear patrones causales en el mundo a un nivel de resolución que sea lo suficientemente grueso para permitirnos hacer progresos en varios dominios de la cognición sin quedar agobiados por los detalles del mecanismo en algún dominio" (Keil et al., 2004, p. 241). Por ejemplo, se sabe que los niños entre los dos y tres años usan con más frecuencia la textura como criterio de categorización en el dominio biológico que en el de los artefactos (Booth, Waxman, & Huang, 2005). Esto ocurre porque típicamente la textura está conectada causalmente con varias propiedades centrales de los seres vivos (e.g., fisiología, hábitat, dieta, nicho ecológico, origen filogenético, etc.) mientras que, en el caso de los artefactos, existen menos conexiones causales con otros rasgos (es menos probable que el mecanismo de funcionamiento de un dispositivo dependa de su textura). En la misma línea, se he encontrado que los niños desde la edad preescolar atribuyen experticia en el dominio de los artefactos a quienes los describen refiriéndose a su forma, tamaño y fragilidad, pero no a quienes mencionan su color, marcas superficiales o número exacto de partes. Cuando los niños evalúan descripciones de seres vivos, este patrón de preferencia se invierte (Keil, 2010, Experimento 3). Así, las teorías intuitivas permiten identificar de manera ágil la relevancia causal de una o varias propiedades en un dominio determinado. Esta habilidad es útil para establecer si un objeto novedoso pertenece a dicho dominio, incluso cuando no se dispone de conocimiento causal específico sobre el objeto en cuestión.

De manera importante, el reconocimiento de estos patrones generales no requiere construir representaciones exhaustivas de todos los procesos causales relacionados con las propiedades relevantes. Por ejemplo, reconocer la relevancia causal de la textura en el dominio de los seres vivos no equivale a saber cómo algunos animales pueden modificar la textura de su piel ante la presencia de un predador o una presa potencial. En este sentido, las personas creen comprender el funcionamiento de un objeto o fenómeno cuando logran identificar propiedades causales que se ajustan a los patrones típicos de algún dominio general (e.g., objetos físicos, entidades biológicas, agentes intencionales, grupos sociales, etc.; Gelman & Legare, 2011), aunque carezcan de representaciones explícitas de los mecanismos a través de los cuales estas conexiones causales se hacen efectivas (Johnson & Ahn, 2017).

Desde esta perspectiva, la IPE ocurre cuando la detección de propiedades y patrones causales relevantes es confundida con una representación exhaustiva de un mecanismo o proceso causal particular. Por ejemplo, desde edades tempranas, las personas entienden que los miembros de un mismo grupo social comparten intenciones (Noyes & Dunham, 2017) y tienden a comportarse de manera similar (Buttelmann, Zmyj, Daum, & Carpenter, 2013; Powell & Spelke, 2013), que las interacciones causales entre agentes intencionales no dependen esencialmente del contacto físico o de la contigüidad temporal de las acciones (Keil, 1995; Spelke, Phillips, & Woodward, 1995), que los objetos físicos pueden tener propiedades causales distintas a las de los materiales de los cuales están hechos (Rips & Hespos, 2015), que las funciones de objetos institucionales como el dinero dependen más de acuerdos sociales que de sus propiedades físicas (Noyes, Keil, & Dunham, 2018), y que una explicación mecanicista del funcionamiento de un objeto implica dar cuenta tanto de

las conexiones físicas como de las relaciones de dependencia causal que existen entre sus componentes (Lombrozo, 2009, 2010). Eventualmente, las personas pueden tener dificultades para diferenciar esta comprensión general del conocimiento en un dominio, y su comprensión particular de los mecanismos de funcionamiento de objetos concretos dentro de éste (Keil, 2012). Respaldando esta hipótesis, Alter et al. (2010) encontraron que la magnitud de la IPE varía cuando se induce a los participantes a adoptar un determinado estilo representacional (*construal*) de los objetos a explicar, por medio de la manipulación de la consigna de la tarea. En uno de sus experimentos, Alter y sus colaboradores le pidieron a un grupo de participantes explicar cómo funcionaban ciertos dispositivos (estilo abstracto), mientras que a otro grupo se le pidió explicar cómo las partes de tales dispositivos hacían posible su funcionamiento (estilo concreto). Se encontró que las personas estimaron de manera más precisa su habilidad para explicar los dispositivos al emplear un estilo representacional concreto de su funcionamiento y evidenciaron una IPE mayor al ser inducidas al estilo abstracto[1]. En consecuencia, las personas fueron más proclives a confundir el conocimiento de la función global de un dispositivo con la comprensión de su mecanismo concreto cuando adoptaron un estilo representacional abstracto. Rozenblit y Keil (2002) reportaron además que las personas son más propensas a sobrestimar su habilidad para explicar el funcionamiento de un objeto cuando saben cómo usarlo, creen que podrían extraer información relevante sobre él a partir de su manipulación, o son capaces de visualizar sus partes o reconocer sus nombres. En suma, la IPE tiende a ser mayor cuando los participantes tienen dificultades para distinguir el

[1]Aunque la magnitud de la IPE fue menor en el grupo que fue inducido al estilo representacional concreto, el efecto no desapareció completamente: entre el 30% y el 50% de los participantes en este grupo exhibieron la IPE en algún grado (Alter et al., 2010, experimentos 2, 3 y 6)

conocimiento representado del que potencialmente pueden adquirir a partir de sus interacciones en tiempo real con el objeto o fenómeno. Esto se debe a que la comprensión causal es estimada, entre otras cosas, a través de procesos que son influenciados por propiedades perceptuales de los objetos (e.g., visibilidad de sus componentes, facilidad de uso o manipulación, etc.) y por el reconocimiento de rasgos superficiales que parecen relevantes para su explicación (e.g., los nombres de sus partes o su marca comercial; Graeff, 2003).

Las teorías intuitivas no solo permiten reconocer las propiedades causales más relevantes de un objeto o proceso determinado; también son útiles para estimar la complejidad y el tipo de experticia que se requiere para entender mecanismos y fenómenos específicos (Kominsky, Zamm, & Keil, 2017). En lugar de recopilar información detallada sobre mecanismos concretos, las teorías intuitivas proporcionan representaciones esquemáticas que permiten identificar el tipo de experticia necesaria para conocer ciertos fenómenos y entender cómo este conocimiento se distribuye al interior de cada cultura. Empíricamente, se ha encontrado que adultos y niños en edad escolar usan de manera implícita información causal abstracta para inferir la probabilidad de que un experto en un área determinada (e.g., física) sea también experto en un fenómeno perteneciente a un dominio relativamente "cercano" (e.g., química) o "lejano" (e.g., psicología cognitiva; Keil, Stein, Webb, Billings, & Rozenblit, 2008). De manera similar, adultos y niños desde segundo grado elemental juzgan más probable que existan expertos en categorías basadas en rasgos causalmente relevantes (e.g., perros de cacería) que en otras creadas a partir de rasgos más superficiales (e.g., perros con collares rojos; Keil, 2010). Estos resultados indican que las personas reconocen desde edades tempranas las propiedades centrales en

distintos dominios explicativos y la distribución de la experticia sobre ellos o "división del trabajo cognitivo" dentro de su cultura (Aguiar, Stoess, & Taylor, 2012; Keil & Kominsky, 2015). Así, la IPE se incrementaría cuando la capacidad para mapear los aspectos esenciales y el tipo de experticia requerida para comprender un proceso causal es confundida con la habilidad para explicar dicho proceso de manera detallada.

Reconocer intuitivamente la división social del conocimiento en áreas de experticia permite a las personas hacer inferencias sobre lo que alguien debería saber al desempeñar un rol en un contexto social determinado. Por ejemplo, es probable que los habitantes de una ciudad consideren muy deseable que la mayoría de los votantes conozcan el programa de gobierno del candidato que eligen como alcalde, pero quizás juzguen menos importante que todos los conductores puedan explicar cómo funciona el sistema eléctrico de sus automóviles. Esto se debe a que la experticia sobre mecánica automotriz es adjudicada por la sociedad a un grupo específico de personas con base en ciertos criterios (e.g., formación académica, experiencia profesional, certificación oficial, etc.), mientras que, en las democracias participativas, la competencia para elegir a los gobernantes se adjudica por defecto a todos los ciudadanos aptos para votar (Brennan, 2016; Somin, 2016). En el siguiente apartado, se examinarán algunos hallazgos empíricos que muestran cómo las expectativas sobre lo que uno debería saber también influyen en el proceso de estimación del conocimiento causal.

Deseabilidad social del conocimiento e ilusión de profundidad explicativa. Fisher y Keil (2016) indagaron si la experticia en un dominio de conocimiento se asocia a una estimación más precisa de la habilidad para explicar objetos y fenómenos en dicho dominio. Para ello, estos investigadores distinguieron entre experticia pasiva y formal. La

experticia pasiva surge de “la exposición a través de la experiencia vital y la posición que uno ocupa en una sociedad o cultura” (p.1251; e.g., la asociada al género o la edad), mientras que la experticia formal es resultado del estudio sistemático, continuado y deliberado de un dominio específico, con hitos de logro definidos (e.g., grados académicos). A la luz de esta distinción, Fisher y Keil (Experimento 1) identificaron en un estudio preliminar diversos temas que las personas relacionaron con la experticia pasiva asociada al sexo (e.g., temas que los hombres creían poder explicar mejor que las mujeres y viceversa) y a la edad (e.g., temas que los jóvenes creían poder explicar mejor que los adultos mayores y viceversa). En el estudio principal, la edad y el sexo se utilizaron como indicadores de experticia pasiva y la obtención de un grado universitario como indicador de experticia formal. Los participantes fueron asignados a una de dos condiciones experimentales: en una de ellas, se les pidió explicar temas consistentes con su experticia pasiva (e.g., los hombres explicaron temas de pesca o mecánica automotriz), mientras que en la condición alternativa los temas no estuvieron relacionados con dicha experticia (e.g., los adultos mayores explicaron el funcionamiento de teléfonos inteligentes o de préstamos estudiantiles). Siguiendo el paradigma básico de la IPE, todos los participantes estimaron su comprensión de los temas antes y después de explicarlos. Se encontró que los participantes sin experticia formal (sin título universitario) sobrestimaron más su comprensión cuando los temas fueron consistentes con su experticia pasiva. Sin embargo, esta diferencia no se replicó en el grupo con experticia formal. Por otro lado, en estudios posteriores (Experimentos 2a y 2b), los participantes con experticia formal sobrestimaron más su comprensión en los temas consistentes con su área de especialidad que en los temas no consistentes con ella o relacionados con su experticia pasiva. Estos resultados indican que las personas con menos educación sobrestiman su habilidad para explicar temas

relacionados con su experticia pasiva, mientras que quienes tienen más educación lo hacen con temas vinculados con su experticia formal. Por consiguiente, uno de los factores desencadenantes de la IPE parece ser que las personas creen que deberían saber más sobre ciertos temas en razón a su tipo de experticia, con independencia de si ese conocimiento se obtuvo sistemáticamente a través de la experticia formal, o simplemente a través de su experiencia vital, es decir, experticia pasiva.

Utilizando un paradigma similar al de la IPE, Fisher y Keil (2014) pidieron a sus participantes estimar qué tan bien podrían justificar su posición sobre varios temas políticamente controversiales (e.g., la contribución de la actividad humana al calentamiento global, la universalidad del aseguramiento en salud en Estados Unidos, la relación entre el uso de teléfonos móviles y el cáncer, la legalización de la marihuana, etc.) antes y después de escribir sus argumentos sobre ellos. Adicionalmente, los participantes juzgaron qué tan importantes consideraban cada uno de estos temas. De manera similar a lo ocurrido con las explicaciones, se encontró una ilusión de justificación argumentativa (IJA): los participantes disminuyeron la estimación de su capacidad para justificar su opinión después de hacer explícitos sus argumentos. Por otra parte, la importancia que los participantes atribuyeron a cada tema se asoció positivamente con las estimaciones previas y posteriores de la capacidad para justificar argumentativamente su propio punto de vista. En otras palabras, los participantes evaluaron mejor la calidad de sus argumentos sobre los temas que consideraron más importantes. No obstante, este patrón no fue replicado cuando los argumentos fueron evaluados por un grupo distinto de personas. Por consiguiente, la IJA parece ser más acentuada en los temas que las personas consideran más relevantes o sobre los cuales creen que deberían tener buenos argumentos. Replicando este resultado en un

diseño entre-sujetos, Johnson, Tynan, Cuthbert y O'Quinn (2017) encontraron que el involucramiento emocional de los participantes con un tema (e.g., el debate sobre el control de armas en Estados Unidos) aumentó la sobrestimación de la calidad de los argumentos elaborados para defender sus posiciones sobre el mismo. Esta relación fue significativa incluso después de controlar el efecto de la complejidad y la familiaridad percibida con el tema. Estos resultados indican que la relevancia social y afectiva de un tema hace más probable que las personas sobrestimen su habilidad para argumentar sobre él.

Un resultado aparentemente incompatible con la evidencia presentada anteriormente es el hallazgo de Rozenblit y Keil (2002, estudio 11), quienes reportaron que la deseabilidad social percibida de ciertos tipos de conocimiento no estuvo asociada significativamente con la magnitud de la IPE en cada uno de ellos. No obstante, existe al menos una diferencia crucial entre los estudios de Rozenblit y Keil (2002) y Fisher y Keil (2014) que pueden dar cuenta de esta discrepancia: mientras que Rozenblit y Keil comparan las diferencias en la sobrestimación entre *tipos* de conocimiento (e.g., explicaciones, narrativas, hechos y procedimientos), Fisher y Keil contrastan la magnitud de la IJA entre temas dentro de un mismo tipo de conocimiento (i.e., debates políticos actuales) que difieren en la importancia que los participantes les atribuyen. Desde el punto de vista metodológico, los tipos de conocimiento pueden no ser las categorías de comparación más apropiadas para determinar si existe alguna relación entre la deseabilidad social del conocimiento y la IPE, ya que este último es un fenómeno característico de las explicaciones. En este sentido, puede resultar más informativo contrastar la magnitud de la IPE entre temas más o menos deseables dentro de dominios particulares, que entre

explicaciones y otros tipos de conocimiento. Explorar esta posibilidad es uno de los objetivos principales de esta disertación.

Los hallazgos anteriores indican que las estimaciones de la comprensión causal pueden verse afectadas no solo por la confusión entre el conocimiento general derivado de las teorías intuitivas y el conocimiento detallado de mecanismos específicos, sino también por juicios sobre factores contextuales que no se relacionan directamente con las propiedades causales de los objetos y fenómenos (e.g., qué tan importante es saber sobre ellos). En el siguiente apartado, se revisarán resultados análogos provenientes de la investigación sobre juicios metacognitivos en tareas de memoria y aprendizaje que pueden ayudar a develar los mecanismos psicológicos subyacentes a la IPE. De manera general, estos resultados sugieren que las personas no estiman su capacidad para recordar contenidos declarativos específicos con base en la inspección directa en la memoria a largo plazo, sino de manera indirecta a través de claves inferenciales como la familiaridad, la accesibilidad de la información relacionada, la fluidez del procesamiento (*processing fluency*), las creencias sobre lo que otros saben al respecto y la deseabilidad del conocimiento. A la luz de estos hallazgos, se propone la hipótesis de que la IPE no depende únicamente de la naturaleza simplificada de las representaciones causales, sino también de las inferencias metacognitivas a partir de las cuales se realizan los juicios de comprensión causal.

Deseabilidad Social del Conocimiento y *Feeling of Knowing* (FOK).

Memoria y juicios metacognitivos: ¿Cómo sabemos que sabemos? Los procesos cognitivos que tienen por objeto otros procesos cognitivos (es decir, la aplicación recursiva de las funciones cognitivas) se denominan procesos metacognitivos (Efklides, 2008;

Flavell, 1979; Shea et al., 2014). En general, estos mecanismos suponen la subordinación de procesos que ocurren en el nivel objeto (e.g., recordar cierta información previamente memorizada) a procesos que ocurren en el meta-nivel correspondiente (e.g., evaluar la dificultad para recordar la información solicitada). Los procesos metacognitivos pueden ser de naturaleza implícita o explícita y cumplen dos funciones principales: monitorear el estado y avance de los procesos cognitivos del nivel objeto, y controlar, regular e intervenir en el curso de dichos procesos. En los procesos de monitoreo, la información proveniente del nivel objeto es procesada en el meta-nivel, mientras que en los procesos de control la información fluye del meta-nivel al nivel objeto. Estas dos rutas de procesamiento crean un circuito de retroalimentación en el cual el resultado de los procesos de control se convierte en insumo fundamental para los procesos de monitoreo y, de manera recíproca, la efectividad del control depende de un monitoreo preciso (Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Koriat, 2015; Nelson & Narens, 1990; Schwarz, 2015; Serra & Metcalfe, 2009; Shea et al., 2014).

Los contenidos de los procesos metacognitivos pueden ser agrupados en tres categorías o "facetas" distintas: conocimientos, experiencias y habilidades (Efklides, 2008; Flavell, 1979). El conocimiento metacognitivo es el conjunto de representaciones declarativas relacionadas con el funcionamiento de los procesos cognitivos, incluyendo las propiedades de las tareas, de los agentes que las realizan, de las estrategias de resolución y de sus metas (e.g., el nivel de complejidad de las tareas, sus semejanzas y diferencias con otras tareas, la habilidad de uno mismo u otros para resolverlas, las condiciones en las cuales es conveniente aplicar una estrategia de resolución determinada y los objetivos que pueden ser alcanzados al implementar dicha estrategia). Por su parte, las experiencias

metacognitivas se refieren a las sensaciones y percepciones de los participantes sobre su desempeño actual o potencial en tareas específicas. Estas experiencias dan lugar a juicios metacognitivos, tales como la facilidad de aprendizaje o "*ease of learning"* (Leonesio & Nelson, 1990), el juicio de aprendizaje o "*judgment of learning*" (JOL, Bjork et al., 2013), la sensación de saber o "*feeling of knowing*" (FOK, Hart, 1965; Koriat, 2007), y los juicios de confianza o certeza en las respuestas generadas (Nelson & Narens, 1990; Thompson, 2009). Finalmente, las habilidades metacognitivas abarcan el conocimiento procedural relativo a la orientación, planeación, regulación e intervención sobre procesos cognitivos en curso (e.g., asignar más tiempo al estudio de la información que resulta más difícil de recordar). Por lo general, las habilidades metacognitivas son desplegadas en respuesta a la información proveniente del conocimiento y las experiencias cognitivas. Así, mientras que el conocimiento y las experiencias metacognitivas suelen estar asociadas con las funciones de monitoreo, las habilidades metacognitivas son esenciales para el control cognitivo (Efklides, 2008). Si bien los conocimientos, experiencias y habilidades metacognitivas pueden ser implementados en procesos simultáneos o secuenciales, la taxonomía descrita anteriormente y representada en la Figura 1 resulta útil para identificar los distintos tipos de información y de acciones implicadas en el monitoreo y control de la cognición.

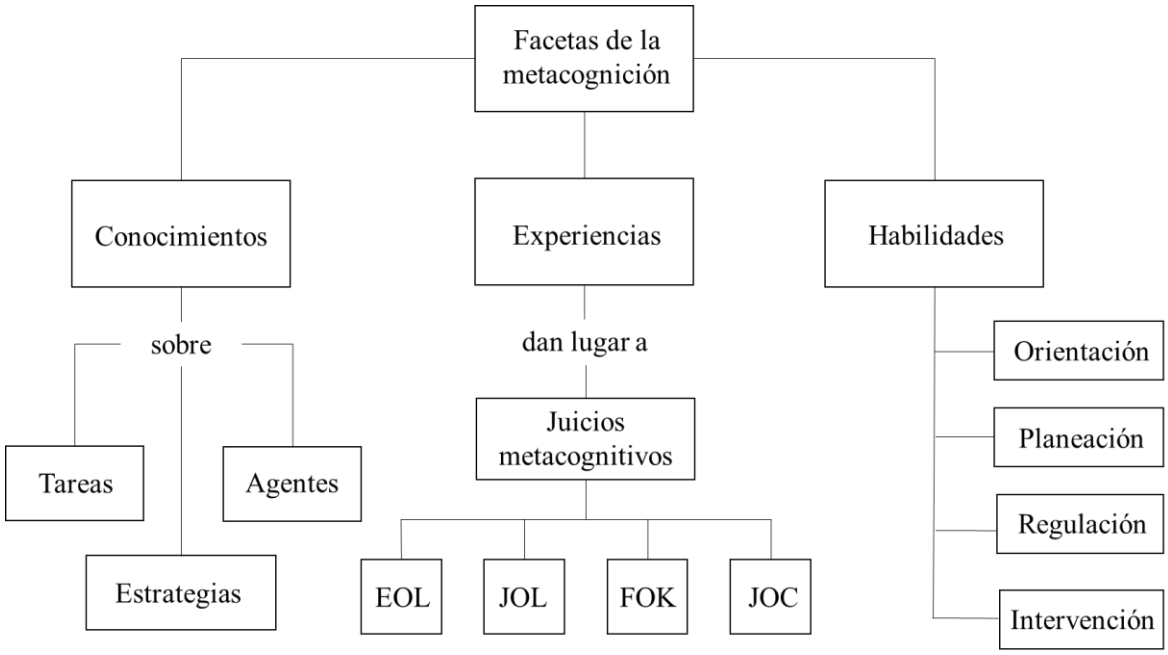


Figura 1. Clasificación de las facetas metacognitivas. Nota: EOL= Ease of learning; JOL= Judgment of learning; FOK= Feeling of knowing; JOC= Judgment of confidence. Fuente: elaboración propia, basado en Efklides (2008).

De manera consistente, la investigación empírica ha demostrado que los procesos metacognitivos no son simples rutinas accesorias o subproductos fenomenológicos, sino que influyen de manera crucial en la regulación de la actividad cognitiva (Dunlosky & Thiede, 2013; Metcalfe, 2009; Metcalfe & Finn, 2008; Veenman, Van Hout-Wolters, & Afflerbach, 2006). En particular, las relaciones entre experiencias metacognitivas y funcionamiento de la memoria han sido exploradas ampliamente en la psicología experimental y la neuropsicología (Schwarz, 2015; Thomas, Lee, & Hughes, 2016). Durante las últimas décadas, los estudios sobre juicios metacognitivos han examinado las fuentes de información que las personas utilizan para juzgar si un contenido particular puede o no ser aprendido o recordado. Específicamente, se ha investigado de manera extensiva sobre el FOK que experimentan las personas cuando creen tener cierto

conocimiento almacenado en su memoria y la capacidad de evocarlo o reconocerlo en el futuro, aun cuando no puedan hacerlo en el presente (Koriat, 2007). En el paradigma clásico de FOK (Hart, 1965), los participantes tratan de recordar información general (e.g., ¿cuál es el continente más grande?) o memorizada en una tarea previa a solicitud del experimentador (e.g., una palabra pareada con otra usada como clave). Si el recuerdo no es exitoso, se le pide al participante estimar la probabilidad de que pueda identificar acertadamente la respuesta correcta si le fuese presentada (es decir, su FOK). A continuación, se evalúa la precisión del juicio anterior solicitando al participante que identifique la respuesta correcta entre varias alternativas. Por último, los participantes expresan su nivel de confianza en que la opción seleccionada por ellos es la correcta. Es importante destacar que el FOK es similar al juicio de comprensión causal que se realiza antes de elaborar las explicaciones en el paradigma de IPE, ya que en ambos casos las personas juzgan de manera *prospectiva* si disponen o no de cierto conocimiento que aún no ha sido recuperado. Sin embargo, mientras que en el paradigma de IPE este conocimiento corresponde a la explicación causal de un tema específico, en el de FOK puede referirse a cualquier tipo de conocimiento declarativo.

El paradigma experimental de "recuerdo-juicio-reconocimiento" (*recall-judgment-recognition*), que predominó en la investigación temprana sobre FOK, no ha estado exento de críticas metodológicas (e.g., Schwartz, Boduroglu, & Tekcan, 2015; Schwartz & Metcalfe, 1994). Por esta razón, en décadas más recientes se han desarrollado paradigmas alternativos. Reder y Ritter (1992), por ejemplo, presentaron a sus participantes diferentes problemas aritméticos (e.g., 17 x 23). A continuación, les pidieron elegir una de dos estrategias para resolver el problema: recordar la respuesta correcta o realizar el cálculo

correspondiente. Los participantes debían tomar esta decisión en un tiempo máximo de 850 milisegundos, de manera que no contaran con el tiempo suficiente para calcular o recordar la respuesta correcta. En este sentido, la elección de la estrategia de recuerdo fue interpretada como indicadora de un FOK alto. De acuerdo con el tipo de problema y la estrategia seleccionada, los participantes disponían de un tiempo preestablecido para dar su respuesta en voz alta. Posteriormente, la respuesta correcta fue presentada durante 2 segundos a manera de retroalimentación. A diferencia del procedimiento de Hart (1965), la estimación del FOK en este arreglo experimental se realizaba *antes* del intento de recuerdo de la respuesta correcta, de manera que no fuese afectada por dicho intento. Tanto en este paradigma como en el de Hart, se encontró que los indicadores de FOK se asociaron al reconocimiento/recuerdo exitoso de las respuestas correctas. En general, los estudios sobre FOK han empleado diferentes tipos de contenido (e.g., semántico, autobiográfico, testimonial, etc.) y de juicio solicitado (e.g., ítem por ítem vs. estimación global; para una revisión, ver Schwartz & Metcalfe, 1994). Con frecuencia se ha encontrado que la correlación entre el FOK y el éxito posterior en el reconocimiento o evocación es positiva y moderada (Nelson, 1984; Schwarz, 2015), aunque esta asociación puede llegar a ser nula o negativa bajo ciertas condiciones (Koriat, 1995, 2012).

¿Acceso directo o claves heurísticas?: Mecanismos metacognitivos del FOK.

Para explicar por qué el FOK predice el desempeño posterior en tareas de reconocimiento, Hart (1967) propuso una hipótesis basada en el acceso a trazas de memoria. Según esta aproximación teórica, las personas cuentan con un sistema interno que monitorea la activación de los contenidos en la memoria de largo plazo. En consecuencia, la magnitud del FOK depende del nivel de activación de la representación de la respuesta correcta

cuando la pregunta es formulada o la clave asociada es presentada. Si esta activación supera cierto umbral pero no es suficiente para recordar la respuesta correcta, las personas juzgarán que podrán reconocerla entre varias alternativas. Para contrastar empíricamente esta hipótesis, Yaniv y Meyer (1987) presentaron a sus participantes definiciones de palabras poco comunes y les pidieron tratar de recordar dichas palabras. Cuando el recuerdo no fue exitoso, se les pidió formular el FOK respectivo. Después de esto, los participantes resolvieron una tarea de decisión léxica en la que debían juzgar si la palabra presentada era real o una pseudopalabra. En esta tarea se incluyeron las palabras correspondientes a las definiciones presentadas al inicio. Yaniv y Meyer encontraron que quienes hicieron juicios altos de FOK con respecto a una palabra inusual también fueron capaces de reconocerla correctamente en la tarea de decisión léxica. Este resultado podría ser interpretado como una evidencia a favor del modelo de Hart, ya que la activación de la palabra generada durante el intento de recuerdo previo al FOK produciría un efecto de facilitación en su reconocimiento posterior. No obstante, Connor, Balota y Neely (1992) demostraron que este efecto de facilitación ocurría incluso cuando la tarea de decisión léxica se realizaba una semana antes que la de FOK, descartando así que el efecto fuese generado por el intento fallido de evocación. En consecuencia, Connor y sus colaboradores adjudicaron este efecto de facilitación a la familiaridad con el dominio de conocimiento al cual pertenecía la pregunta, más que a la fuerza de la activación de palabras específicas. En la misma línea, Nelson y Narens (1990) reportaron que los juicios de FOK correlacionaron de manera más robusta con las estimaciones subjetivas del número de intentos exitosos de recuerdo realizados que con las frecuencias reales de dichos intentos. Ambos hallazgos apuntan a que los juicios metacognitivos como el FOK son generados a partir de atributos o claves inferenciales relacionadas con los elementos a recordar (e.g., su familiaridad o la

accesibilidad de la información relacionada), convirtiendo en poco parsimoniosa la hipótesis de un sistema específico con acceso privilegiado a las trazas de dichos elementos en la memoria de largo plazo.

Consolidando esta visión de los juicios metacognitivos como procesos inferenciales, en décadas más recientes se ha demostrado empíricamente que múltiples claves informativas afectan dichos juicios. Por ejemplo, se han establecido diversos factores que afectan la estimación de la probabilidad de recordar en el futuro información previamente estudiada o "*judgment of learning*" (JOL; para una revisión, ver Bjork et al., 2013). Así, la fluidez de procesamiento es entendida como "una sensación subjetiva de facilidad o dificultad asociada con algún tipo de procesamiento mental" (Graf, Mayer, & Landwher, 2018, p. 394). Se ha encontrado que la baja fluidez de procesamiento de un contenido particular, normalmente se asocia a mayores tiempos de aprendizaje, menores estimaciones de JOL y una menor probabilidad de recordarlo en el futuro (Koriat et. al, 2006). De manera similar, se ha mostrado que diversos factores superficiales afectan la estimación metacognitiva, aumentando la discrepancia de esta estimación y el desempeño real. Por ejemplo, el tamaño de las letras de las palabras presentadas durante los ensayos de aprendizaje afecta la fluidez y modula la estimación del JOL pero no tiene un efecto significativo en la probabilidad de recordarlas (Rhodes & Castel, 2008). En la misma línea, cuando se induce a los participantes a creer que el color de las palabras afecta su procesamiento, la magnitud del JOL cambia en función de esta variable a pesar de que el desempeño en las pruebas de memoria no se ve alterado (Mueller & Dunlosky, 2017). Por otro lado, algunos factores asociados con el desempeño no son ponderados correctamente en la estimación del JOL. Por ejemplo, aunque el tiempo de presentación de los estímulos

durante la fase de aprendizaje y el intervalo de retención correlacionan positivamente con la probabilidad del recuerdo, la magnitud del JOL es poco sensible a estas relaciones (Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004). Consistentemente, la influencia de ciertas claves inferenciales en la estimación del JOL puede ser modificada a partir de la experiencia con la tarea (Tauber & Rhodes, 2010) o del tipo de juicio solicitado (comprensión vs. predicción del desempeño; Rawson, Dunlosky, & McDonald, 2002). En conjunto, estos resultados sugieren que los juicios metacognitivos son influenciados por factores que pueden o no estar asociados con el desempeño posterior. En consecuencia, la precisión de dichos juicios depende críticamente del valor predictivo y del peso asignado a las claves inferenciales utilizadas durante su estimación.

Al igual que en el caso del JOL, se ha encontrado que el FOK no se deduce a partir de una única clave heurística, sino que múltiples factores pueden afectarlo en mayor o menor medida, dependiendo del momento en el cual se solicita y de las restricciones de la tarea (Koriat, 2015; Thomas et al., 2016). Por ejemplo, mientras que la familiaridad percibida de los ítems puede aumentar el FOK antes de la fase de evocación, la facilidad para recordar información relacionada con ellos (accesibilidad) tiene una mayor influencia cuando dicha evocación no es exitosa. Concretamente, Koriat y Levy-Sadot (2001) encontraron que la relación entre accesibilidad de la información y magnitud del FOK depende de la familiaridad previa con los ítems: cuando la familiaridad con la pregunta antes del intento es alta y la evocación no es exitosa, la accesibilidad y el FOK tiende a estar más asociadas; en contraste, cuando la familiaridad es baja, la asociación entre accesibilidad y FOK es menor. Estos hallazgos respaldan la hipótesis de que el mecanismo de estimación del FOK no tiene las propiedades de un módulo encapsulado de acceso

directo a los contenidos almacenados en la memoria (Burke, MacKay, Worthley, & Wade, 1991; Hart 1965, 1967), sino que depende de procesos inferenciales que operan a partir de claves generadas con anterioridad o de manera simultánea con el proceso de evocación (Koriat, 1997, 2015; Thomas et al., 2016).

Desde una perspectiva más amplia, Koriat y sus colaboradores (Koriat & Levy-Sadot, 1999; Koriat, Nussinson, Bless, & Shaked, 2008) postulan que los procesos de estimación de los juicios metacognitivos pueden estar basados o bien en información explícita (e.g., las creencias relacionadas con el propio desempeño o el de otros) o en sensaciones y experiencias asociadas a la implementación de los procesos cognitivos (e.g., accesibilidad, familiaridad, facilidad o fluidez de procesamiento, etc.). Cuando los juicios metacognitivos se derivan de información declarativa, estos contenidos son recobrados de la memoria de largo plazo de manera consciente y deliberada. En contraste, los juicios basados en experiencias dependen de claves heurísticas que son generadas de forma implícita en el transcurso de los procesos cognitivos y cuya activación no puede ser controlada de manera consciente. Empíricamente, se ha encontrado evidencia robusta de que ambos tipos de procesos no solo afectan la magnitud del FOK (Costermans et al., 1992; Hosey, Peynircioğlu, & Rabinovitz, 2009; Koriat, 1995), sino también la de otros juicios metacognitivos (Ackerman & Thompson, 2015; Bjork et al., 2013; Dunlosky & Tauber, 2014; Koriat et al., 2004). En suma, la investigación reciente en metamemoria ha mostrado que los juicios metacognitivos no son generados a partir de un sistema de monitoreo con acceso directo a los contenidos de la memoria a largo plazo, sino a partir de procesos inferenciales basados en rasgos superficiales de los estímulos presentados y en claves relacionadas con el proceso de recuerdo. En la siguiente sección, se mostrará cómo algunos

de estos procesos inferenciales pueden estar basados en juicios de comparación social y expectativas sobre el propio desempeño.

La influencia de claves sociales en el FOK. Algunas investigaciones sobre metamemoria sugieren que ciertas claves sociales pueden ser usadas en los procesos inferenciales que dan lugar a los juicios de FOK. Con el propósito de estudiar las fuentes de información declarativa que pueden afectar el FOK, Costermans et al. (1992, Experimento 2), examinaron la relación entre este tipo de juicios metacognitivos y diversas claves de origen social. En la primera fase del estudio, los participantes recibieron una lista de 160 preguntas sobre conocimientos generales. Cuando la pregunta fue respondida, los participantes estimaron el nivel de confianza en su respuesta; en caso de que declararan no saber la respuesta, se les solicitaba la estimación del FOK. A continuación, se les pidió completar un cuestionario con respecto a las preguntas de la primera fase, el cual indagaba cuatro aspectos: (a) el recuerdo de las circunstancias en las cuales el contenido fue memorizado, (b) qué tan reciente era este recuerdo, (c) la familiaridad del participante con el dominio y (d) la proporción estimada de personas que podrían saber la respuesta a la pregunta. Por último, se llevó a cabo una prueba de reconocimiento para determinar la precisión de las estimaciones de confianza y del FOK. Se encontró que la confianza fue un mejor predictor de la precisión de la respuesta que el FOK. Adicionalmente, tanto la familiaridad con la pregunta como la cantidad estimada de personas que sabían la respuesta correcta se asociaron positivamente con la magnitud del FOK. Estos resultados son consistentes con los obtenidos por Koriat (2008, 2012), los cuales indican que la correlación entre confianza y precisión del FOK es positiva para las preguntas en las que más personas logran identificar la respuesta correcta; en contraste, dicha correlación es

negativa cuando existe un consenso similar sobre una respuesta errónea. En este sentido, los juicios de FOK parecen ser sensibles a las claves que hacen que una opción de respuesta sea seleccionada por la mayoría, independientemente de si se trata o no de la respuesta correcta.

Las expectativas sobre el desempeño de otros también afectan la magnitud del FOK y de los juicios de comprensión de textos. De Carvalho Filho y Yuzawa (2001) encontraron que los juicios de FOK en estudiantes universitarios con niveles bajos de habilidad metacognitiva aumentaron al informarles que otros estudiantes (ficticios) lograron desempeños altos con las mismas preguntas en una tarea similar y disminuyeron cuando se les informó que dichos estudiantes obtuvieron desempeños bajos. En la misma línea, Zhao y Linderholm (2010) reportaron que las personas utilizaron información sobre el rendimiento de pares ficticios como puntos de anclaje (*anchors*) al estimar su propio desempeño en pruebas de comprensión de textos leídos previamente, incluso cuando dichas estimaciones fueron realizadas después de responder las pruebas. Por otra parte, Karabenick (1996) encontró que las personas expresaron más dudas sobre su comprensión de mensajes videograbados cuando se les comunicó que otros participantes formularon preguntas sobre dichos materiales. Más recientemente, Zeveney y Marsh (2016) utilizaron el paradigma de IPE para mostrar que las personas tienden a sobrestimar más su comprensión causal de algunos trastornos mentales cuando creen que la sociedad tiene un conocimiento sofisticado sobre ellos y que este se distribuye de manera equitativa entre la población. En conjunto, los resultados anteriores apoyan la hipótesis de que tanto el FOK como los juicios de comprensión no son producto de una búsqueda exhaustiva de contenidos en la memoria, sino que son estimados a partir de diversas claves inferenciales,

algunas de las cuales son de origen social (e.g., la distribución del conocimiento entre grupos de pares, el nivel de consenso, la cantidad estimada de personas que conocen la respuesta correcta, el desempeño de otros en la misma tarea, etc.).

El efecto de la deseabilidad social de saber la respuesta correcta a una pregunta no ha sido investigado directamente desde el paradigma de FOK, probablemente debido a las dificultades metodológicas que implica su manipulación experimental. En particular, el conocimiento sobre un tema específico se hace más deseable como resultado de múltiples interacciones sociales en contextos diversos (e.g., culturales, académicos, profesionales, etc.), los cuales son difíciles de reproducir en condiciones de laboratorio a partir de interacciones de corta duración. Adicionalmente, la investigación sobre sesgos de respuesta en pruebas de auto-reporte de personalidad e inteligencia ha mostrado que la deseabilidad social no es un constructo unidimensional, sino que agrupa distintas formas de comportamiento tendientes a exagerar los atributos positivos y ocultar los negativos (Paulhus, 2017). En particular, Paulhus y Trapnell (2008) proponen dos componentes principales de las respuestas socialmente deseables: por un lado, aquellas orientadas a la auto-promoción (*self-enhancement*) y el auto-engaño (*self-deception*), basadas en la motivación para elaborar un autoconcepto que incluya rasgos de agencia personal como dominancia, proactividad, auto-eficacia, regulación emocional, etc. El segundo componente se relaciona con el manejo de las percepciones que otros tienen sobre uno mismo (*impression management*), a partir de la mención de características prosociales y la negación de comportamientos que pueden ser interpretados como inapropiados o inmorales (e.g., mentir, tener emociones negativas sobre otros, irrespetar normas sociales cuando no se es vigilado, etc.). Paulhus (2002) ha encontrado que estas dos dimensiones de la

deseabilidad social se asocian a distintos rasgos de personalidad y pueden implicar mayores o menores grados de esfuerzo consciente para sesgar positivamente las descripciones propias. Debido a que los juicios sobre la propia comprensión son por definición medidas de auto-reporte, es necesario controlar posibles sesgos de respuesta no relacionados con los procesos metacognitivos que generan dichos juicios. La dificultad para encontrar medidas objetivas que permitan distinguir las respuestas distorsionadas de los reportes genuinos en dominios específicos ha llevado a la creación de técnicas psicométricas como las pruebas de *over-claiming* (Paulhus, Harms, Bruce, & Lysy, 2003), las cuales serán descritas en el Capítulo 2 y se utilizarán en algunos experimentos de la presente disertación como medida de control de sesgos potenciales de auto-promoción.

Desde los inicios de la investigación sobre juicios metacognitivos, Gruneberg, Monks y Sykes (1977) postularon la deseabilidad social como una de las posibles claves a partir de las cuales las personas infieren la probabilidad de disponer de cierto conocimiento en su memoria. Esta idea es consistente con algunas investigaciones previas que destacan la influencia de motivaciones sociales en los juicios metacognitivos y las estimaciones sobre el propio desempeño (Nisbett & Wilson, 1977). Por ejemplo, Soderstrom y McCabe (2011) encontraron que estudiantes universitarios sobrestimaron su habilidad para recordar pares de palabras no relacionadas semánticamente cuando su evocación exitosa fue mejor recompensada. De manera similar, las predicciones de los estudiantes universitarios sobre sus calificaciones después de realizar un examen se vieron afectadas por el nivel de desempeño que deseaban tener (Serra & DeMarree, 2016). Por su parte, Ehrlinger y Dunning (2003, Experimento 2) aplicaron una prueba idéntica a dos grupos de estudiantes de psicología; los participantes en uno de los grupos recibieron la indicación de que la

prueba medía una habilidad que ellos previamente habían afirmado tener y que consideraron más deseable (razonamiento abstracto), mientras que al grupo restante se le informó que el test evaluaba una habilidad en la que ellos declararon ser menos competentes y juzgaron menos deseable (programación de computadores). A pesar de que los resultados de ambos grupos fueron similares, los participantes creyeron haber obtenido puntajes más altos cuando se les indicó que la prueba medía la habilidad que consideraron más deseable y en la que afirmaron ser más competentes.

Si bien en los paradigmas experimentales de FOK se han utilizado tradicionalmente conocimientos declarativos distintos a las explicaciones (e.g., sucesos, fechas, nombres, lugares, etc.), es posible que la IPE involucre mecanismos inferenciales comunes a otros juicios metacognitivos. Por ejemplo, Fisher y Keil (2016) sugieren que las personas con experticia formal exhiben una IPE mayor en temas relacionados con sus áreas, porque subestiman la tasa de olvido del conocimiento causal especializado que adquirieron durante su formación académica. Por otra parte, en la investigación sobre metamemoria se ha encontrado que este "sesgo de estabilidad" del recuerdo se asocia a la sobrestimación de la capacidad de recordar información en el futuro (Kornell & Bjork, 2009). En ambos casos, las personas parecen estimar su conocimiento a partir de creencias, sensaciones y experiencias relacionadas, omitiendo el escrutinio del contenido representado y la reflexión explícita sobre los factores que pueden facilitar o inhibir su evocación. Estos hallazgos muestran claras similitudes entre los procesos de estimación de la comprensión causal y los de otros juicios metacognitivos, respaldando la hipótesis de que dichos juicios son generados a partir de un mecanismo único (Nelson & Narens, 1990; Dunlosky & Tauber, 2014). Las implicaciones teóricas de esta hipótesis serán examinadas a continuación.

La hipótesis de mecanismo único de los juicios metacognitivos. Dunlosky & Tauber (2014) postulan la existencia de procesos comunes a todos los juicios metacognitivos y analizan las posibilidades de capturar este mecanismo general en un marco conceptual unificado. Según estos investigadores, la plausibilidad de dicho mecanismo se sustenta en dos rasgos comunes a todos los juicios metacognitivos: (a) la dependencia de claves inferenciales explícitas o implícitas y (b) la naturaleza analítica/controlada o no analítica/automática del procesamiento de dichas claves. Una consecuencia empírica de esta hipótesis de mecanismo único o "isomecanismo" es que una clave inferencial que afecte cierto juicio metacognitivo (como el FOK) tendrá más probabilidades de influir en otros juicios (como el asociado a la IPE). Confirmando esta predicción, se ha encontrado que la accesibilidad percibida de una palabra (i.e., el número estimado de palabras compuestas que la incluyen) se asocia a estimaciones más altas de la facilidad para aprenderlas y de la probabilidad de recordarlas en el futuro (Ackerman & Beller, 2017, Experimento 3).

De otra parte, una clave inferencial determinada podría afectar los juicios metacognitivos a través de diferentes rutas de procesamiento. Por un lado, dichos juicios pueden depender del procesamiento implícito de una clave inferencial basada en experiencias metacognitivas (como en el caso de la familiaridad o la fluidez). Dunlosky y Tauber (2014, p. 454) representan este escenario por medio de la ecuación 1, en la cual *j* representa el juicio metacognitivo, *h* denota el procesamiento no analítico o heurístico, y c_i se refiere a una clave inferencial implícita.

$$j = h(c_i) \quad (1)$$

De manera alternativa, el juicio puede ser generado a partir de una clave informativa procesada explícitamente (e.g., las creencias sobre el desempeño de otros o sobre el funcionamiento de la memoria). Esta posibilidad se representa en la ecuación 2, en la en la cual *j* representa el juicio metacognitivo, *a* denota el procesamiento analítico y c_e se refiere a una clave inferencial explícita.

$$j = a(c_e) \tag{2}$$

Estas dos formas esquemáticas dan lugar a múltiples combinaciones posibles de procesamientos "híbridos". Por ejemplo, el efecto de la activación de una clave implícita puede mitigarse a través del procesamiento analítico (e.g., ajustar la estimación de la probabilidad de recordar en el largo plazo un contenido recién aprendido con el fin de controlar el efecto de la familiaridad y la accesibilidad actuales). Este caso se representa en la ecuación 3, en donde los corchetes indican la aplicación de un procesamiento analítico al resultado de una inferencia heurística[2].

$$j = a[h(c_i)] \tag{3}$$

Entender los mecanismos que conectan las inferencias analíticas y no analíticas en el procesamiento de una clave determinada es crucial para diseñar intervenciones que ayuden a las personas a mejorar la precisión de sus juicios metacognitivos y la efectividad del control cognitivo subsiguiente. Sin embargo, la evidencia disponible en la actualidad no es suficiente para dilucidar dichos mecanismos (Dunlosky & Tauber, 2014). En la presente

[2] Existen múltiples posibilidades de procesamiento híbrido, tales como el escrutinio analítico de los resultados de inferencias heurísticas contradictorias, la aplicación recursiva de inferencias analíticas o la competencia dinámica entre los dos sistemas, al estilo de los modelos de doble ruta en el procesamiento lingüístico (Coltheart, Curtis, Atkins, & Haller, 1993). No obstante, estas alternativas no serán examinadas en detalle en esta sección, ya que su propósito es ilustrar la posibilidad de interacción entre procesamientos analíticos y heurísticos en la generación de los juicios metacognitivos.

disertación, se indaga si la deseabilidad social del conocimiento es utilizada como clave inferencial al momento de estimar la comprensión subjetiva de mecanismos físicos y sociales, en el marco de procesos metacognitivos análogos a los que hacen posible el FOK. Si la hipótesis del mecanismo único de los juicios metacognitivos es correcta, será necesario establecer de qué manera estos procesos inferenciales se acoplan a la arquitectura general del sistema de monitoreo y control cognitivo. En particular, se requiere determinar si estas inferencias están relacionadas con el tipo de procesamiento (analítico o heurístico) que las personas realizan al momento de formular sus juicios metacognitivos. Para responder a este interrogante, en la siguiente sección se introducirán las teorías de procesamiento dual y se examinará evidencia empírica que muestra cómo el tipo de procesamiento está relacionado con los procesos metacognitivos involucrados en la estimación de la comprensión causal.

Teorías de Procesamiento Dual y Metacognición

Encrucijadas inferenciales. Desde la alegoría platónica de la razón como auriga de las pasiones hasta las teorías psicoanalíticas, la distinción entre pensamiento reflexivo e intuitivo ha sido un tema recurrente en la tradición filosófica occidental (Frankish & Evans, 2009). Durante las últimas décadas del siglo XX, se desarrolló una familia de teorías psicológicas que postulaban la existencia de dos formas alternativas de procesar la información y de realizar inferencias. Dichas teorías surgieron de manera relativamente independiente en el contexto de la investigación experimental sobre atención (Schneider & Shiffrin, 1977), razonamiento (Stanovich & West, 2000; Wason & Evans, 1975), toma de decisiones (Barbey & Sloman, 2007; Kahneman, 2011; Kahneman & Frederick, 2005), aprendizaje (Reber & Squire, 1994; Sun, Slusarz, & Terry, 2005), formación de actitudes

(Chaiken, 1980; Petty & Cacciopo, 1986), cognición social (Devine, 1989; Epstein, Pacini, Denes-Raj, & Heier, 1996; Strack & Deutsch, 2015) y desarrollo cognitivo (Barrouillet, 2011).

En líneas generales, los modelos de procesos duales plantean dos rutas o sistemas de procesamiento de la información: una de naturaleza intuitiva, autónoma, rápida, implícita, asociativa, con componentes afectivos, no accesible a la introspección, basada en claves contextuales y con bajo consumo de recursos cognitivos (tipo 1), y otra de carácter analítico, secuencial, lento, consciente, controlado, de alto poder computacional, asociada a la inteligencia fluida, basada en reglas explícitas y pensamiento hipotético (tipo 2; Evans, 2008; Kahneman, 2003; Sloman, 1996; Stanovich, 2011). Aunque esta diferenciación basada en los atributos más frecuentes de cada tipo de procesamiento es compatible con la mayoría de las teorías de procesos duales, no ha estado exenta de críticas (Melnikoff & Bargh, 2018). Específicamente, el hecho de que algunos fenómenos posean atributos asociados a los dos sistemas, como la flexibilidad en el aprendizaje implícito (Eitam, Hassin, & Schul, 2008) o el razonamiento lógico rápido (Newman, Gibb, & Thompson, 2017) ha sido interpretado por algunos autores como una evidencia contra la validez teórica de esta distinción (Keren & Schul, 2009). Con el fin de resolver esta aparente ambigüedad conceptual, Evans y Stanovich (2013) postularon dos propiedades esenciales de los procesos duales: (a) el procesamiento de tipo 1 opera de manera autónoma con respecto a las funciones ejecutivas de propósito general, en tanto que no requiere atención controlada para iniciarse e involucra una participación mínima del sistema ejecutivo central (Baddeley, 2007); (b) el procesamiento de tipo 2 implica el pensamiento hipotético y el "desacoplamiento cognitivo" (*cognitive decoupling*), entendido como la habilidad para

distinguir entre las representaciones de la realidad y las representaciones de escenarios o "mundos posibles" que son utilizadas en el contexto de las simulaciones mentales (Stanovich, 2011). De manera consistente con esta definición, De Neys y sus colegas (2006; 2006a; Johnson, Tubau, & De Neys, 2014) han demostrado que la manipulación experimental de la sobrecarga de la memoria de trabajo a través de tareas concurrentes afecta sustancialmente el desempeño en la tarea de selección de Wason (1966), problemas de tasas base (Kahneman & Tversky, 1973) y el CRT (*Cognitive Reflection Test*, Frederick, 2005). Estas tareas experimentales han sido tradicionalmente utilizadas en la investigación sobre procesos duales en razonamiento, debido a que logran activar una respuesta intuitiva errónea que debe ser inhibida a través del procesamiento controlado para obtener la respuesta correcta (Pennycook, 2017; Pennycook & Thompson, 2016; Thompson, Evans, & Campbell, 2013). Por ejemplo, uno de los tres ítems de la versión clásica del CRT es el conocido problema del bate y la pelota: "Un bate y una pelota cuestan $1.10 en total. El bate cuesta $1.00 más que la pelota. ¿Cuánto cuesta la pelota?" (Frederick, 2005, p. 26). Para resolver correctamente este problema, los participantes deben percatarse de que la respuesta intuitiva ($0.10) es errónea (e.g., $1.10 + $0.10=$1.20). El cálculo aritmético requerido para deducir la respuesta correcta puede ser representado en la ecuación 4, donde *x* es el precio de la pelota:

$$x + (x + \$1.00) = \$1.10 \quad (4)$$

Al plantearse en estos términos, es fácil notar que la respuesta correcta al problema es $0.05. En este sentido, la razón por la cual muchas personas responden equivocadamente a estas tareas radica más en la dificultad para inspeccionar y descartar la respuesta intuitiva antes de aceptarla que en problemas con el cómputo necesario para inferir la respuesta

correcta. ¿Cómo interactúan los tipos de procesamiento 1 y 2 en la resolución correcta o incorrecta de este tipo de problemas? En el siguiente apartado se examinarán los modelos de procesos duales que abordan esta pregunta.

Modelos de procesamiento dual. La explicación de cómo interactúan los tipos de procesamiento 1 y 2 cuando las personas resuelven tareas de razonamiento ha sido unas de las fuentes principales de debate en las teorías de procesos duales. Al respecto, se han propuesto tres clases de modelos de procesamiento: serial, paralelo e híbrido. La Figura 2 ilustra los diferentes patrones temporales de activación de los sistemas 1 y 2, postulados en cada uno de estos modelos. El más influyente de ellos ha sido el denominado modelo serial o "intervencionista por defecto" (*default-interventionist*, Kahneman 2011; Evans & Stanovich, 2013). Dicho modelo asume que el procesamiento tipo 1 es la respuesta predeterminada del sistema cognitivo ante las tareas inferenciales. Si se detecta alguna inconsistencia en esta respuesta inicial, el procesamiento tipo 2 puede activarse para realizar un escrutinio analítico del problema a resolver (ver Figura 2, panel A). No obstante, con frecuencia las personas tienen dificultades para detectar que la respuesta intuitiva es errónea, debido a su tendencia a minimizar el consumo de recursos ejecutivos durante las tareas inferenciales (Toplak, West, & Stanovich, 2014). Esto explicaría la alta tasa de errores que se presenta en las tareas en las cuales la respuesta normativa no coincide con la respuesta más intuitiva. Por su parte, los modelos paralelos (Epstein, 1994; Sloman, 1996) plantean que ambos tipos de procesamiento entran en operación desde la fase inicial del proceso inferencial; en caso de que exista un conflicto entre las respuestas generadas por los dos sistemas, ambos se activarán de manera simultánea hasta alcanzar un estado coherente, sin que uno de ellos logre suprimir completamente el funcionamiento del otro

(ver Figura 2, panel B). Finalmente, el modelo híbrido o de intuición lógica (De Neys, 2017) propone que el procesamiento tipo 1 puede dar lugar a la activación de dos clases de respuestas intuitivas: una heurística y otra basada en conocimiento lógico y probabilístico de tipo implícito. Si las dos respuestas son incompatibles, el conflicto entre ellas puede ser detectado sin que se lleven a cabo procesamientos tipo 2 (ver Figura 2, panel C). El modelo de intuición lógica combina así propiedades de los modelos anteriores, ya que asume que el procesamiento tipo 2 es posterior al de tipo 1 (como el modelo serial) y que la activación de más de una respuesta intuitiva puede ocurrir desde el inicio del proceso (al igual que los modelos paralelos). Como evidencia empírica a favor de este modelo, se ha encontrado que las personas recuerdan mejor las versiones de las tareas en las cuales la respuesta heurística entra en conflicto con la respuesta normativa, incluso cuando su memoria de trabajo es sobrecargada a través de una tarea concurrente (Franssens & De Neys, 2009). En un estudio relacionado, Bago y De Neys (2017) encontraron que el 66% de quienes respondieron correctamente este tipo de tareas lograron hacerlo en condiciones de tiempo limitado y de sobrecarga de la memoria de trabajo, es decir, cuando solo era posible realizar procesamientos tipo 1. Ambos hallazgos sugieren que las respuestas normativas también pueden ser generadas a partir del procesamiento intuitivo, sin que se requiera la intervención “correctiva” del procesamiento tipo 2.

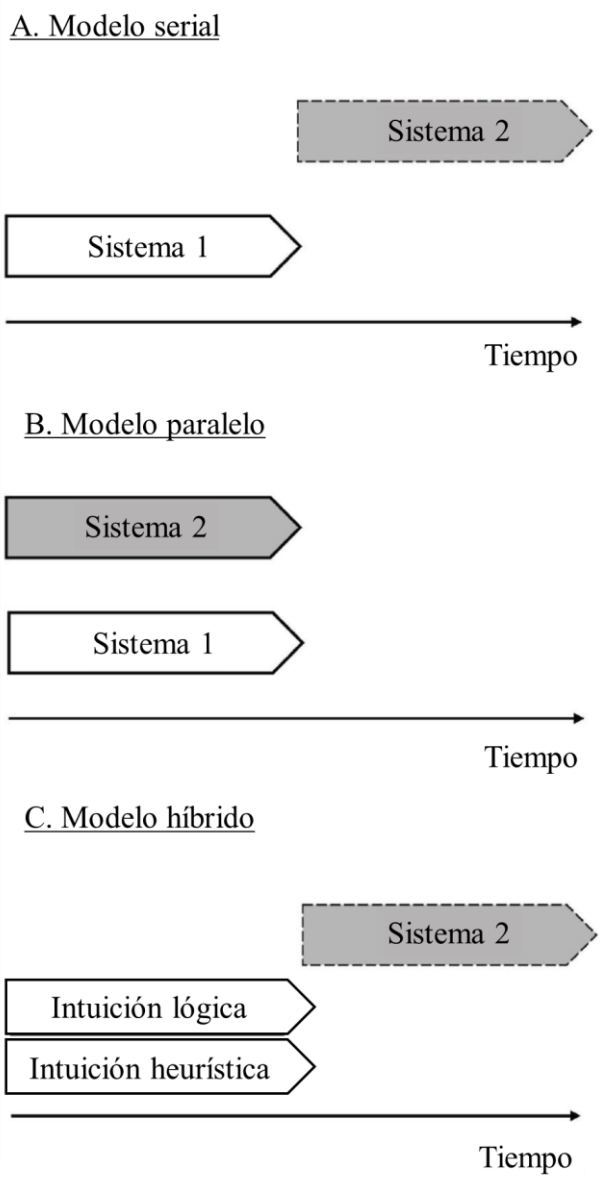


Figura 2. Secuencias de activación de los sistemas 1 y 2 en los modelos de procesamiento serial, paralelo e híbrido. Adaptado de "Bias, conflict, and fast logic" de W. De Neys, en W. De Neys (Ed.), *Dual process theory 2.0* (p.49), 2017, London, UK: Routledge. El eje horizontal representa el tiempo; las líneas discontinuas en los modelos A y C indican que la activación del sistema 2 depende del resultado de los procesamientos previos de tipo 1.

Una consecuencia teórica importante del modelo híbrido es que cuestiona la idea del procesamiento tipo 2 como un mecanismo de respaldo ante las limitaciones del procesamiento intuitivo, propuesta por el modelo serial. En otras palabras, si los conflictos entre respuestas correctas y heurísticas pueden ser detectados de manera intuitiva, ¿por qué las personas necesitan invertir sus recursos cognitivos de propósito general iniciando procesos analíticos? Si el mecanismo de detección de conflicto entre respuestas intuitivas fuese totalmente eficiente, ciertamente muchas de las funciones del procesamiento tipo 2 serían redundantes. Sin embargo, Pennycook, Fugelsang y Koehler (2015b, Experimentos 2-4) encontraron que solo un pequeño porcentaje de participantes en sus estudios fue capaz de detectar consistentemente el conflicto entre respuestas intuitivas en problemas de tasas base cuando el contraste entre las proporciones descritas en ellos fue moderado. Para explicar este hallazgo, Pennycook y sus colaboradores plantean la existencia de diferencias individuales en la capacidad de las personas para iniciar el procesamiento tipo 2 ante la detección del conflicto entre respuestas intuitivas y formulan un modelo híbrido de tres etapas (ver Figura 3): Inicialmente, una o varias respuestas (e.g., R_1, $R_2 \ldots R_n$) son generadas a partir del procesamiento tipo 1 (i.e., etapa de generación de respuestas). Estas respuestas no emergen de manera simultánea, sino que algunas se activan de manera más rápida o "fluida" que otras. En la segunda etapa, un mecanismo de monitoreo metacognitivo se encarga de detectar posibles conflictos entre dichas respuestas y activar el procesamiento tipo 2. La detección de un conflicto puede dar lugar a dos tipos de procesos analíticos: o bien la búsqueda de razones explícitas que justifiquen la elección de la respuesta generada de manera más fluida, o bien la inhibición de esta respuesta y el análisis de posibles alternativas. En caso de que no se detecte algún conflicto, la respuesta intuitiva procesada con mayor fluidez será seleccionada de manera automática. En la última fase, se formula la

respuesta obtenida a partir de los procesos efectuados. En este modelo tripartito, la rapidez o fluidez de procesamiento de las respuestas intuitivas durante la fase inicial del proceso inferencial aumenta la probabilidad de detectar posibles conflictos entre ellas. Por lo tanto, cuando dos respuestas inconsistentes se activan de manera casi simultánea, es más probable que el conflicto sea detectado y que se inicie un proceso de escrutinio analítico. En una propuesta teórica relacionada, Thompson (2009) señala que los procesamientos tipo 1 generan dos tipos de resultados: uno cognitivo (i.e., la respuesta derivada del proceso) y una sensación metacognitiva sobre la certeza o corrección del proceso (*feeling of rightness* o FOR). En este modelo, el mecanismo de control metacognitivo responsable de la activación del procesamiento tipo 2 depende de la información suministrada por el FOR: mientras que un FOR bajo se asocia una alta probabilidad de detección de conflictos entre respuestas y tiempos de respuesta más prolongados, un FOR alto se relaciona con una disminución de dicha probabilidad y respuestas más rápidas (Thompson & Johnson, 2014; Thompson, Prowse-Turner, & Pennycook, 2011). En este sentido, los dos modelos postulan la existencia de claves heurísticas (la fluidez de procesamiento y la sensación de certeza) asociadas con la eventual detección de inconsistencias entre respuestas intuitivas y la activación posterior del procesamiento tipo 2.

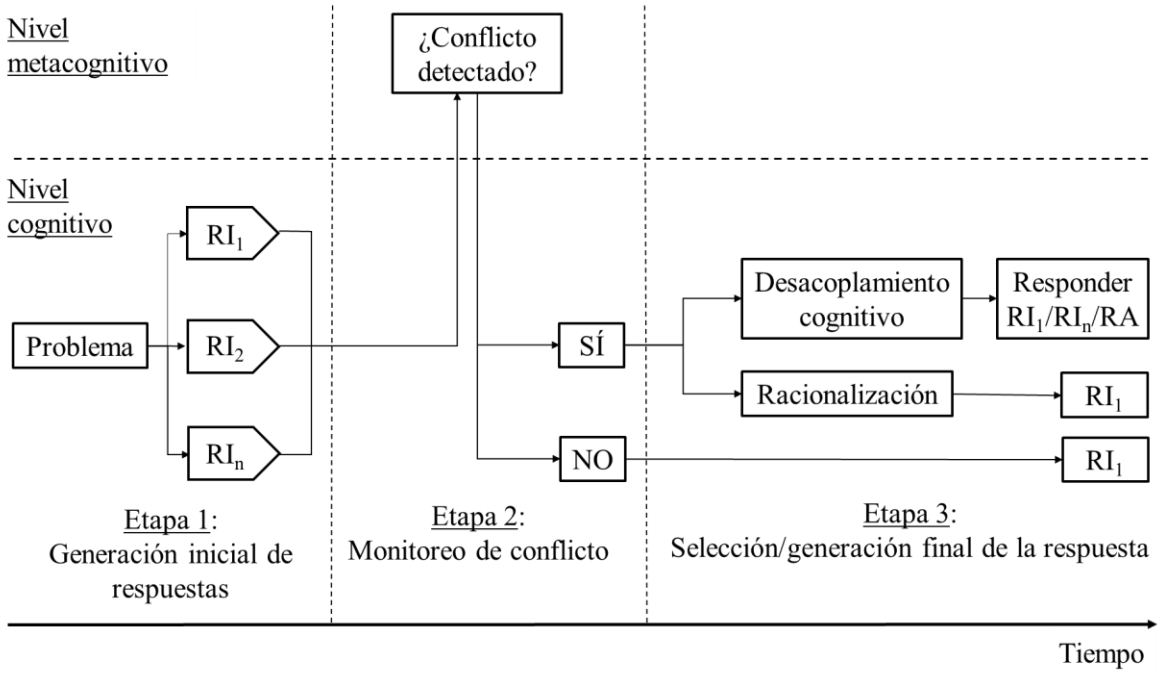


Figura 3. Modelo híbrido de tres etapas. RI=Respuesta intuitiva; RA= Respuesta alternativa; Las líneas discontinuas verticales dividen las fases del proceso inferencial; La línea discontinua horizontal separa los niveles cognitivo y metacognitivo. Adaptado de "What make us think? A three-stage dual-process model of analytic engagement" de G. Pennycook, J. A. Fugelsang y D. J. Koehler, 2015b, *Cognitive Psychology*, *80*, p. 39. Ver la descripción del modelo en el texto principal.

De manera análoga a los modelos que explican los juicios metacognitivos en tareas de memoria, los modelos de interacción entre procesos duales sugieren que dichos juicios están conectados a través de un mecanismo de monitoreo que puede ser influenciado por creencias, sensaciones y experiencias metacognitivas (Ackerman & Thompson, 2017). A partir de esta coincidencia teórica, es posible proponer que la capacidad para inhibir el efecto de ciertas claves heurísticas (e.g., accesibilidad, familiaridad, fluidez de procesamiento), tanto en la resolución de tareas de razonamiento como en la estimación del

conocimiento sobre temas específicos, depende de mecanismos de monitoreo que suponen altas demandas de memoria de trabajo y recursos atencionales. De este modo, es posible que la arquitectura propuesta por los modelos interactivos en el nivel cognitivo (i.e., procesos metacognitivos que controlan la activación del procesamiento analítico con base en información declarativa y claves heurísticas) se replique en el meta-nivel (i.e., procesos que controlan la generación de juicios metacognitivos a partir del monitoreo de conocimientos y experiencias metacognitivas). En el siguiente apartado, se revisarán algunos antecedentes teóricos y empíricos que apuntan en esta dirección, vinculando la habilidad para iniciar procesamientos analíticos con los juicios metacognitivos.

La dualidad de los procesos metacognitivos. La capacidad para procesar información de manera analítica y controlada no solo tiene efectos en el desempeño en tareas de razonamiento, sino que también se asocia a comportamientos y disposiciones más generales. Por ejemplo, se ha encontrado que las personas que tienen más dificultades para llevar a cabo procesamientos tipo 2 tienden a asumir creencias religiosas más tradicionales (Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012), exhibir orientaciones morales más asociadas a la identificación y el arraigo a un grupo social (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014), cooperar más en interacciones económicas (Rand et al., 2014), abstenerse de votar (Pennycook & Rand, 2018) y aceptar noticias falsas difundidas en redes sociales (Pennycook & Rand, 2017). Adicionalmente, el desempeño en el CRT ha demostrado ser un predictor más robusto de la comprensión de conceptos y teorías científicas que la habilidad para analizar información estadística o el conocimiento sobre los métodos de la ciencia (Shtulman & McCallum, 2014). En este sentido, es posible que las inferencias tipo 2 no solo se relacionen con una mejor comprensión de conceptos más

complejos, sino también con la capacidad de monitorear mejor la calidad de dicha comprensión. En esta misma línea, Mata, Ferreira y Sherman (2013) plantean que las personas capaces de realizar procesamientos tipo 2 exhiben cierta "ventaja metacognitiva", ya que la toma de conciencia de la respuesta normativa y de las alternativas intuitivas incorrectas, les permite estimar de manera más precisa la probabilidad de que ellos mismos y otros respondan correctamente. Respaldando esta idea, se ha encontrado que las personas que obtienen puntajes altos en pruebas de inteligencia verbal y no verbal tienden a realizar juicios de confianza más precisos sobre la corrección de sus respuestas en tareas de razonamiento y toma de decisiones (Jackson & Kleitman, 2014; Jackson, Kleitman, Howie, & Stankov, 2016) y que las personas con puntajes más bajos en el CRT son más susceptibles de sobrestimar su desempeño en dichas tareas (Pennycook, Ross, Koehler, & Fugelsang, 2017). Por su parte, Fernbach, Sloman et al. (2013, Experimento 4) reportan que las personas con puntajes altos en el CRT tienden a sobrestimar menos su comprensión causal de productos comerciales, exhibiendo una menor IPE que quienes obtienen puntajes bajos en esta prueba. Este hallazgo es especialmente relevante para los objetivos de esta disertación porque demuestra que la habilidad para realizar procesamientos tipo 2 se asocia a una mayor precisión en los juicios de comprensión causal que subyacen a la IPE. Como se verá a continuación, entender el vínculo entre la precisión de los juicios metacognitivos y el procesamiento analítico, requiere explorar en detalle las variedades del procesamiento metacognitivo.

Durante la última década, se ha desarrollado un extenso debate teórico sobre el alcance del concepto de metacognición y sus relaciones con la conciencia, el lenguaje y la teoría de la mente (e.g., Carruthers, 2009; 2017; Metcalfe, 2008; Metcalfe & Son, 2012;

Proust, 2012, 2013; Zohar & Ben David, 2009). En este contexto, algunos investigadores han diferenciado dos niveles de procesamiento metacognitivo, en clara analogía con las teorías de procesamiento dual en razonamiento y toma de decisiones (Proust, 2012, 2013; Shea et al., 2014). En particular, Arango-Muñoz (2011) ha distinguido entre metacognición de bajo y alto nivel. Los procesos de bajo nivel son aquellos que requieren únicamente de la capacidad de tener sensaciones metacognitivas y construir representaciones no conceptuales de ciertas propiedades de los procesos cognitivos (e.g., facilidad o rapidez), utilizándolas para el control de estos últimos. Por otra parte, los procesos de alto nivel implican además la capacidad para representar los estados mentales de otros agentes y disponer de una base de conocimiento explícito sobre el funcionamiento cognitivo[3]. Por su naturaleza declarativa, los procesos metacognitivos de alto nivel se asocian al procesamiento tipo 2, mientras que los de bajo nivel están relacionados con las inferencias tipo 1. Adicionalmente, Arango-Muñoz (2011) plantea tres posibles formas en las cuales ambos tipos de procesos pueden interactuar: (a) la activación de procesos de alto nivel como producto de la detección de un conflicto entre procesos de bajo nivel; (b) la facilitación de experiencias metacognitivas de bajo nivel como resultado de la activación de creencias o conocimientos declarativos de nivel alto (e.g., creer que dos palabras están asociadas aumenta la magnitud de los juicios de aprendizaje de dichas palabras; Mueller, Tauber, & Dunlosky, 2013); (c) la inhibición de los juicios generados a partir de procesos de bajo nivel debido a la influencia de procesos de alto nivel (e.g., la activación del

[3] La distinción entre metacognición de alto y bajo nivel es compatible con la formulada por Koriat et al. (2008) entre juicios metacognitivos basados en información y juicios basados en claves heurísticas (ver la sección “Deseabilidad Social del Conocimiento y Feeling of Knowing”) y con la hipótesis del mecanismo único de los juicios metacognitivos (Dunlosky & Tauber, 2014).

concepto de olvido genera estimaciones más conservadoras del JOL; Finn, 2008; Koriat et al., 2004).

Siguiendo esta interpretación dual de los procesos metacognitivos, la deseabilidad social del conocimiento podría afectar la estimación del conocimiento causal al menos de dos maneras alternativas: (a) de forma implícita, a través de un proceso ascendente (*bottom-up*) que genere la sensación metacognitiva de tener un conocimiento más detallado sobre los temas más deseables, o (b) de manera explícita, a partir un proceso descendente (*top-down*) en el cual la creencia en que se dispone del conocimiento socialmente deseable afecte la experiencia metacognitiva ligada al juicio de comprensión causal.

¿Cómo establecer empíricamente cuál de estos dos procesos conecta la deseabilidad social del conocimiento con los juicios de comprensión causal? Una alternativa metodológica consiste en utilizar el paradigma de doble respuesta, creado en el contexto de la investigación sobre procesos metacognitivos y razonamiento (Thompson, Prowse-Turner, & Pennycook, 2011). Este paradigma experimental se divide en dos fases: durante la fase inicial, se presenta a los participantes un problema de razonamiento y se les pide resolverlo tan rápido como puedan, de forma que no dispongan del tiempo suficiente para realizar inferencias analíticas. En la segunda fase, los participantes disponen de tiempo ilimitado para reflexionar sobre el problema y decidir si modifican o no su respuesta inicial. Para prevenir que las personas realicen inferencias analíticas "rápidas" antes de generar la primera respuesta, Bago y De Neys (2017) agregaron a la fase inicial una tarea concurrente de memoria visoespacial que aumenta la demanda de recursos ejecutivos centrales. Una ventaja importante de este paradigma es que permite distinguir claramente las respuestas intuitivas de las respuestas reflexivas y deliberadas.

Utilizando un principio análogo, el paradigma de IPE podría ser adaptado para controlar el tipo de procesamiento que las personas llevan a cabo al formular la estimación inicial de comprensión causal: si dicho juicio se realiza en un tiempo restringido y en condiciones de sobrecarga ejecutiva, es probable que en su generación únicamente intervengan procesamientos tipo 1. En contraste, si este juicio es realizado sin restricciones de tiempo o de recursos centrales, tanto los procesamientos tipo 1 como tipo 2 podrán confluir en su estimación. Por consiguiente, si se comparan los efectos de la deseabilidad social sobre la IPE entre estas dos condiciones, será posible determinar el nivel del procesamiento metacognitivo en el cual esta clave inferencial tiene un efecto predominante.

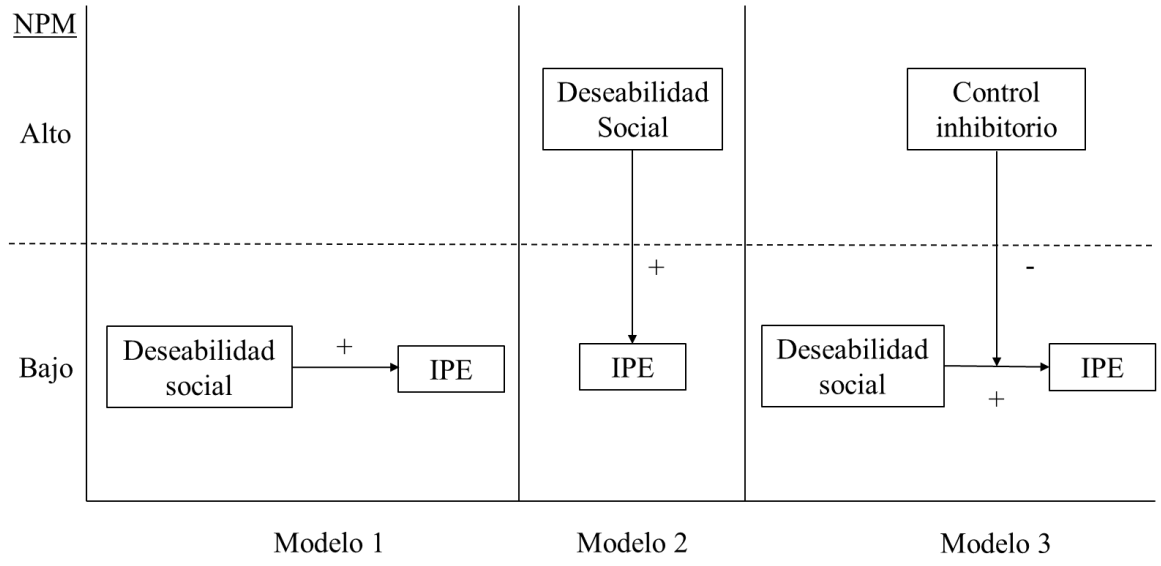


Figura 4. Modelos de procesamiento metacognitivo del efecto de la deseabilidad social del conocimiento en la IPE. Nota: Los signos representan la dirección de la asociación. Las flechas que conectan dos variables representan regresiones. La flecha que apunta hacia una línea indica una moderación. NPM = Nivel de procesamiento metacognitivo. Fuente: elaboración propia.

La Figura 4 ilustra tres posibles mecanismos metacognitivos de influencia de la deseabilidad social del conocimiento en la IPE. Si el efecto de esta clave inferencial ocurre exclusivamente a través de procesos de bajo nivel (modelo 1), se esperaría que la demanda de recursos atencionales y de memoria de trabajo no altere la influencia de la deseabilidad social en la IPE. Por otra parte, si las creencias explícitas sobre la deseabilidad social de ciertos conocimientos afectan los procesos metacognitivos de nivel bajo (modelo 2), la sobrecarga en los procesos ejecutivos centrales disminuirá el efecto de esta clave inferencial sobre el proceso de estimación del conocimiento causal. Finalmente, si los procesos de control metacognitivo en el nivel alto pueden inhibir el uso de la deseabilidad social como clave inferencial de nivel bajo, la sobrecarga ejecutiva fortalecerá la asociación entre deseabilidad social e IPE (modelo 3).

Desde un punto de vista más amplio, la manipulación de la sobrecarga del sistema ejecutivo central puede ser aplicada como estrategia metodológica para determinar el nivel de procesamiento en el cual influyen distintas claves inferenciales asociadas a los juicios metacognitivos. Esta información puede resultar útil en el diseño de intervenciones que minimicen los efectos perjudiciales de los sesgos metacognitivos en el aprendizaje de nuevos contenidos y en la formación de opiniones y actitudes sobre temas de interés público. En general, si un contenido explícito (e.g., una creencia) distorsiona los juicios metacognitivos al activarse en el nivel de procesamiento alto, facilitar que dichos juicios se generen de manera rápida e intuitiva podría ser una estrategia óptima para mitigar los sesgos metacognitivos. En contraste, si se trata de una clave heurística que afecta el nivel de procesamiento bajo (e.g., una sensación metacognitiva), disminuir la sobrecarga

ejecutiva durante la generación de los juicios metacognitivos sería una alternativa más efectiva para aumentar su precisión.

Adicionalmente, la comprensión de estos mecanismos permitiría entender cómo funcionan algunas estrategias de reducción de la IPE que ya han sido implementadas exitosamente. Por ejemplo, se sabe que la reflexión explícita sobre la complejidad causal de ciertos mecanismos permite hacer estimaciones más precisas de la comprensión de aquellos que son considerados altamente complejos (Johnson, Murphy, & Messer, 2016). Este tipo de reflexión podría disminuir la IPE a través de dos rutas distintas de procesamiento: o bien generando sensaciones metacognitivas asociadas a estimaciones más conservadoras de la comprensión causal (e.g., la sensación de no saber o "*feeling of not knowing*"; Glucksberg & McCloskey, 1981; Kolers & Palef, 1976; Liu, Su, Xu, & Chan, 2007), o bien activando inferencias analíticas que corrijan el sesgo inducido por ciertas claves heurísticas generadas durante la estimación de la comprensión causal. Si esta estrategia opera en el nivel bajo de procesamiento metacognitivo, podría ser eficaz incluso en condiciones de tiempo limitado y sobrecarga ejecutiva. Si, por el contrario, la estrategia afecta los procesos metacognitivos de nivel alto, su efectividad dependerá de que se realice en condiciones favorables al procesamiento tipo 2. Este ejemplo ilustra la importancia de entender los mecanismos inferenciales de los juicios metacognitivos para desarrollar intervenciones que mejoren la precisión del monitoreo y la eficacia del control cognitivo.

Formulación del Problema

Preguntas de investigación e hipótesis. La evidencia analizada hasta el momento sugiere una conexión entre deseabilidad social, tipos de procesamiento y procesos metacognitivos. En la primera parte de la revisión teórica, se ha mostrado que la IPE

depende no solo de la naturaleza simplificada de las representaciones de los mecanismos causales, sino también de claves heurísticas como lo que las personas creen saber en función de su experticia y de la importancia atribuida a temas específicos. En la segunda sección, se han resaltado algunas similitudes entre los juicios de FOK y las estimaciones de comprensión causal que se realizan en el paradigma de IPE. Adicionalmente, se ha mostrado cómo las creencias sobre el propio desempeño, la información sobre el desempeño de otros y la motivación para responder correctamente en una tarea influyen tanto en el FOK como en otros juicios metacognitivos. Con base en estos antecedentes, es plausible proponer que la deseabilidad social del conocimiento sobre temas específicos se asociará a una mayor sobrestimación de su comprensión causal, del mismo modo que algunas claves sociales similares generan sesgos en otros juicios metacognitivos.

Por otra parte, en la tercera sección se presentó evidencia de que las personas capaces de inhibir inferencias intuitivas erróneas tienden a estimar de manera más precisa su comprensión causal. Adicionalmente, se examinaron algunas propuestas teóricas que sugieren que la interacción entre procesamientos analíticos y heurísticos puede producirse tanto en el nivel cognitivo como en el metacognitivo. En este sentido, es posible que la capacidad para realizar procesamientos analíticos permita controlar el efecto de claves heurísticas como la deseabilidad social en los juicios metacognitivos. Así, las manipulaciones experimentales que alteren la capacidad de llevar a cabo inferencias tipo 2 afectarán la precisión de la estimación del propio conocimiento, de acuerdo con los modelos presentados en la Figura 3. En consecuencia, el efecto de la deseabilidad social del conocimiento sobre la IPE podría ser modulado por la manipulación de los recursos

ejecutivos disponibles a través de una tarea concurrente y la limitación del tiempo con el que cuentan los participantes para realizar sus juicios de comprensión causal.

En síntesis, la presente investigación se propone establecer si la deseabilidad social del conocimiento sobre un tema y la magnitud de la IPE están relacionadas, y si la capacidad para realizar procesamientos tipo 2 modera esta relación. Específicamente, en esta investigación se intenta dar respuesta a dos preguntas empíricas relacionadas, a través de cinco experimentos. En correspondencia con estas preguntas, es posible formular dos hipótesis específicas a la luz de la literatura revisada anteriormente:

1. ¿Existe una asociación entre la deseabilidad social percibida del conocimiento y la IPE? (Experimentos 1, 2 y 2a)

   Hipótesis 1: La deseabilidad social percibida del conocimiento sobre un tema se asociará positivamente con la sobrestimación de su comprensión, es decir, con la magnitud de la IPE.
2. ¿La asociación entre deseabilidad social percibida e IPE es moderada por la capacidad para realizar inferencias tipo 2? (Experimentos 3 y 4)

   Hipótesis 2: El procesamiento heurístico (tipo 1) se asociará a una mayor sobrestimación de la comprensión de temas más deseables socialmente (una mayor IPE), mientras que el procesamiento analítico (tipo 2) se asociará a una estimación más precisa de la comprensión sobre los temas cuyo conocimiento sea más deseable socialmente (ver Figura 4, modelo 3).

La Figura 5 resume las dos hipótesis formuladas anteriormente (líneas punteadas), junto con las asociaciones que ya han sido reportadas en la literatura y mencionadas en la revisión teórica (líneas continuas).

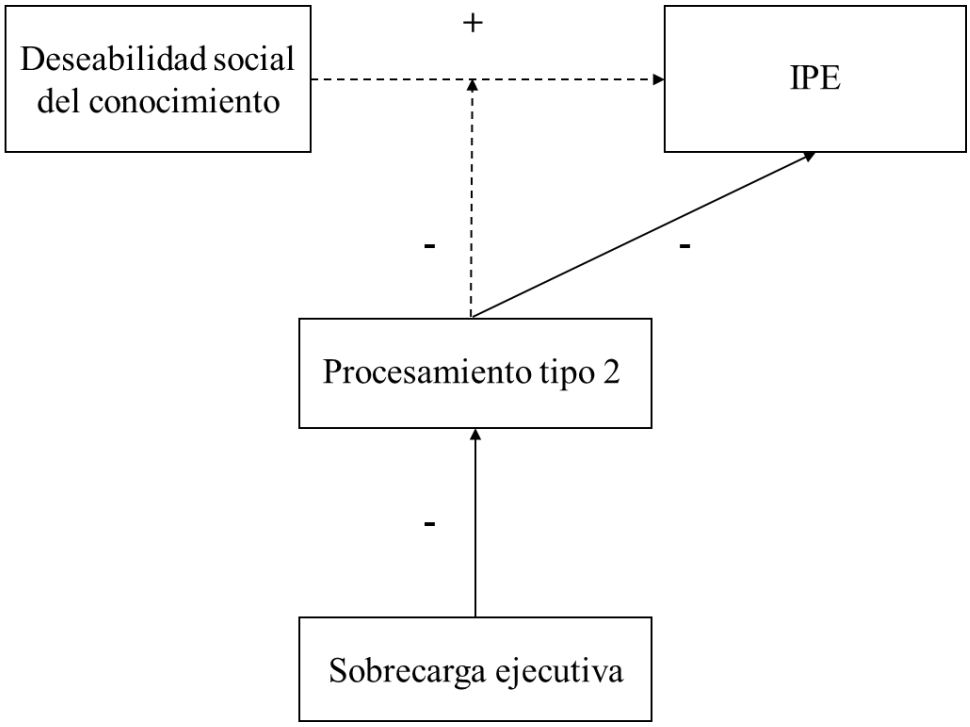


Figura 5. Modelo de las relaciones entre las variables de interés. Nota: Los signos representan el sentido de la relación; las flechas que conectan dos variables representan regresiones; las flechas que apuntan hacia una línea indican moderaciones. Fuente: elaboración propia.

Capítulo 2: Método

Resumen de los Experimentos

Los cinco estudios que hacen parte de esta disertación pueden ser divididos en dos grupos: en el primero se encuentran los Experimentos 1, 2 y 2a, cuyo objetivo fue examinar la asociación entre la IPE y la deseabilidad social del conocimiento sobre temas específicos. En el Experimento 1, se utilizó una manipulación intra-sujeto para examinar las diferencias en la magnitud de la IPE entre temas de alta y baja deseabilidad social, pertenecientes a tres dominios distintos (histórico, económico y dispositivos). Para obtener evidencia más

robusta del efecto de la deseabilidad social en la IPE, en los Experimentos 2 y 2a la asignación de temas históricos de alta y baja deseabilidad social fue manipulada en un diseño entre-sujetos. Adicionalmente, se verificó que este efecto no fuese producido por la baja familiaridad con los temas menos deseables o por un sesgo de auto-promoción. El propósito del segundo grupo de Experimentos (3 y 4) fue determinar si el efecto de la deseabilidad social sobre la IPE depende del tipo de procesamiento que realizan los participantes. Concretamente, en el Experimento 3 se analizaron las relaciones entre la deseabilidad social asignada a diferentes propuestas legislativas, la capacidad para inhibir inferencias intuitivas (i.e., medida a través del CRT) y la magnitud de la IPE. En el Experimento 4, se manipuló directamente el tiempo disponible para realizar las estimaciones de comprensión y el nivel de sobrecarga ejecutiva con el fin de establecer si el tipo de procesamiento afecta la relación entre deseabilidad social del conocimiento e IPE. En este capítulo, se describirá en detalle la metodología empleada en cada uno de estos estudios.

Tamaños de Muestra

En esta serie de experimentos participaron 1.077 personas en total, distribuidas así: el Experimento 1 contó con 194 participantes en el estudio preliminar y 184 en el estudio principal; en los Experimentos 2 y 2a, participaron 178 y 189 personas, respectivamente; el Experimento 3 contó con 70 participantes, mientras que en el Experimento 4 participaron 262 personas.

El número de participantes por condición experimental se determinó a partir del análisis de potencia estadística del efecto de IPE realizado por Johnson et al. (2016). En este análisis, se utilizaron las medias y desviaciones estándar de las estimaciones de

comprensión causal de los estudios 2 y 3 de Rozenblit y Keil (2002). A partir de estos valores, se estableció que el número estimado de participantes para observar el efecto de IPE en muestras independientes fue de n=64, con una potencia estadística de 1-β=.80 y α=.05. Teniendo en cuenta este resultado, Johnson et al. (2016) reclutaron entre 55 y 65 participantes por cada condición experimental en sus nueve estudios. Siguiendo este criterio, se intentó reclutar 60 participantes por condición en cada uno de los experimentos. Esta prescripción se cumplió en todos los casos, con excepción del Experimento 3, el cual contó con 35 participantes por condición debido a la tasa de deserción. En los Experimentos 2, 2a y 4, los participantes recibieron extra-crédito en una asignatura u otros beneficios académicos como incentivo por su colaboración. Con el fin de garantizar que todos los estudiantes inscritos en los cursos correspondientes pudiesen obtener estos beneficios, el número de participantes en dichos estudios superó el umbral definido.

Experimento 1: Deseabilidad Social e Ilusión de Profundidad Explicativa

El objetivo del primer experimento fue determinar si la deseabilidad social del conocimiento sobre un tema específico está asociada a una mayor sobrestimación de su comprensión causal. Además del dominio de funcionamiento de dispositivos tradicionalmente utilizado en los estudios sobre IPE, se seleccionaron temas pertenecientes a los dominios histórico y económico. Estos temas fueron incluidos en este experimento por dos razones: en primer lugar, no se han reportado estudios previos sobre la magnitud de la IPE en ellos; en segundo lugar, algunos de estos temas (e.g., la historia del conflicto armado interno o el funcionamiento de la economía del país) se relacionan directamente con los debates públicos de las recientes campañas electorales en Colombia, lo cual puede potenciar su familiaridad, accesibilidad y el valor social atribuido a su conocimiento. Esto

permitió registrar la magnitud de la IPE sobre temas de actualidad cuando existe algún grado de consenso social sobre la deseabilidad de comprenderlos. Adicionalmente, los participantes fueron asignados de manera aleatoria a un único dominio y estimaron su comprensión causal de temas de alta y baja deseabilidad correspondientes, antes y después de explicarlos. El dominio de conocimiento fue manipulado entre grupos para prevenir efectos de comparación de la habilidad para explicar temas de distintos dominios, tales como los reportados por Roeder y Nelson (2015, Experimento 3). En particular, estos investigadores encontraron que los participantes que elaboraron explicaciones sobre un mecanismo concreto (e.g., cómo cocer un huevo), disminuyeron sus estimaciones de comprensión sobre mecanismos pertenecientes a otros dominios (e.g., cómo funciona un helicóptero).

Desde el punto de vista metodológico, el hecho de que un participante estime en la misma sesión experimental su comprensión causal de un tema y la deseabilidad social de dicho conocimiento podría ser una fuente de sesgo sistemático. Por ejemplo, si una persona califica el conocimiento causal sobre un tema determinado como altamente deseable, es posible que se sienta inducida a reportar que comprende dicho tema en detalle, aunque en realidad crea que su comprensión es mucho más limitada. Si este es el caso, los efectos encontrados podrían ser artefactos producidos por la influencia recíproca de las preguntas y no por el uso de la deseabilidad social como clave inferencial. Para controlar esta potencial interacción entre preguntas, la deseabilidad social percibida de cada tema fue evaluada por una muestra distinta de participantes en un sondeo previo. Los temas que presentaron mayores diferencias en esta variable dentro de cada dominio fueron seleccionados para el estudio principal.

Participantes

Ciento ochenta y cuatro estudiantes de la Universidad Nacional de Colombia (88 mujeres) de diferentes programas de pregrado y niveles de formación, con edades entre 18 y 42 años (*M*=20.7, *D.E.*=2.93). Para controlar el efecto de la experticia, ningún estudiante de historia o economía participó en este estudio. Los participantes recibieron extra-crédito académico en una asignatura por su colaboración en este estudio.

Diseño

Se utilizó un diseño experimental mixto de 3x2, con el dominio de explicación (histórico, económico y funcionamiento de dispositivos) como variable entre-sujetos; la deseabilidad social del conocimiento sobre el tema (alta y baja) fue la variable intra-sujeto. La variable dependiente fue la diferencia entre las estimaciones de comprensión realizadas antes y después de explicar los temas asignados.

Teniendo en cuenta la posibilidad de generar un efecto de demanda (Zizzo, 2010) al pedir a los mismos participantes realizar estimaciones de su conocimiento sobre un tema particular y juzgar qué tan avergonzados se sentirían de no saber sobre él, se siguió un procedimiento similar al de Rozenblit y Keil (2002, estudio 11) para seleccionar los temas de alta y baja deseabilidad dentro de cada dominio. En un estudio preliminar, ciento noventa y cuatro participantes (119 mujeres, *M*=21.2 años, *D.E.*=3.47) estimaron la deseabilidad social del conocimiento sobre siete temas en cada uno de los tres dominios, para un total de 21 ítems. Los participantes indicaron qué tan avergonzados se sentirían de no tener una buena comprensión de cada uno de estos temas en una escala Likert de 1 a 7, siendo 1 = “Si yo no tuviera una buena comprensión o conocimiento de este tema, para nada me sentiría avergonzado” y 7 = “Si yo no tuviera una buena comprensión o

conocimiento de este tema, me sentiría muy avergonzado" (ver Anexo A). Esta escala de ítem único fue idéntica a la utilizada en el único estudio reportado sobre deseabilidad social del conocimiento e IPE (Rozenblit & Keil, 2002, estudio 11). Los dos temas entre los cuales se encontraron mayores diferencias dentro de cada dominio (6 ítems en total) fueron seleccionados para el estudio principal (ver Anexo B).

Materiales y procedimiento

Test de verificación de la manipulación de las instrucciones (*Instructional manipulation check*). Esta prueba está diseñada para verificar que los participantes en un experimento leen cuidadosamente las instrucciones presentadas**.** Con este propósito, se le presentó a cada participante una pregunta en un formato similar al de otras preguntas del estudio (e.g., selección múltiple con única respuesta) junto con un párrafo que incluía instrucciones complementarias. En este párrafo, se le pedía al participante ignorar la pregunta presentada y se le indicaba la acción que debía llevar a cabo para continuar (ver Anexo C). Cuando los participantes no realizaban la acción solicitada en las instrucciones, se generaba un mensaje de advertencia que los invitaba a leerlas con más cuidado antes de responder. En un estudio previo, Oppenheimer, Meyvis y Davidenko (2009, estudio 2) encontraron que los participantes a los cuales se les invitó a leer de nuevo las instrucciones en una prueba de este tipo posteriormente fueron sensibles a las mismas manipulaciones experimentales que los participantes que siguieron correctamente las instrucciones desde el inicio.

Procedimiento. En una sala aislada y libre de ruido, los participantes realizaron las tareas experimentales en computadores personales ubicados en cubículos separados. Para la presentación de las tareas y el registro de las respuestas se utilizó la plataforma de

encuestas Qualtrics (2005). Después de diligenciar el consentimiento informado e información demográfica, los participantes completaron un test de verificación de lectura de las instrucciones de la tarea. Responder correctamente esta prueba fue un requisito indispensable para avanzar a la tarea principal.

Las siguientes tres fases del experimento se ajustaron al paradigma básico de IPE: inicialmente, los participantes estimaron su conocimiento sobre 12 objetos/temas utilizando una escala de 1 a 7, donde 1 significa "comprensión vaga o pobre" y 7 "comprensión detallada" (fase de estimación previa). Las instrucciones previas para utilizar esta escala de calificación fueron adaptadas de Rozenblit y Keil (2002) y Fernbach, Rogers, et al. (2013; ver Anexo D). A continuación, los participantes fueron asignados aleatoriamente a uno de tres dominios (histórico, económico y dispositivos) y se les pidió explicar paso a paso el mecanismo causal del tema de alta o baja deseabilidad dentro de dicho dominio (fase de explicación). Una vez completada esta tarea, los participantes estimaron nuevamente su comprensión sobre dicho tema (fase de estimación post). La secuencia explicación-estimación post fue repetida para el tema restante (de alta o baja de deseabilidad). De este modo, cada participante explicó y estimó de nuevo su comprensión sobre los dos temas del dominio asignado que presentaron mayores diferencias en las puntuaciones de deseabilidad social en el estudio preliminar (ver Anexo B). El orden de presentación de los temas de alta y baja deseabilidad para cada participante fue asignado aleatoriamente, mientras que la secuencia explicación-estimación post se mantuvo constante para de cada uno de los dos temas presentados. La persona a cargo del estudio verificó que los participantes no accedieran a otras páginas web durante la sesión. Al finalizar el experimento, se agradeció a

los participantes por su colaboración y se registraron algunos datos personales para asignar el respectivo extra-crédito académico.

Experimento 2: Controlando el Sesgo de Auto-promoción (*Self-enhancement*) y el Efecto de la Familiaridad.

El objetivo del Experimento 2 fue realizar un test empírico robusto del efecto de la deseabilidad social del conocimiento sobre la IPE y descartar que dicho efecto fuese ocasionado por la baja familiaridad con los temas menos deseables o por un sesgo deliberado de respuesta. A diferencia del Experimento 1, los temas de alta y baja deseabilidad social fueron asignados a grupos distintos de participantes, y se emplearon únicamente temas relacionados con la historia del conflicto armado interno colombiano.

Una limitación del Experimento 1 es que no es posible saber si los participantes en el estudio principal hicieron estimaciones de deseabilidad social del conocimiento consistentes con las reportadas por los participantes en el sondeo previo. Para superar esta limitación, se requiere registrar las estimaciones de deseabilidad y comprensión de cada participante sin crear un efecto de demanda. Por esta razón, en el Experimento 2 los participantes asistieron a dos sesiones separadas entre sí por al menos 15 días. En la primera de ellas, los participantes respondieron una escala de deseabilidad social del conocimiento sobre los temas mencionados, compuesta por cuatro preguntas. Después de la primera sesión, se seleccionaron las parejas de temas cuya comprensión fue calificada como más deseable y menos deseable. En la segunda sesión, los participantes realizaron el paradigma de IPE con dos de estos temas, según la condición experimental a la cual fueron asignados.

Por otro lado, es posible que el efecto de la deseabilidad social en la IPE sea una consecuencia de un sesgo de auto-promoción (*self-enhancement*). Específicamente, los participantes podrían distorsionar sus estimaciones iniciales de comprensión causal con el fin de mostrarse como conocedores de temas socialmente relevantes, incluso cuando son plenamente conscientes de su desconocimiento de dichos temas. Si este es el caso, la deseabilidad social no operaría como una clave inferencial en el nivel metacognitivo, sino como una señal de activación de un sesgo de respuesta deliberado. Para descartar esta explicación alternativa del efecto de la deseabilidad social sobre la IPE, los participantes en los estudios 2 y 2a realizaron una prueba adicional que medía la tendencia a exagerar el conocimiento sobre un dominio particular. Este test se diseñó utilizando la técnica de *overclaiming*, originada en la investigación sobre sesgos de deseabilidad social de la respuesta en pruebas de personalidad (Paulhus, 2012; Paulhus et al., 2003) y evaluación del logro académico (Clariana, Castelló, & Cladellas, 2016; Paulhus & Dubois, 2014). En este tipo de pruebas, los participantes califican qué tan familiarizados se encuentran con distintos términos pertenecientes a un dominio específico (e.g., biología). Mientras que el 80% de estos términos son reales (e.g., mitosis), el 20% restante son ficticios (e.g., metatoxina). Comparando las respuestas de los participantes a los ítems reales y ficticios mediante un análisis de detección de señal, es posible medir simultáneamente la habilidad para identificar el conocimiento real (Precisión=Aciertos-Falsas alarmas) y la tendencia a exagerar el conocimiento reportado (Sesgo=[Aciertos + Falsas alarmas] /2; Paulhus & Dubois, 2014). Investigaciones previas han mostrado que estas medidas son parcialmente independientes de la inteligencia cristalizada y de los rasgos de personalidad asociados a los sesgos de respuesta (Bensch, Paulhus, Stankov, & Ziegler, 2017), y que las personas que creen tener más conocimientos sobre dominios específicos tienden a reportar más

familiaridad con ítems ficticios asociados a tales dominios (Atir et al., 2015; Hülür, Wilhelm, & Schipolowski, 2011). Si el efecto de la deseabilidad social sobre la IPE puede ser explicado a partir de los puntajes obtenidos en una prueba de *overclaiming* en el dominio correspondiente, es más probable que dicho efecto sea resultado de un sesgo de auto-promoción que de un proceso inferencial metacognitivo.

De otra parte, existe evidencia de que la magnitud del FOK tiende a ser menor cuando existe poca familiaridad con el contenido de las preguntas (Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992). En este sentido, es posible que los temas de baja deseabilidad social utilizados en el Experimento 1 tuviesen una menor IPE asociada debido a que resultaron poco familiares para los participantes. Así, la baja familiaridad con los temas menos deseables podría generar menores estimaciones iniciales de comprensión causal y, por consiguiente, menores diferencias con las estimaciones posteriores. En la medida en que la familiaridad con un tema y la deseabilidad social de su comprensión pueden estar asociadas, manipular ortogonalmente estas variables supone un desafío metodológico importante. En particular, es difícil encontrar temas de alta relevancia social que al mismo tiempo sean poco familiares o, de manera inversa, temas considerados poco relevantes y que resulten muy familiares para sectores amplios de la sociedad. Una alternativa para controlar el posible efecto de la baja familiaridad consiste en garantizar que los temas menos deseables no fuesen completamente desconocidos para los participantes. En este sentido, al seleccionar los temas de baja deseabilidad para la tarea de IPE del Experimento 2 se tuvieron en cuenta únicamente los temas cuyas puntuaciones se ubicaron por encima del percentil 30 (i.e., 1) en la escala de familiaridad que completaron los participantes durante la primera sesión.

Participantes

Ciento setenta y ocho estudiantes (97 mujeres) de pregrado en psicología y de varios programas de ciencias de la salud de la Universidad Nacional de Colombia, con edades entre 18 y 42 años (*M*=20.51 años, *D.E.*=3.22). Al igual que en el Experimento 1, ningún estudiante de historia o economía participó en este estudio. No obstante, 24 participantes afirmaron estar cursando alguna asignatura universitaria sobre historia de Colombia. Los participantes recibieron extra-crédito académico en cursos ofertados por el departamento de psicología como incentivo por su colaboración. Veintiún participantes fueron excluidos de los análisis debido a que no asistieron a la segunda sesión.

Diseño

Se utilizó un diseño experimental mixto de 2x2, con la deseabilidad social del conocimiento (alta o baja) como variable entre-sujetos; los dos temas presentados en cada condición fueron tratados como variable intra-sujeto. La variable dependiente fue la diferencia en las estimaciones de comprensión sobre los temas asignados, realizadas antes y después de explicarlos. El puntaje en la prueba de *overclaiming* se utilizó como indicador del sesgo de auto-promoción y fue incluido como covariada en el análisis de resultados.

Materiales y Procedimiento

Escala de deseabilidad social del conocimiento. Esta escala incluyó cuatro preguntas tipo Likert de 7 puntos para cada uno de los 16 temas históricos, las cuales indagaban el valor social y la importancia atribuida a su comprensión (ver Anexo E). A diferencia del Experimento 1, en dichas preguntas se incluyeron tanto indicadores positivos (e.g., “qué tan importante es para usted tener una buena comprensión de…”) como negativos (e.g., “qué tan avergonzado se sentiría de no saber sobre…”). Los puntajes totales

para cada tema se calcularon promediando las respuestas de cada participante a las cuatro preguntas correspondientes. Este procedimiento se agregó para tener un estimador confiable que pudiera ser defendible frente a una audiencia psicométrica. La consistencia interna de esta escala en los cuatro temas seleccionados estuvo dentro de un rango óptimo (α de Cronbach≥.7; ver Anexo H).

Escala de familiaridad. Los participantes indicaron qué tan familiarizados estaban con cada uno de los 16 temas históricos en una escala Likert de 7 puntos, siendo 1= "Para nada estoy familiarizado(a) con el tema" y 7= "Estoy muy familiarizado(a) con el tema". Este tipo de escala de ítem único ha sido utilizada previamente en estudios sobre FOK, IPE e IJA (Costermans et al., 1992; Johnson et al., 2017; Koriat & Levy-Sadot, 2001; Rozenblit & Keil, 2002).

Prueba de *overclaiming* sobre el conflicto armado colombiano. En esta prueba, cada participante calificó su familiaridad con 20 acontecimientos específicos de la historia del conflicto armado interno en Colombia, de los cuales cuatro eran ficticios (Anexo J). Los índices de consistencia interna de las subescalas de 16 sucesos reales (α de Cronbach=.88) y cuatro eventos falsos (α de Cronbach=.69) estuvieron dentro de límites aceptables.

Procedimiento. Las tareas experimentales fueron realizadas por los participantes en el mismo lugar y a través de la misma plataforma utilizada en el Experimento 1. En la primera sesión, los participantes diligenciaron el consentimiento informado y el test de verificación de lectura de instrucciones. A continuación, respondieron la prueba de *overclaiming* sobre el conocimiento del conflicto armado colombiano y completaron las preguntas de la escala de deseabilidad social. Los dos pares de temas que registraron las puntuaciones promedio más altas y más bajas en esta escala fueron seleccionados para las

condiciones de alta y baja deseabilidad del estudio principal, respectivamente (Anexo F). Posteriormente, cada participante respondió la escala de familiaridad. Para garantizar que los participantes tuviesen algún grado de familiaridad con los temas menos deseables, se verificó que los temas seleccionados para el estudio principal tuviesen medias superiores a 2 en esta escala (Anexo G). Adicionalmente, los participantes reportaron si estaban cursando alguna asignatura universitaria sobre historia de Colombia y diligenciaron información demográfica.

Una vez transcurridas dos semanas desde la primera sesión, los participantes fueron citados nuevamente. Al inicio de la segunda sesión, los participantes estimaron su comprensión previa con respecto a 12 de los 16 temas presentados en la primera sesión (fase de estimación previa). Las instrucciones de uso de esta escala fueron similares a las proporcionadas en el Experimento 1. Posteriormente, cada participante fue asignado de manera aleatoria a una de las dos condiciones experimentales (deseabilidad alta o baja) y se les pidió explicar uno de los temas pertenecientes a dicha condición. Después de escribir su explicación, los participantes estimaron nuevamente su comprensión causal de cada tema (fase de estimación post). Esta secuencia fue repetida para el segundo tema. El orden de presentación de los temas dentro de cada condición fue asignado aleatoriamente a cada participante, mientras que la secuencia explicación-estimación post se mantuvo constante para de cada uno de los dos temas. La persona a cargo del estudio verificó que los participantes no accedieran a otros sitios web durante la sesión experimental. Por último, los participantes calificaron nuevamente la deseabilidad social del conocimiento en la misma escala utilizada en la primera sesión y registraron algunos datos personales para obtener el extra-crédito académico respectivo.

Experimento 2a: Controlando el Efecto de Exposición Previa.

Como se explicará en la sección de resultados, no hubo un efecto entre-sujetos de la deseabilidad social del conocimiento. Esto pudo deberse o bien a las características específicas de las dos combinaciones de temas históricos utilizados, o a la doble exposición al paradigma de IPE. Por esta razón, en el Experimento 2a cada participante respondió el paradigma de IPE con respecto a un único tema, el cual fue seleccionado aleatoriamente de un conjunto de cuatro temas posibles por condición. Estos ocho temas fueron seleccionados a partir de los resultados obtenidos en las escalas de deseabilidad social y familiaridad del Experimento 2. Para verificar que las estimaciones de deseabilidad social del conocimiento realizadas por los participantes en ambos experimentos fuesen similares, los participantes en el Experimento 2a completaron la escala correspondiente después de realizar el paradigma de IPE.

Participantes

Ciento ochenta y nueve estudiantes (109 mujeres) de programas de pregrado de tres universidades colombianas, con edades entre 18 y 44 años (M=21.59 años, $D.E.$=3.38). Al igual que en el Experimento 1, ningún estudiante de historia o economía participó en este estudio. Treinta y cuatro participantes afirmaron haber curso o estar cursando asignaturas universitarias sobre historia de Colombia. La mayoría de los participantes recibieron extra-crédito académico en una asignatura como incentivo por su colaboración.

Diseño

Se utilizó un diseño experimental de dos grupos, en el cual la presentación de temas de alta o baja deseabilidad social fue manipulada entre sujetos. La variable dependiente fue la diferencia en las estimaciones de comprensión causal solicitadas antes y después de

explicar el tema asignado. Los indicadores de sesgo de auto-promoción derivados de los puntajes en la prueba de *overclaiming* fueron incluidos como covariadas en el análisis de resultados.

Este diseño fue similar al del Experimento 2, con dos excepciones: en primer lugar, el Experimento 2a se realizó en una única sesión. En segundo lugar, los participantes realizaron el paradigma de IPE sobre un único tema, el cual fue seleccionado a partir de un conjunto de cuatro temas posibles dentro de cada condición experimental. Los dos grupos de temas fueron seleccionados a partir de las estimaciones de familiaridad y deseabilidad social reportadas por los participantes del Experimento 2 (ver Anexos F y G).

Materiales y Procedimiento

Los instrumentos utilizados fueron los mismos descritos en el procedimiento el Experimento 2, con excepción de los temas históricos adicionales (i.e., paramilitarismo, intensidad del conflicto en las zonas rurales, alianzas entre paramilitares y políticos, movimiento constituyente de 1990, genocidio de la Unión Patriótica, Plan Colombia). El experimento se realizó en una única sesión en espacios similares al utilizado en los Experimentos 1 y 2. La persona a cargo del estudio verificó que los participantes no accedieran a otros sitios web durante la sesión experimental. Después de completar el consentimiento informado y el test de verificación de lectura de las instrucciones, los participantes estimaron su comprensión previa con respecto a 12 temas históricos (fase de estimación previa). Las instrucciones de uso de esta escala fueron idénticas a las utilizadas en el Experimento 2. A continuación, los participantes fueron asignados aleatoriamente a una de las dos condiciones experimentales (deseabilidad alta o baja) y se les pidió explicar uno de los temas pertenecientes a dicha condición. Después de escribir su explicación, los

participantes estimaron nuevamente su comprensión causal de cada tema (fase de estimación post). El tema presentado a cada participante fue seleccionado aleatoriamente del grupo correspondiente a cada condición experimental. Posteriormente, los participantes respondieron la prueba de *overclaiming* sobre el conocimiento del conflicto armado colombiano y calificaron la deseabilidad social del conocimiento en la misma escala utilizada en el Experimento 2. Antes de finalizar el experimento, los participantes reportaron si se encontraban cursando alguna asignatura universitaria relacionada con la historia de Colombia y registraron datos personales para la asignación del extra-crédito académico respectivo.

Experimento 3: Deseabilidad Social, Ilusión de Profundidad Explicativa y Tipo de Procesamiento

El propósito principal del tercer experimento fue establecer si existe una interacción entre la habilidad para realizar un procesamiento analítico/controlado (tipo 2) y el efecto de la deseabilidad social del conocimiento sobre la IPE. Un objetivo adicional fue el de replicar los resultados del Experimento 2 con participantes provenientes de distintas regiones del país y con diferentes niveles de educación formal. Al igual que en los Experimentos 2 y 2a, los temas de alta y baja deseabilidad social fueron asignados a grupos distintos de participantes. En este caso, los temas seleccionados fueron dos de los puntos del acuerdo de paz entre el gobierno nacional y las FARC, y dos propuestas legislativas en trámite en el Congreso de la República durante el segundo semestre de 2016. Al igual que en el Experimento 2, los mismos participantes que juzgaron la deseabilidad social del conocimiento sobre cada tema realizaron dos semanas después la tarea experimental de IPE. La versión original del *Cognitive Reflection Test* (CRT; Frederick, 2005) fue utilizada

como indicador de la habilidad para realizar procesamientos tipo 2. El puntaje obtenido por los participantes en esta prueba fue incluido como covariada en el análisis principal.

Participantes

Setenta personas (44 mujeres) entre los 18 y los 59 años participaron en las dos fases de este estudio (*M*=28.75 años, *D.E.*=12.11). Con el fin maximizar la variabilidad de las orientaciones ideológicas y el nivel de conocimiento sobre los temas, la muestra incluyó personas de diferentes edades, niveles educativos y regiones del país: 24% de ellos contaban únicamente con educación secundaria o técnica, 50% con estudios de pregrado en curso o finalizados y 17% con formación de posgrado. Los participantes fueron contactados a través de listas de correos y redes sociales en cinco ciudades de Colombia (Bogotá, Medellín, Manizales, Bucaramanga y Yopal). De las 130 personas que aceptaron participar en el experimento inicialmente, 60 (46%) no completaron la segunda sesión, de modo que sus respuestas no fueron incluidas en el análisis de resultados.

Diseño

Se implementó un diseño experimental mixto de 2x2, en el cual la asignación de propuestas legislativas de alta o baja deseabilidad social fue manipulada entre sujetos. Las dos propuestas legislativas presentadas en cada condición fueron tratadas como variable intra-sujeto. La variable dependiente fue la diferencia en las estimaciones de comprensión de las propuestas legislativas asignadas, antes y después de formular explicaciones sobre ellas. El estudio fue dividido en dos sesiones, utilizando un procedimiento similar al del Experimento 2. En la primera sesión, cada participante estimó la deseabilidad social del conocimiento sobre un conjunto de propuestas legislativas. Durante la segunda sesión, realizada dos semanas después, las parejas de propuestas con puntajes de deseabilidad más

altos y bajos en promedio fueron presentadas a grupos distintos de participantes. El desempeño de cada participante en la versión original del CRT (Frederick, 2005) fue utilizado como indicador de la capacidad para realizar procesamientos tipo 2 e incluido como covariada en el análisis de resultados.

Materiales y Procedimiento

***Cognitive Reflection Test* (CRT).** Creada originalmente por Frederick (2005), esta prueba mide la habilidad de las personas para realizar procesamientos analíticos a partir de sus respuestas a tres problemas aritméticos (ver Anexo N). Estos problemas se caracterizan por tener una respuesta intuitiva incorrecta saliente y una respuesta correcta que requiere de procesamiento controlado. El puntaje total equivale a la suma de las respuestas correctas de cada participante (entre cero y tres). Se ha encontrado que la capacidad para responder correctamente el CRT se asocia a una menor probabilidad de incurrir en sesgos e inferencias heurísticas erróneas en tareas de razonamiento (Toplak, West, & Stanovich, 2011) y a una disposición general hacia un estilo de pensamiento reflexivo (Pennycook, Fugelsang, & Koehler, 2015a). Aunque el contenido de esta prueba se ha divulgado en los últimos años a través de internet, existe evidencia de que su validez predictiva no disminuye a pesar de las exposiciones repetidas (Bialek & Pennycook, 2017). La consistencia interna del CRT fue aceptable (α de Cronbach = .699)

Procedimiento. Todas las tareas fueron presentadas a través la plataforma Qualtrics y los enlaces de acceso correspondientes fueron enviados vía correo electrónico. En la primera sesión del experimento, 130 participantes (82 mujeres) calificaron la deseabilidad social de 24 propuestas políticas seleccionadas de la agenda legislativa del Congreso de la República de Colombia durante la legislatura de 2016 (ver Anexo K), a través de la misma

escala utilizada en los Experimentos 2 y 2a. Los participantes respondieron estas cuatro preguntas para cada una de las propuestas políticas presentadas. Los índices de consistencia interna de esta escala (α de Cronbach) estuvieron en un intervalo de variación aceptable en los 24 ítems (*M*=.79, *D.E.*=.053; I.C. 95%= [.70 - .74], ver Anexo K). A partir de estos resultados, se seleccionaron dos propuestas cuya comprensión fue juzgada como altamente deseable y dos que fueron consideradas menos deseables (ver Anexo L).

Dos semanas después de completar la primera fase del estudio, los participantes recibieron vía correo electrónico el enlace de acceso a la segunda. Al inicio de esta sesión, los participantes realizaron el test de verificación de lectura de instrucciones usado en el Experimento 1 y diligenciaron un breve cuestionario demográfico. A continuación, los participantes estimaron su comprensión previa sobre 12 propuestas políticas (fase de estimación previa). Las instrucciones para utilizar la escala de comprensión fueron adaptadas a partir de las empleadas en los experimentos anteriores (Anexo M). En este punto, los participantes fueron asignados aleatoriamente a uno de dos grupos: a uno de ellos se les pidió explicar paso a paso el contenido de las propuestas políticas más deseables (i.e. acuerdo sobre víctimas y jurisdicción especial para la paz/ acuerdo sobre solución al problema de las drogas ilícitas) de la manera más completa posible, evitando utilizar fuentes adicionales de información (e.g., páginas web, revistas, periódicos, libros, etc.). Por su parte, el grupo restante formuló explicaciones sobre las propuestas cuyo conocimiento fue juzgado como menos deseable en la primera fase (i.e. ley de derechos laborales para los actores/reglas de convalidación de títulos extranjeros de cirugía plástica). A continuación, los participantes calificaron nuevamente su comprensión sobre la propuesta (estimación post). La secuencia explicación-estimación post fue repetida para el segundo tema. El orden

de presentación de los temas fue asignado al azar, al igual que el momento en el cual los participantes resolvieron el CRT (antes o después de realizar el paradigma de IPE). Al finalizar el experimento, se agradeció al participante por su colaboración y sus datos personales fueron registrados para el envío posterior de información sobre los objetivos del experimento (ver Anexo O).

Experimento 4: Manipulación Experimental de la Sobrecarga Ejecutiva

El objetivo del Experimento 3 fue determinar si existe una interacción entre el tipo de procesamiento y la asociación entre deseabilidad social y sobrestimación del conocimiento causal. Sin embargo, por las características de su diseño, este estudio solo permitió obtener evidencia correlacional de dicha interacción. Por otro lado, es posible que los participantes que se consideraban menos informados sobre los temas presentados en la primera sesión del Experimento 3 fuesen los más propensos a abandonar el estudio; si este sesgo de selección (*selective attrition*) causó la alta tasa de abandono de los participantes, la validez interna del experimento se vería amenazada (Zhou & Fishbach, 2016).

En este contexto, el objetivo del Experimento 4 fue obtener evidencia robusta de efecto moderador del tipo de procesamiento sobre la asociación entre deseabilidad social e IPE. El nivel de sobrecarga ejecutiva y el tiempo disponible para realizar la estimación inicial de comprensión causal fueron manipulados con el fin de facilitar o inhibir la activación de procesamientos tipo 2. En el grupo de sobrecarga ejecutiva alta, los participantes realizaron dos tareas de memoria concurrentes (una de tipo visoespacial y otra de recuerdo de sílabas) durante un tiempo limitado. Estas dos tareas fueron incluidas con el fin generar un efecto de sobrecarga tanto en el componente visual (agenda viso-espacial) como en el verbal (bucle fonológico) de la memoria de trabajo (Baddeley, 2007). Por su

parte, en el grupo de sobrecarga ejecutiva baja, los participantes resolvieron una versión más sencilla de la tarea de memoria visoespacial sin límites de tiempo. La manipulación simultánea de la deseabilidad social y el nivel de sobrecarga ejecutiva se llevó a cabo para determinar si la dificultad para realizar inferencias analíticas promueve que los participantes dependan más de claves heurísticas como la deseabilidad social al momento de estimar su comprensión causal.

Participantes

Doscientos sesenta y dos estudiantes universitarios (164 mujeres) de diferentes programas de pregrado de universidades públicas y privadas de Bogotá y Cundinamarca, con edades entre 18 y 50 años (*M*=20.9 años, *D.E.*=3.6). Al igual que en los experimentos anteriores, ningún estudiante de historia o economía participó en este estudio. Del total de participantes, 103 afirmaron haber cursado o estar cursando alguna asignatura universitaria sobre historia de Colombia. La mayoría de los participantes recibieron extra-crédito académico en alguna asignatura o certificados de apoyo a actividades académicas como incentivo por su colaboración. Diez participantes fueron excluidos de los análisis de resultados debido a que sus puntajes en las tareas de memoria asignadas fueron muy bajos (i.e., tres desviaciones estándar o más por debajo del promedio de su grupo experimental).

Diseño

Se implementó un diseño experimental mixto de 2x2. La deseabilidad social del conocimiento sobre los temas a evaluar (alta/baja) y el nivel de sobrecarga ejecutiva (alta/baja) fueron manipuladas entre sujetos. Se utilizó el mismo conjunto de temas de alta y baja deseabilidad social que en el Experimento 2a. La variable dependiente fue la diferencia en las estimaciones de comprensión causal antes y después de escribir las

explicaciones sobre el tema asignado. El puntaje de cada participante en el CRT, el porcentaje de aciertos en las tareas de memoria y el haber cursado asignaturas sobre historia de Colombia fueron incluidos como covariadas en el análisis de resultados.

Materiales y Procedimiento

Tarea de memoria visoespacial. Se presentaron matrices de 3x3 con tres o cuatro casillas coloreadas a los participantes, dependiendo de la condición experimental asignada. En la condición de sobrecarga baja, tres casillas coloreadas formaban líneas horizontales o verticales; en la condición de sobrecarga alta, cuatro de las casillas formaban patrones más complejos (ver Figura 6). Estas matrices fueron presentadas durante dos segundos. En un momento posterior, los participantes reprodujeron el patrón observado haciendo click en las casillas correspondientes de una matriz en blanco. Para evitar efectos de facilitación por exposición repetida, se utilizaron ocho matrices distintas para cada una de las condiciones experimentales (ver Anexo P). Esta tarea fue adaptada a partir de la tarea de memorización de puntos (*dot memory task*) de De Neys (2006b). En investigaciones previas, se ha encontrado que la memorización de los patrones complejos de cuatro casillas demanda recursos ejecutivos centrales (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) e interfiere con el desempeño en tareas que requieren inferencias analíticas, como el CRT (Johnson et. al, 2014).

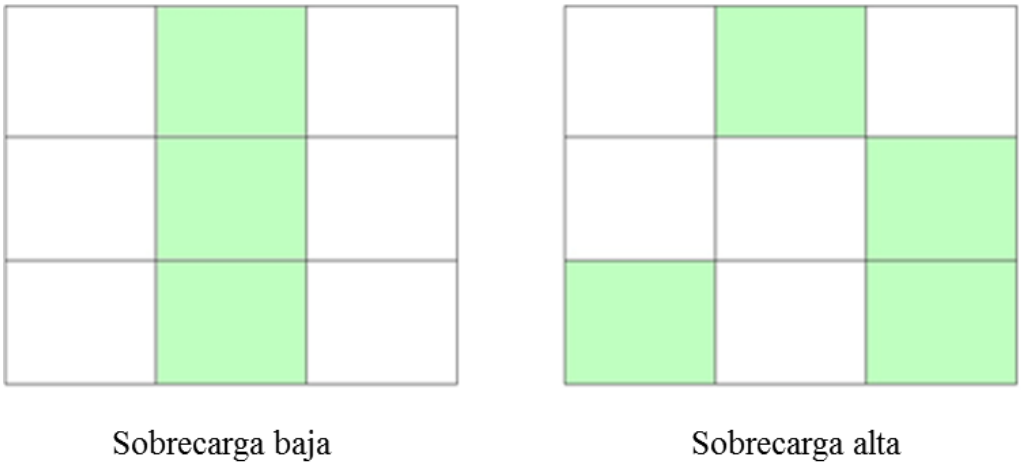


Figura 6. Ejemplos de matrices utilizadas en la tarea de memoria visoespacial en las condiciones de sobrecarga baja y sobrecarga alta del Experimento 4.

Tarea de memorización de sílabas. Se presentaron dos sílabas sin sentido a los participantes y se les pidió memorizarlas. Con el fin de evitar efectos de facilitación por exposición repetida, se utilizaron ocho pares de sílabas. Después de realizar la estimación inicial de comprensión causal de cada tema y la tarea de memoria visoespacial, se les pidió escribir las dos sílabas en el orden en el cual fueron mostradas (ver Anexo P). Las sílabas fueron presentadas durante dos segundos. Esta tarea fue presentada únicamente en la condición de sobrecarga ejecutiva alta.

Procedimiento. La sesión experimental se llevó a cabo en salas similares a la utilizada en los Experimentos 1, 2 y 2a, empleando la plataforma de encuestas Qualtrics. Después de diligenciar el consentimiento informado, los participantes realizaron el test de verificación de lectura de las instrucciones utilizado en estudios anteriores. A continuación, los participantes recibieron indicaciones sobre cómo realizar las tareas de memoria y llevaron a cabo dos ejercicios de práctica de cada una. Cuando los participantes cometieron algún error en los ejercicios de práctica, se les pidió leer nuevamente las instrucciones y

repetir el ejercicio. Responder correctamente los ítems de práctica fue un requisito indispensable para continuar con la tarea principal. Después de esto, los participantes recibieron instrucciones sobre cómo utilizar la escala de comprensión previa, las cuales fueron similares a las utilizadas en los experimentos anteriores (Anexo D).

Al avanzar a la tarea principal, cada participante fue asignado a una de dos condiciones experimentales: sobrecarga ejecutiva alta o baja. En la condición de sobrecarga alta, se le pidió los participantes memorizar la matriz de casillas coloreadas y una pareja de sílabas. Cada uno de estos estímulos fue presentado durante dos segundos. A continuación, los participantes tuvieron cuatro segundos para estimar su comprensión previa de uno de los temas históricos seleccionados para el estudio. Después de realizar este juicio, se les pidió reproducir el patrón de casillas coloreadas en una matriz en blanco y escribir las sílabas memorizadas previamente. Como control de lectura de la tarea, en la siguiente pantalla cada participante seleccionó en una lista el tema histórico cuya comprensión había estimado. Esta secuencia se repitió para cada uno de los ocho temas históricos seleccionados, los cuales fueron presentados aleatoriamente (fase de estimación previa). Los participantes asignados a la condición de sobrecarga baja realizaron una secuencia de tareas similares a la del grupo de sobrecarga alta, con tres excepciones: memorizaron únicamente las matrices con patrones lineales, no llevaron a cabo la tarea de memorización de sílabas y no tuvieron restricciones en el tiempo disponible para realizar la tarea.

Después de completar sus juicios de comprensión previa y las tareas de memoria, se pidió a los participantes explicar las causas de uno de los temas históricos pertenecientes a la condición de deseabilidad asignada (alta o baja) y estimar nuevamente su comprensión sobre el tema (fase de estimación post). A continuación, los participantes completaron la

escala de deseabilidad social utilizada en los estudios 2 y 2a a propósito de los ocho temas históricos presentados. Por último, los participantes respondieron el CRT y diligenciaron información demográfica.

La Tabla 1 recapitula las preguntas de investigación y los diseños de los cinco experimentos descritos en este capítulo. En dicha tabla se especifican los factores fijos, factores aleatorios, covariadas y variables dependientes incluidas en los análisis de resultados.

Tabla 1.

Resumen de los experimentos realizados.

Experimento	Pregunta de investigación	Factores fijos	Factores aleatorios (Medidas repetidas)	Covariadas	Variable dependiente
1	¿La deseabilidad social se asocia a la IPE en dominios de conocimiento natural y social?	Dominio de conocimiento (Histórico, económico, dispositivos) Orden de presentación	Deseabilidad social (Alta y baja)	Ninguna	Diferencia entre las estimaciones de comprensión pre y post-explicación
2	¿La deseabilidad social se asocia a la IPE cuando se controla el efecto de la baja familiaridad y del sesgo de auto-promoción? (diseño entre-sujetos)	Deseabilidad social (Alta y baja)	Tema presentado (2 temas por participante)	Puntajes del test de *overclaiming*	Diferencia entre las estimaciones de comprensión pre y post-explicación

2a	¿La deseabilidad social se asocia a la IPE cuando varían los temas presentados y se elimina la doble exposición al paradigma experimental? (diseño entre-sujetos).	Deseabilidad social (Alta y baja)	Ninguno	Puntajes del test de *overclaiming* Cursos de historia de Colombia	Diferencia entre las estimaciones de comprensión pre y post-explicación
3	¿La habilidad para realizar procesamientos tipo 2 modera (interactúa con) la asociación entre deseabilidad social e IPE?	Deseabilidad social (Alta y baja)	Tema presentado (2 temas por participante)	CRT	Diferencia entre las estimaciones de comprensión pre y post-explicación
4	¿El nivel de sobrecarga ejecutiva modera la asociación entre deseabilidad social e IPE?	Deseabilidad social (Alta y baja) Sobrecarga ejecutiva (Alta y baja)	Ninguno	CRT Desempeño en las pruebas de memoria Cursos de historia de Colombia	Diferencia entre las estimaciones de comprensión pre y post-explicación

Capítulo 3: Resultados

Experimento 1

Plan de análisis. En el Experimento 1 se empleó un diseño experimental mixto de 3x2, con el dominio de explicación como variable entre-sujetos; la deseabilidad social del conocimiento sobre los temas presentados (alta y baja) fue la variable intra-sujeto. La diferencia en las estimaciones de comprensión sobre cada tema fue la variable dependiente.

El objetivo de este experimento fue determinar si existen diferencias en la magnitud de la IPE entre temas de alta y baja deseabilidad social en tres dominios de conocimiento: histórico, económico y dispositivos. Debido a que cada participante estimó su conocimiento causal sobre dos temas con distintos niveles de deseabilidad social en cada dominio, la estructura de este diseño es anidada. En otras palabras, las estimaciones de comprensión realizadas por un mismo participante sobre los temas de alta y baja deseabilidad social en el dominio asignado están relacionadas entre sí. Así, el efecto de la deseabilidad social en este diseño experimental es aleatorio, porque varía entre participantes. Para tener en cuenta esta dependencia estadística de los residuales, se construyeron modelos lineales jerárquicos (MLJ, en adelante) en los cuales se incluyó como factor aleatorio la deseabilidad social, mientras que los factores fijos fueron el dominio de conocimiento y el orden de presentación. Esta última variable se incluyó en los modelos para verificar que el efecto de la deseabilidad social no fuese un artefacto producido por la comparación entre temas más y menos deseables dentro de cada dominio. La variable dependiente fue la diferencia entre las estimaciones de comprensión previas y posteriores a la generación de explicaciones (i.e., la magnitud de la IPE). Este modelo (ver Tabla 3, Modelo 3) se representa en la ecuación 5, donde IPE_{ij} es el valor de la IPE para el participante *i* en el tema *j*, $b_0$es el

intercepto del modelo general, b_1 es el coeficiente del efecto de deseabilidad social (*DS*), b_2 y $b_3$son los coeficientes de los efectos fijos de dominio (*DOM*) y orden de presentación (*ORD*), respectivamente. Dentro de los paréntesis se han agrupado los términos aleatorios del modelo: u_{oj} es la estimación de la variabilidad de los interceptos, u_{1j} es la estimación de la variabilidad del coeficiente de *DS* y ε_{ij} es el término de error.

$$IPE_{ij} = b_0 + b_1 DS_{ij} + b_2 DOM_{ij} + b_3 ORD_{ij} + (u_{oj} + u_{1j} DS_{ij} + \varepsilon_{ij}) \quad (5)$$

Se prefirió construir MLJ a realizar ANOVA mixtos de medidas repetidas por dos razones. En primer lugar, en los MLJ es posible especificar la estructura anidada de los datos. Por otra parte, estos modelos son más robustos que el ANOVA con respecto a dos fuentes potenciales de error: las violaciones al supuesto de esfericidad (i.e., la igualdad de varianzas de todos los pares posibles de niveles de las variables intra-sujeto) y la inflación del error tipo I debido a la subestimación del error estándar de la media (Quené & van den Bergh, 2004). Para realizar estos análisis, se utilizó el paquete *lme4* (versión 1.1-21; Bates, et al., 2019) del entorno R (versión 3.4.2, R Core Team, 2017).

Como primer paso, se verificó el efecto global de IPE a partir de las diferencias entre las estimaciones de comprensión de cada tema. Posteriormente, en el análisis principal se construyeron modelos sucesivos para incluir los efectos de las variables independientes. En el primer modelo, la deseabilidad social de los temas asignados a cada participante fue introducida como un factor predictor de la IPE; en los modelos subsiguientes, se agregaron como efectos fijos el dominio de conocimiento, el orden de presentación de los temas (i.e., deseabilidad baja-deseabilidad alta o a la inversa) y las interacciones entre estas variables (ver Tabla 3).

IPE y deseabilidad social del conocimiento. Replicando el fenómeno de IPE, se observó que las estimaciones de comprensión previas a la elaboración de explicaciones fueron mayores que las posteriores $M_{pre\text{-}post}$= .61, *EE*=.09, I.C. 95%= [.42, .79], *t*(367)=6.45, *p*<.01, *d*=.33. La tabla 2 muestra las diferencias observadas entre las estimaciones pre y post para cada uno de los temas del estudio principal.

Tabla 2.

Magnitud de la IPE para cada tema del Experimento 1.

Dominio	Tema	Deseabilidad	*M* (*EE*)	I. C. 95 %
Histórico	Creación de las FARC-EP	Baja	.46 (.18)	[.10, .83]
	Prolongación del conflicto armado	Alta	1.24 (.21)	[.81, 1.66]
Económico	Bolsa de Valores	Baja	.04 (.152)	[-.37, .47]
	Inflación	Alta	1.24 (.19)	[.86, 1.62]
Dispositivos	Turbina jet	Baja	-.56 (.19)	[-1.05, -.08]
	Caña de pescar	Alta	1.21 (.25)	[.71,1.72]

EE = Error estándar; I. C.= Intervalo de confianza de la media.

Por otra parte, la deseabilidad social interactuó con la diferencia entre estimaciones pre y post, $\chi^2(5) = 45.97$, $p<.001$[4] (ver Tabla 3, Modelo 1). Específicamente, la disminución

[4] El ajuste de los MLJ puede compararse utilizando las transformaciones logarítmicas de la función de verosimilitud (*log-likelihood*) del modelo actual y del modelo de comparación, de acuerdo con la siguiente ecuación (Field, Miles, & Field, 2012, p.868):

$$\chi^2_{cambio} = -2(log\text{-}likelihood_{comparación} - log\text{-}likelihood_{actual})$$

Este índice de comparación sigue una distribución chi- cuadrado con grados de libertad iguales a:

$$gl = Número\ de\ parámetros_{comparación} - Número\ de\ parámetros_{actual}$$

en las estimaciones de comprensión después de la formulación de explicaciones se observó únicamente en los temas de deseabilidad alta (ver Figura 7). Por otra parte, se encontraron diferencias significativas en la magnitud de la IPE entre dominios de conocimiento, $\chi^2(7) = 6.00$, $p<.05$. En particular, la magnitud de la IPE fue menor en el dominio de los dispositivos, $b= -.53$, $t(181)= -2.43$, $p<.05$ (ver Tabla 3, Modelo 2). Estas diferencias serán interpretadas a la luz de la interacción entre deseabilidad social y dominio de conocimiento. Aunque el orden de presentación no tuvo un efecto principal sobre la IPE (ver Tabla 3, Modelo 3), este factor interactuó con la deseabilidad social y el dominio de conocimiento, como se verá más adelante.

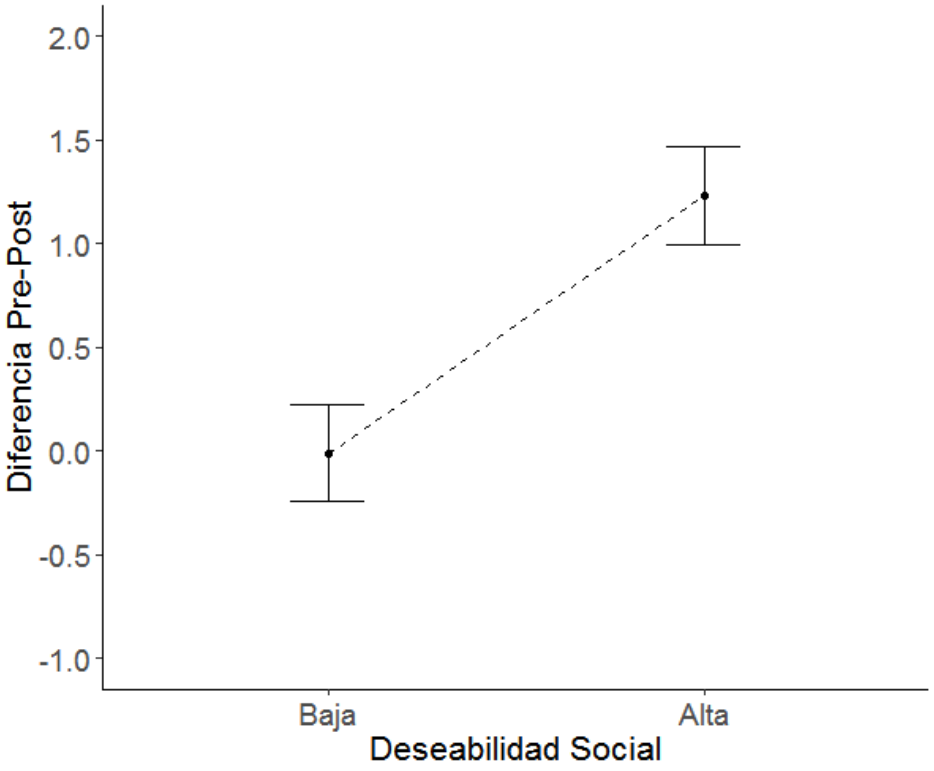


Figura 7. Diferencias en la magnitud de la IPE entre los temas de deseabilidad alta y baja (Experimento 1). Las barras de error representan los intervalos de confianza de la media al 95%.

Interacción entre deseabilidad social y dominio de conocimiento. Se encontró una interacción significativa entre la deseabilidad social y el dominio de conocimiento, b=1.01, t(181)=2.32, p<.05 (ver Tabla 3, Modelo 4). Específicamente, mientras que en el tema de deseabilidad alta en el dominio de los dispositivos (caña de pescar) se evidenció la IPE, en el tema de deseabilidad baja (motor de turbina jet) se encontró el efecto inverso: las estimaciones de comprensión posteriores a las explicaciones fueron más altas que las previas (ver Tabla 2).

Tabla 3.

Modelos lineales jerárquicos de las relaciones entre deseabilidad social, dominio de conocimiento, orden de presentación e IPE (Experimento 1).

Modelo	Predictores	*b*	*t*	*gl*	*r*	AIC	BIC
Modelo 0	Intercepto	.61	6.45**	184	.43	1491.41	1507.04
Modelo 1	Intercepto	-.01	-.08	183	.01	1447.43	1466.97
	Deseabilidad	1.24	6.97**	183	.46		
Modelo 2	Intercepto	.23	1.31	183	.10	1445.43	1472.79
	Deseabilidad	1.24	7.01**	183	.46		
	Dominio (Dispositivos)	-.53	-2.43*	181	.18		
	Dominio (Económico)	-.21	-.97	181	.07		
Modelo 3	Intercepto	.26	1.32	183	.10	1147.31	1478.58
	Deseabilidad	1.24	7.01**	183	.46		
	Dominio (Dispositivos)	-.53	-2.43*	180	.18		

	Dominio (Económico)	-.21	-.97	180	.07		
	Orden	-.06	-.33	180	.02		
Modelo 4	Intercepto	.50	2.14*	181	.16	1445.81	1484.89
	Deseabilidad	.77	2.54*	181	.19		
	Dominio (Dispositivos)	-1.03	-3.37**	180	.24		
	Dominio (Económico)	-.42	-1.37	180	.10		
	Orden	-.06	-.34	180	.03		
	Deseabilidad*Dominio (Dispositivos)	1.01	2.32*	181	.17		
	Deseabilidad*Dominio (Económico)	.42	.97	181	.07		
Modelo 5	Intercepto	.95	3.99**	180	.29	1419.19	1462.19
	Deseabilidad	-.14	-.42	180	.03		
	Dominio (Dispositivos)	-1.03	-3.50**	180	.25		
	Dominio (Económico)	-.42	-1.43	180	.11		
	Orden	-.98	-4.06**	180	.29		
	Deseabilidad*Dominio (Dispositivos)	1.01	2.41*	180	.18		
	Deseabilidad*Dominio (Económico)	.42	1.01	180	.08		
	Deseabilidad*Orden	1.83	5.39**	180	.37		
Modelo 6	Intercepto	.67	2.35*	180	.17	1410.33	1468.95
	Deseabilidad	.45	1.11	176	.08		

Dominio (Dispositivos)	-.41	-.99	180	.07
Dominio (Económico)	-.19	-.47	180	.04
Orden	-.42	-1.03	180	.08
Deseabilidad*Dominio (Dispositivos)	-0.52	-.89	176	.07
Deseabilidad*Dominio (Económico)	.13	.22	176	.02
Deseabilidad*Orden	.64	1.12	176	.08
Deseabilidad*Dominio (Dispositivos)*Orden	1.81	3.11**	176	.23
Deseabilidad*Dominio (Económico)*Orden	.13	.22	176	.02

Nota: *=p<.05; **=p<.01; gl=grados de libertad.

Interacción entre orden de presentación y deseabilidad social. Se obtuvo una interacción inesperada entre el orden de presentación y la deseabilidad de los temas, $\chi^2(11)=28.61$, $p< .001$ (ver Modelo 5, Tabla 3). En particular, cuando los temas menos deseables fueron explicados por los participantes antes que los temas más deseables, la magnitud de la IPE fue similar en los temas de alta y baja deseabilidad (ver Figura 8).

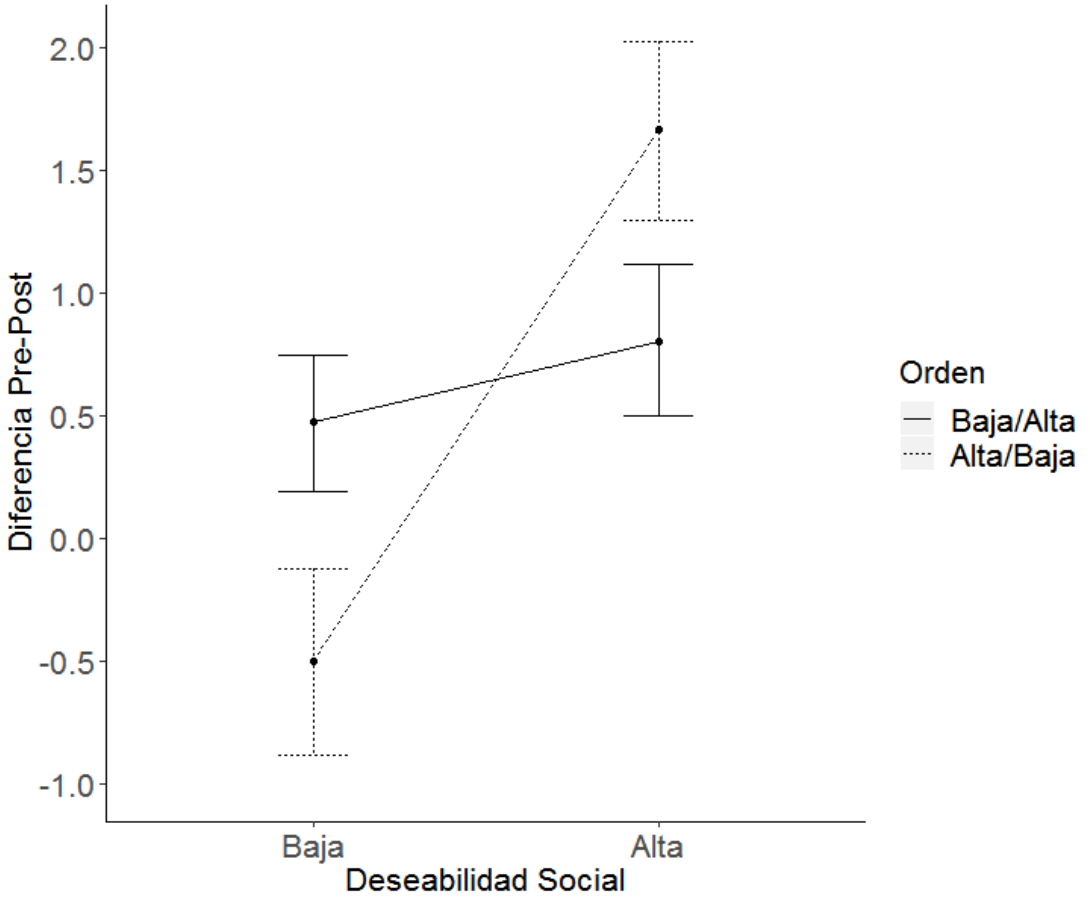


Figura 8. Diferencias en la magnitud de la IPE en función de la deseabilidad social y el orden de presentación de los temas. Las barras de error representan los intervalos de confianza de la media al 95%.

Interacción entre orden de presentación, dominio de conocimiento y deseabilidad social. El análisis de la interacción hallada entre el dominio de conocimiento, la deseabilidad de los temas y el orden de presentación, $\chi^2(15) = 16.86$, $p<.001$ (ver Tabla 3, Modelo 6), reveló que el patrón descrito en la Figura 8 se presentó principalmente en el dominio de dispositivos (ver Tabla 4). En particular, cuando el tema menos deseable de este dominio (i.e., motor de turbina jet) fue explicado en primer lugar, las estimaciones posteriores de comprensión del tema más deseable (i.e., caña de pescar) fueron similares a las iniciales

(i.e., se eliminó el efecto de IPE). En contraste, cuando el tema de alta deseabilidad fue explicado inicialmente, la estimación de la comprensión posterior del tema menos deseable fue mayor a la inicial (i.e., el valor de la IPE fue negativo).

Tabla 4.

Diferencias en la magnitud de la IPE según el dominio de conocimiento, orden de presentación y deseabilidad social de los temas.

Dominio	Orden de presentación	Deseabilidad social del tema	*M* (*EE*)	I. C. 95%
Histórico	Bajo/Alto	Baja	.67 (.23)	[.19, 1.16]
		Alta	1.12 (.30)	[.50, 1.75]
	Alto/Bajo	Baja	.25 (.27)	[-.30, .82]
		Alta	1.35 (.29)	[.75, 1.95]
Económico	Bajo/Alto	Baja	.48 (.25)	[-.03, 1]
		Alta	1.06 (.23)	[.59, 1.53]
	Alto/Bajo	Baja	-.38 (.27)	[-1.06, .28]
		Alta	1.41 (.29)	[.80, 2.03]
Dispositivos	Bajo/Alto	Baja	.26 (.25)	[-.25, .78]
		Alta	.20 (.24)	[-.29, .69]
	Alto/Bajo	Baja	-1.41 (.35)	[-2.12, -.67]
		Alta	2.23 (.36)	[1.49, 2.97]

Nota: *M* = Media; *EE*= Error estándar. I. C.=Intervalo de confianza de la media.

Experimento 2

Plan de análisis. En el Experimento 2 se utilizó un diseño experimental mixto de 2x2, con la deseabilidad social del conocimiento como variable entre-sujetos; los dos temas históricos presentados en cada condición experimental fueron tratados como variables intra-sujeto. La variable dependiente fue la diferencia en las estimaciones de comprensión sobre cada tema, antes y después de escribir las explicaciones. Las puntuaciones en la prueba de *overclaiming* se incluyeron como covariadas.

El propósito de este experimento fue replicar el efecto de la deseabilidad social sobre la IPE encontrado en el Experimento 1, utilizando un diseño entre-sujetos. Adicionalmente, se buscó verificar que este efecto no fuese producido por un sesgo deliberado de auto-promoción. Al igual que en el análisis de resultados del Experimento 1, se construyeron MLJ en los cuales se introdujeron como medidas repetidas (factor aleatorio) la relación entre las estimaciones realizadas por cada participante. En el modelo inicial, la deseabilidad social de los temas asignados fue introducida como factor entre-sujetos; en los modelos posteriores, se adicionaron las estimaciones de deseabilidad social realizadas en la primera sesión como factor fijo y los puntajes en la prueba de *overclaiming* como covariada.

IPE y deseabilidad social. Al igual que en el Experimento 1, se encontró que las estimaciones de comprensión previas a la elaboración de explicaciones fueron mayores que las posteriores, $M_{pre\text{-}post}$=.59, *EE*=.08, I.C. 95% [.42, .76], $t(313)=7.02$, $p < .001$, d=.39. La tabla 5 muestra las diferencias entre las estimaciones pre y post para cada uno de los temas del estudio principal. La IPE se observó tanto en los temas de alta como de baja

deseabilidad social, es decir, las diferencias entre las estimaciones previas y posteriores fueron mayores que cero en todos los temas.

Tabla 5.

Magnitud de la IPE en cada uno de los temas del Experimento 2.

Deseabilidad	Tema	*M* (*EE*)	I. C. 95%
Baja	Finalización de la dictadura de Gustavo Rojas Pinilla	.75 (.16)	[.43-1.08]
	Financiación del Plan Colombia	.40 (.17)	[.05-.75]
Alta	Prolongación del conflicto armado	.47 (.16)	[.13-.81]
	Mayor intensidad del conflicto en las zonas rurales	.74 (.17)	[.40-1.08]

Nota: *M*=Media; *EE*=Error estándar; I. C.=Intervalo de confianza de la media.

Con el fin de corroborar la asociación entre la deseabilidad social del conocimiento e IPE, se construyó un modelo en el que la condición experimental (alta o baja deseabilidad) fue incluida como factor predictor de la IPE (ver Tabla 6, modelo 1). Este efecto no fue significativo, $\chi^2(4) = .02$, p=.88, debido a que la magnitud de la IPE fue similar en los temas de las dos condiciones experimentales (ver Tabla 5). Una posible razón de este resultado es que los puntajes en los ítems dentro de cada nivel de deseabilidad fueron diferentes y, al agruparlos como factores, se perdía parte de la variabilidad en la variable de deseabilidad. Por este motivo, se elaboró un modelo alternativo en el que se introdujo como factor predictor la puntuación en la escala de deseabilidad social que fue aplicada dos semanas antes de la tarea principal (ver Tabla 6, modelo 2). Se encontró que los puntajes de deseabilidad social del conocimiento sobre los temas asignados se asociaron

con mayores magnitudes de IPE, χ^2 (5) = 5.18, p< .05. La correlación entre las estimaciones de deseabilidad realizadas en la primera y la segunda sesión fue positiva, r=.62, t(312)=14.27, p<.001, lo cual demuestra que las estimaciones de deseabilidad social de cada participante sobre cada tema fueron consistentes entre las dos sesiones. Las estimaciones de familiaridad y de deseabilidad social registradas durante la primera sesión también correlacionaron positivamente, r=.44, t(312)=8.70, p<.001. Por esta razón, la familiaridad fue excluida de los MLJ implementados para prevenir problemas asociados a la multicolinealidad.

IPE, deseabilidad social y *overclaming*. Para verificar que el efecto de la deseabilidad social sobre la IPE no fuese producto de un sesgo de auto-promoción, se incluyeron las puntuaciones en la prueba de *overclaiming* aplicada en la primera sesión. Siguiendo la técnica de calificación propuesta por Paulhus et al. (2003), se calculó para cada participante la proporción de respuestas de familiaridad mayores que uno dadas a cada uno de los cuatro eventos ficticios relacionados con el conflicto. Este parámetro se denomina tasa de falsa alarma (*False Alarm Rate*; TFA en adelante) Así, si un participante reportaba no estar informado sobre alguno de los eventos falsos, su TFA sería igual a cero; en contraste, cuando el participante afirmaba estar muy informado sobre estos cuatro eventos, su TFA era igual a uno. Adicionalmente, se calculó la tasa de aciertos (*Hit Rate*; HR en adelante) a partir de la proporción de juicios de familiaridad mayores que uno dadas a los 16 hechos reales del conflicto colombiano. A partir de estos dos valores, se calcularon los índices de precisión (HR-TFA) y sesgo ([HR + TFA] /2) para cada participante.

Al incluir estas variables en el análisis principal, se encontró una relación positiva entre los puntajes de TFA y la magnitud de la IPE, χ^2(6) = 5.71, p< .05 (ver Tabla 6,

Modelo 3). En otras palabras, las personas que tuvieron más dificultades para distinguir entre los eventos falsos y verdaderos del conflicto armado también sobrestimaron más su comprensión de los temas relacionados con este último. Es importante destacar que la asociación entre deseabilidad social e IPE fue significativa después de introducir las puntuaciones de TFA, lo cual implica que dicha asociación no puede ser explicada como resultado de un sesgo de auto-promoción. Al incluir los índices de precisión y sesgo como predictores de la IPE, no se incrementó significativamente el ajuste de los modelos.

Tabla 6.

Modelos lineales jerárquicos de las relaciones entre deseabilidad social, overclaiming e IPE (Experimento 2).

Modelo	Predictores	*b*	*t*	*gl*	*r*	AIC	BIC
Modelo 0	Intercepto	.59	6.39	157	.45	1146.41	1161.40
Modelo 1	Intercepto	<.01	<.01	157	-	1148.39	1167.13
	Deseabilidad (condición experimental)	.02	.14	155	0.01		
Modelo 2	Intercepto	-.60	-.20	156	.02	1143.22	1161.97
	Deseabilidad (estimación previa)	.14	2.28*	156	.18		
Modelo 3	Intercepto	-.19	-.65	156	.05	1139.51	1162.01
	Deseabilidad (estimación previa)	.13*	2.09	156	.17		
	TFA	1.62*	2.39	155	.19		

Modelo 4	Intercepto	-.87	-2.25	155	.18	1134.32	1160.56
	Deseabilidad (estimación previa)	.28**	3.39	155	.26		
	TFA	7.5**	3.29	155	.26		
	Deseabilidad (estimación previa)*TFA	-1.27**	-2.69	155	.21		

Nota: *=p<.05; **=p<.01; TFA=Tasa de falsa alarma.

Por último, se encontró una interacción significativa entre la deseabilidad social del conocimiento y el TFA como predictores de la IPE, $\chi^2(7) = 7.19$, $p< .01$. Específicamente, las personas que reportaron estar más informadas sobre eventos ficticios del conflicto colombiano tendieron a sobrestimar por igual su comprensión de los temas de alta y baja deseabilidad; por su parte, quienes reportaron correctamente no tener información sobre dichos eventos exhibieron una IPE mayor en los temas más deseables en comparación con los menos deseables (ver Figura 9). En este estudio, el orden de presentación de los temas en las condiciones experimentales y cursar asignaturas universitarias sobre historia de Colombia no fueron predictores significativos de la IPE.

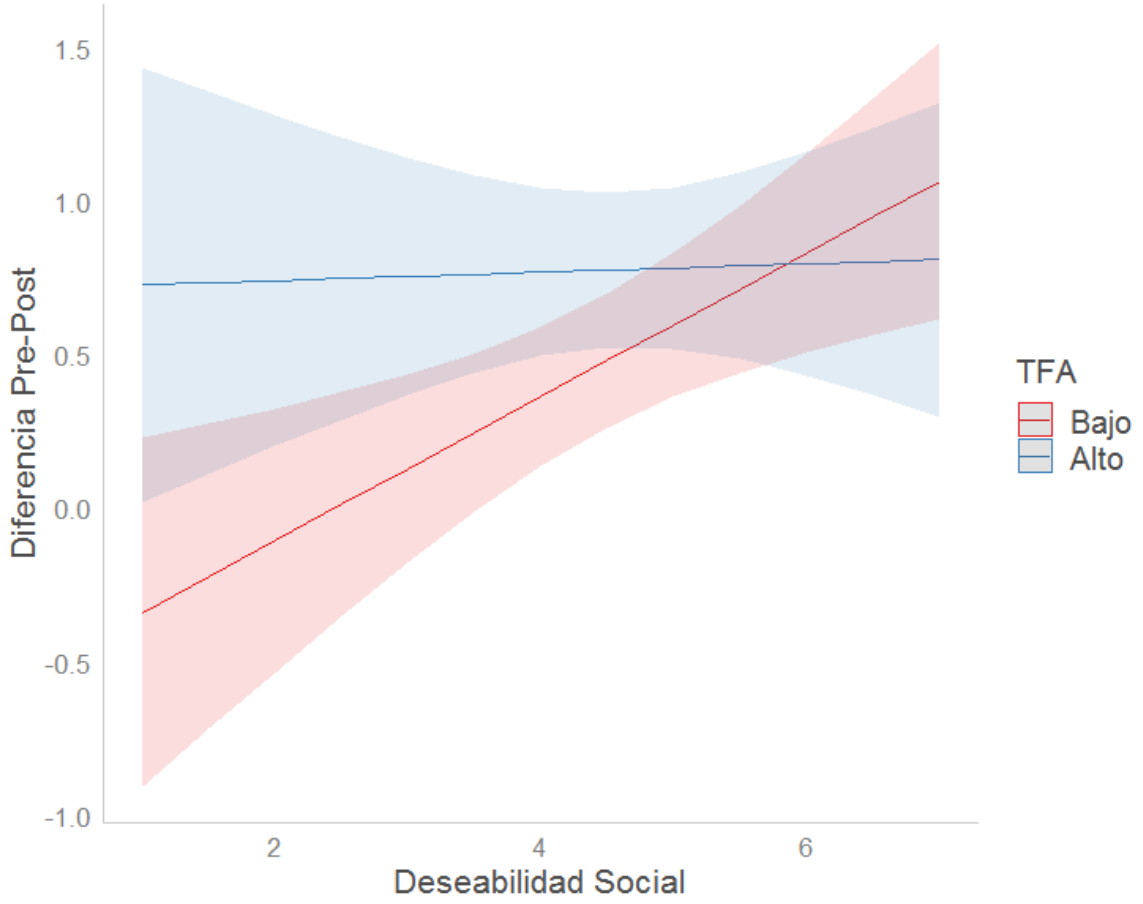


Figura 9. Interacción entre la deseabilidad social y TFA como predictores de la IPE. Nota: El nivel de TFA "Bajo" incluye todos los participantes que tuvieron puntajes por debajo de la media en este índice, mientras que el nivel "Alto" corresponde a quienes tuvieron puntajes por encima de la media.

Experimento 2a

Plan de análisis. El propósito de este estudio fue realizar una réplica del Experimento 2 introduciendo dos modificaciones: primero, se amplió el conjunto de temas que podían ser asignados a los participantes dentro de cada condición; en segundo lugar, cada participante explicó y estimó nuevamente su comprensión sobre un único tema, con el fin de evitar artefactos producidos por la exposición repetida al paradigma experimental.

Debido a que este diseño no fue anidado, se construyeron modelos de regresión lineal múltiple por pasos en lugar de MLJ. Las variables introducidas en dichos modelos fueron idénticas a las reportadas en el Experimento 2.

IPE y deseabilidad social. Replicando los resultados de los Experimentos 1 y 2, las estimaciones de comprensión previas a la elaboración de explicaciones fueron mayores que las posteriores $M_{pre\text{-}post}$=.72, *EE*=.10, I.C. 95% [.51, .93], $t(188)=6.93$, $p < .001$, d=.50. La tabla 7 presenta muestra las diferencias entre las estimaciones de comprensión de cada tema del estudio principal, antes y después de explicarlos. La IPE se observó en todos los temas de alta deseabilidad social y en dos de los temas menos deseables (i.e. parapolítica y genocidio de la Unión Patriótica).

Tabla 7.

Magnitud de la IPE en cada uno de los temas del Experimento 2a.

Deseabilidad	Tema	*M* (*EE*)	I.C. 95%
Baja	Alianzas entre políticos y paramilitares	.64 (.27)	[.08, 1.19]
	Creación del movimiento a favor de la Constituyente de 1991	.45 (.26)	[-.09, 1.01]
	Genocidio de la Unión Patriótica	.68 (.31)	[.03, 1.33]
	Financiación del Plan Colombia	.24 (.35)	[-.48, .96]
Alta	Creación de las FARC-EP	1.43 (.30)	[.81, 2.05]

Creación de grupos paramilitares	.54 (.23)	[.05, 1.03]
Prolongación del conflicto armado	1.04 (0.25)	[.51, 1.56]
Intensidad del conflicto en las zonas rurales	.78 (.32)	[.11, 1.44]

Nota: *M*=Media; *EE*=Error estándar; I. C.=Intervalo de confianza de la media al 95%.

Se encontró un efecto significativo de la condición experimental (ver Tabla 8, Modelo 1), de modo que la magnitud de la IPE fue mayor en los temas de deseabilidad alta, *M*=0.96, *EE*=.14, I. C. 95 % [.67, 1.24], que en los temas de deseabilidad baja, *M*=0.50, *EE*=.15, I. C. 95% [.20, .80]. Confirmando este efecto, las estimaciones de deseabilidad social sobre el tema asignado realizadas por los participantes al final de la sesión se asociaron positivamente con la magnitud de la IPE, $F(1,187)=4.57$, $p<.05$. La consistencia interna de esta escala en los ocho temas seleccionados estuvo dentro de un rango adecuado (α de Cronbach≥.7; ver Anexo I). Adicionalmente, se encontró que la media de deseabilidad social atribuida a los temas de la condición de deseabilidad alta, *M*=4.96, *EE*=.04, I. C. 95 % [4.87, 5.04], fue superior a la de los temas de deseabilidad baja, *M*=4.26, *EE*=.05, I. C. 95% [4.16, 4.35], $t(755)=16.67$, $p<.01$, $d=.53$ (ver Anexo I). Este hallazgo puede ser interpretado como una verificación de la manipulación experimental *(manipulation check*) de la deseabilidad social. Los índices derivados de las pruebas de *overclaiming* (i.e., TFA, HR, Precisión y Sesgo) y cursar asignaturas universitarias sobre historia de Colombia no incrementaron el ajuste de los modelos significativamente.

Tabla 8.

Modelos lineales jerárquicos de las relaciones entre deseabilidad social e IPE (Experimento 2a).

Modelo	Predictores	*b*	*t*	*gl*	*r*	$R^2_{adj.}$
Modelo 0	Intercepto	.72	6.93	189	.45	
Modelo 1	Intercepto	.5	3.44**	187	.24	.02*
	Deseabilidad (condición experimental)	.45	2.21*	187	.16	

Nota: *gl*=grados de libertad. *=*p*<.05; **=*p*<.01; $R^2_{adj.}$=R cuadrado ajustado.

Experimento 3

Plan de análisis. En este experimento se implementó un diseño mixto de 2x2, en el cual la asignación de propuestas legislativas de alta o baja deseabilidad social fue la variable entre-sujetos. Las dos propuestas legislativas presentadas en cada condición fueron tratadas como variable intra-sujeto. La variable dependiente fue la diferencia entre las estimaciones de comprensión sobre los temas asignados, las cuales fueron solicitadas antes y después de explicarlos. El objetivo de este experimento fue establecer si la habilidad para realizar procesamientos tipo 2 modera la relación entre la IPE y la deseabilidad social del conocimiento. Debido a que se utilizó un diseño anidado similar al de los Experimentos 1 y 2, se construyeron MLJ que incluyeron como factor aleatorio la relación entre las estimaciones realizadas por cada participante. En el modelo inicial, se incluyó la condición experimental como factor fijo, mientras que en los modelos subsiguientes se agregaron

como efectos fijos la deseabilidad social estimada, la puntuación en el CRT y la interacción entre estas dos variables.

IPE y deseabilidad social. Replicando los resultados de los experimentos anteriores, se encontró que los juicios iniciales de comprensión fueron mayores que los posteriores a la elaboración de explicaciones de las propuestas políticas $M_{pre\text{-}post}$=.74, *EE*=.13, I.C. 95%=[.48, 1.00], $t(139)=5.58$, $p < .001$, d=.47. El efecto del grupo experimental (deseabilidad alta o baja) sobre la IPE no fue significativo, $\chi^2(5)=2.54$, p=.11. Al igual que en el Experimento 2, una posible explicación de este hallazgo es que los promedios globales de deseabilidad social de las propuestas legislativas seleccionadas (ver Anexo L) no dan cuenta de las diferencias individuales en esta variable. Como se indica en la Tabla 9, las diferencias entre las puntuaciones pre y post fueron positivas en los dos temas de deseabilidad alta (según lo esperado) y en uno de los temas de deseabilidad baja. Específicamente, el efecto de IPE se evidenció en el proyecto de ley sobre derechos laborales de actores y actrices, aunque este ítem tuvo una media relativamente baja en la escala de deseabilidad social inicial (ver Anexo L).

Tabla 9.

Medias de las diferencias entre las estimaciones previas y posteriores para cada uno de los temas del Experimento 2.

Deseabilidad	Tema	*M* (*EE*)	I. C. 95%
Alta	Acuerdo sobre víctimas y jurisdicción especial para la paz.	1.20 (.31)	[.57, 1.84]

	Acuerdo sobre solución al problema de las drogas ilícitas.	.79 (.22)	[.33, 1.25]
Baja	Ley de derechos laborales para actores y actrices.	.66 (.27)	[.10, 1.22]
	Reglas de convalidación de títulos extranjeros de cirugía plástica.	.33 (.23)	[-.13, .79]

Nota: *M* = Media; *EE* = Error estándar; I. C.=Intervalo de confianza de la media.

Se construyó un modelo alternativo en el cual se incluyó como factor predictor la puntuación en la escala de deseabilidad social del tema asignado (ver Tabla 10, modelo 1). Replicando los resultados de los Experimentos 2 y 2a, se encontró una asociación positiva entre el puntaje en esta escala y la magnitud de la IPE correspondiente, $\chi^2(5) = 5.58$, $p < .05$. En otras palabras, los participantes sobrestimaron más su comprensión de las propuestas legislativas cuyo conocimiento consideraron más deseable socialmente.

Tabla 10.

Modelos lineales jerárquicos de las relaciones entre deseabilidad social, puntuaciones en el CRT e IPE (Experimento 3).

Modelo	Predictores	*b*	*t*	*gl*	*r*	AIC	BIC
Modelo 0	Intercepto	.74	4.68	70	.49	516.47	528.24
Modelo 1	Intercepto	-.70	-.2	69	.02	512.89	527.60
	Deseabilidad (estimación previa)	.22	2.38*	69	.28		

Modelo 2	Intercepto	.06	.15	69	.02	513.16	530.81
	Deseabilidad (estimación previa)	.22	2.38*	69	.28		
	CRT	-.20	-1.31	68	.16		
Modelo 3	Intercepto	-.67	-1.4	68	.17	508.82	529.41
	Deseabilidad (estimación previa)	.41	3.51**	68	.39		
	CRT	.59	1.71	68	.20		
	Deseabilidad (estimación previa) * CRT	-.22	-2.53*	68	.29		

Nota: *p <.01; ** p < .001; *gl*= grados de libertad.

Interacción entre deseabilidad social y CRT como predictores de la IPE.

Aunque la asociación entre los puntajes en el CRT y la IPE no fue significativa, $\chi^2(3)=1.73$, p=.18 (ver Tabla 10, Modelo 2), se encontró una interacción entre la deseabilidad social estimada y dichos puntajes como predictores de la IPE, $\chi^2(7) = 7.81$, p<.01 (ver Tabla 10, Modelo 3). Concretamente, los participantes que tuvieron dificultades para realizar procesamientos tipo 2 (i.e., quienes obtuvieron un CRT igual a 0) sobrestimaron en mayor medida su conocimiento causal sobre los temas más deseables en comparación con los participantes que obtuvieron puntajes más altos en el CRT (ver Figura 10).

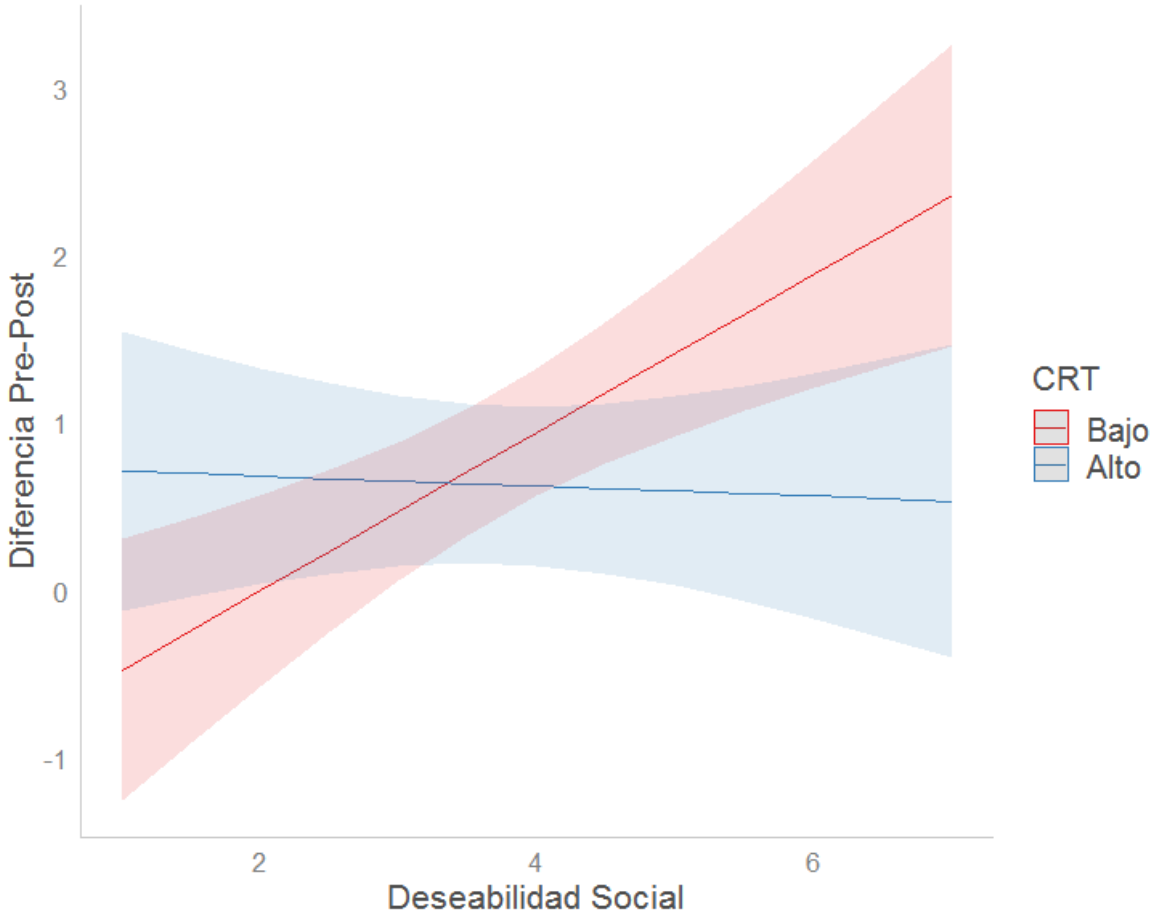


Figura 10. Puntajes en el CRT como moderadores de la asociación entre deseabilidad social del conocimiento y la magnitud de la IPE. Nota: El nivel de CRT "Bajo" incluye todos los participantes que tuvieron puntajes iguales a cero en dicha prueba, mientras que el nivel "Alto" corresponde a quienes tuvieron puntajes mayores de cero.

Experimento 4

Plan de análisis. En el Experimento 4 se utilizó un diseño experimental entre-sujetos de 2x2, donde la deseabilidad social del conocimiento sobre los temas históricos asignados (alta/baja) y el nivel de sobrecarga ejecutiva (alta/baja) fueron manipuladas. Al igual que en los experimentos anteriores, la variable dependiente fue la diferencia en las

estimaciones de comprensión causal antes y después de generar explicaciones sobre el tema asignado. El objetivo de este experimento fue determinar si la capacidad para realizar procesamientos tipo 2, manipulada a través del nivel de sobrecarga sobre el sistema ejecutivo central, afecta la asociación entre deseabilidad social e IPE. El puntaje obtenido en el CRT, el porcentaje de aciertos en las tareas de memoria y el haber cursado asignaturas sobre historia de Colombia fueron incluidos como covariadas en el análisis de resultados. Al igual que en el análisis de resultados del Experimento 2a, se construyeron modelos de regresión lineal múltiple por pasos, introduciendo como predictores las variables mencionadas anteriormente y sus respectivas interacciones.

IPE y deseabilidad social. Al igual que en los experimentos anteriores, se encontró que las estimaciones de comprensión de los temas asignados disminuyeron después de explicarlos, $M_{pre\text{-}post}$= .98, *EE*=.09, I.C. 95%= [.79, 1.16], $t(251)=10.41$, $p < .001$, d=.65. La tabla 11 presenta las diferencias entre las estimaciones pre y post para cada uno de los temas del estudio principal. La IPE se encontró en todos los temas históricos de alta y baja deseabilidad social, excepto en el tema de la prolongación del conflicto armado interno.

Tabla 11.

Magnitud de la IPE en cada uno de los temas del Experimento 4.

Deseabilidad	Tema	*M* (*EE*)	I.C. 95%
Baja	Alianzas entre políticos y paramilitares	1.13 (.30)	[.52, 1.74]
	Creación del movimiento a favor de la Constituyente de 1991	1.27 (.29)	[.67, 1.86]

	Genocidio de la Unión Patriótica	.89 (.21)	[.45, 1.34]
	Financiación del Plan Colombia	.81 (.21)	[.39, 1.24]
Alta	Creación de las FARC-EP	1.56 (.24)	[1.07, 2,06]
	Creación de grupos paramilitares	.60 (.23)	[.13, 1.08]
	Prolongación del conflicto armado	.51 (.32)	[-.14, 1.17]
	Intensidad del conflicto en las zonas rurales	1.03 (.25)	[.52, 1.54]

Nota: *M*=Media; *EE*=Error estándar; I. C.=Intervalo de confianza de la media.

Replicando los resultados de los Experimentos 2 y 3, la condición de deseabilidad social asignada en la forma de grupo no fue un predictor significativo de la IPE, $F(1,250) = .35$, $p = .55$ (ver Tabla 12, Modelo 1). Como se ha indicado anteriormente, este resultado puede deberse a que las diferencias individuales en las estimaciones de la deseabilidad social no pueden ser medidas de manera confiable a través de la asignación experimental de los temas más y menos deseables. Esto es, al usar un diseño de grupos, las diferencias individuales en la estimación de la deseabilidad pueden verse ignoradas en el diseño. Por consiguiente, se construyó un modelo alternativo en el cual se incluyó como factor predictor la deseabilidad social estimada después de realizar el paradigma de IPE. La consistencia interna de esta escala en los ocho temas seleccionados estuvo dentro de un rango aceptable (α de Cronbach ≥.7; ver Anexo Q). La media de deseabilidad social estimada de los temas asignados a la condición de deseabilidad alta, *M*=4.83, *EE*=.04, I. C. 95 % [4.76, 4.91], fue superior a la de los temas de deseabilidad baja, *M*=4.19, *EE*=.05, I.

C. 95% [4.11, 4.28], $t(1.007)=18.31$, $p<.001$, $d=.49$ (ver Anexo Q). Este hallazgo puede ser interpretado como una verificación de la manipulación experimental de la deseabilidad social.

Se encontró una asociación positiva significativa entre la estimación de la deseabilidad social del tema asignado y la IPE, $F(1,250)=3.87$, $p=.05$ (ver Tabla 12, Modelo 2), de modo que los participantes sobrestimaron más su comprensión de los temas históricos cuyo conocimiento juzgaron más deseable socialmente. Este efecto no fue significativo en los modelos subsiguientes que incluyeron el porcentaje de aciertos en las tareas de memoria, los puntajes en el CRT y el haber tomado cursos de historia como covariadas (ver Tabla 12, Modelos 3,4 y 5). Esto indica que, si bien la deseabilidad tiene un efecto, este efecto es absorbido cuando se incluyen en el modelo el desempeño en las tareas de sobrecarga ejecutiva (i.e., el porcentaje de aciertos en las tareas de memoria), el nivel de procesamiento controlado y los conocimientos previos.

Tabla 12.

Modelos lineales jerárquicos de las relaciones entre deseabilidad social e IPE (Experimento 4).

Modelo	Predictores	*b*	*t*	*gl*	*r*	$R^2_{adj.}$
Modelo 0	Intercepto	.98	10.41**	251	.55	
Modelo 1	Intercepto	1.03	7.77**	250	.44	.35
	Deseabilidad (condición experimental)	-.11	-.60	250	.04	

Modelo 2	Intercepto	.38	1.17	250	.07	.01
	Deseabilidad (estimación posterior)	.13	1.87	250	.12	
Modelo 3	Intercepto	3.42	4.23*	249	.26	.07**
	Deseabilidad (estimación posterior)	.11	1.65	249	.10	
	Porcentaje de aciertos (Tareas de memoria)	-3.30	-4.10**	249	.25	
Modelo 4	Intercepto	3.30	4.12	248	.25	.09**
	Deseabilidad (estimación posterior)	.09	1.31	248	.08	
	Porcentaje de aciertos (Tareas de memoria)	-2.83	-3.45**	248	.21	
	CRT	-.23	-2.45*	248	.15	
Modelo 5	Intercepto	3.22	4.06**	247	.25	.11**
	Deseabilidad (estimación posterior)	.08	1.28	247	.08	
	Porcentaje de aciertos (Tareas de memoria)	-2.49	-3.12**	247	.19	
	CRT	-.20	-2.15*	247	.14	
	Curso de historia	-.47	-2.53*	247	.16	

Nota: *gl*=grados de libertad. *=p<.05; **=p<.01; $R^2_{adj.}$=R cuadrado ajustado.

IPE, sobrecarga cognitiva y CRT. El efecto de la condición experimental de sobrecarga ejecutiva en la IPE no fue significativo, $F(1,250) = 2.89$, $p = .09$. No obstante, al igual que en el caso de la deseabilidad social, esto pudo ocurrir porque esta variable dicotómica no brinda información sobre las diferencias individuales en el nivel de sobrecarga inducido por la tarea de memoria asignada. Así, un participante en la condición de sobrecarga baja podría haber experimentado un nivel de sobrecarga ejecutiva similar al de alguien asignado a la condición de sobrecarga alta, en virtud de las diferencias en la capacidad (*span*) de su memoria de trabajo. Es importante recordar que se asignó sobrecarga ejecutiva a ambos grupos siguiendo el diseño de Johnson et al. (2014), ya que en dicho estudio se encontró un efecto significativo de esta manipulación en el desempeño en el CRT. Para obtener un indicador del desempeño individual en las tareas de memoria, se promediaron los porcentajes de aciertos de cada participante en la tarea de memoria visoespacial y en la tarea de control de lectura durante la fase de estimación previa (i.e., la identificación del tema histórico presentado en cada juicio de comprensión). Si la manipulación experimental de la sobrecarga ejecutiva fue efectiva, cabría esperar que el desempeño en estas tareas fuese mayor en el grupo de sobrecarga baja. Corroborando esta hipótesis, se encontró que el porcentaje de aciertos en las pruebas de memoria fue superior en la condición de sobrecarga baja, $M = .97$, $EE = .005$, en comparación con el grupo de sobrecarga alta, $M = .82$, $EE = .009$, $t(191) = 12.59$, $p < .001$.

El porcentaje de aciertos en las tareas de memoria se asoció negativamente con la magnitud de la IPE, $F(1, 249) = 34.52$, $p < .001$ (ver Tabla 12, Modelo 3). En otras palabras, la IPE fue menor en los participantes que tuvieron mejores desempeños en las tareas de memoria, lo cual fue más común en la condición de sobrecarga baja. Aunque la condición

experimental de sobrecarga ejecutiva no tuvo un efecto directo sobre la IPE, es posible que la relación entre estas variables estuviese mediada por el desempeño en las tareas de memoria. Para explorar esta alternativa, se elaboraron dos modelos de ecuaciones estructurales (Gana & Broc, 2019) utilizando el paquete lavaan (Rosseel, 2012) del entorno R. En el primer modelo se incluyeron de manera separada los efectos de la sobrecarga ejecutiva asignada y del porcentaje de aciertos en las tareas de memoria sobre la IPE (ver Figura 11, panel A). En el segundo modelo, el efecto de la condición experimental de sobrecarga sobre la IPE está mediado por el porcentaje de aciertos en las tareas de memoria (ver Figura 11, panel B). En el primer modelo, se encontró una asociación significativa entre el porcentaje de aciertos en las tareas de memoria y la IPE, $\beta = -.31$, $p < .001$, I. C. 95%, [.46;-.16], pero no entre la condición experimental de sobrecarga y la IPE, $\beta = -.09$, $p > .05$. No obstante, en el segundo modelo se encontró una relación significativa entre la condición experimental de sobrecarga y la IPE, la cual está mediada completamente por el desempeño en las tareas de memoria, $\beta = .19$, $p < .001$, I. C. 95% [.10; .29]. Los índices de ajuste del modelo de mediación (AIC=396.34; BIC=413.99) fueron mejores que los del modelo de efectos directos (AIC=902.17; BIC=912.75)

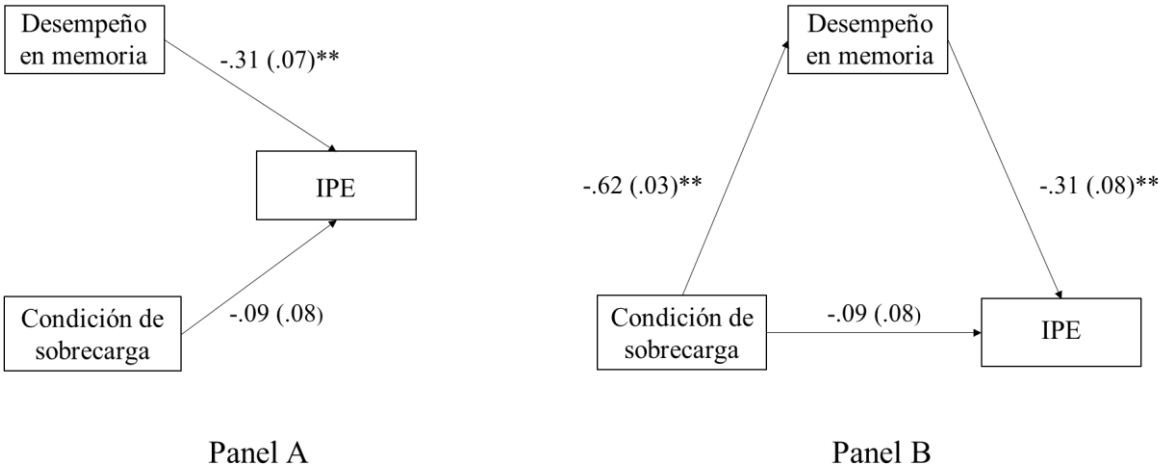


Figura 11. Modelos de ecuaciones estructurales de efectos directos (Panel A) y efecto mediado (Panel B) de la sobrecarga ejecutiva y el desempeño en las tareas de memoria del Experimento 4.

Por otro lado, se encontró una asociación negativa entre los puntajes del CRT y la IPE, $F(1, 249)=6.15$, $p<.05$ (ver Tabla 12, Modelo 4), de manera que quienes tuvieron mejor desempeño en esta prueba estimaron de manera más precisa su conocimiento sobre el tema histórico asignado. Finalmente, quienes habían cursado o se encontraban cursando asignaturas universitarias sobre historia de Colombia demostraron menores diferencias entre sus estimaciones de comprensión antes y después de explicar los temas asignados, $F(1, 249)=6.45$, $p<.05$. Ninguna de las interacciones entre estas variables incrementó significativamente el ajuste de los modelos.

Discusión

Hallazgos principales

Los cinco experimentos reportados en esta disertación fueron diseñados para cumplir dos objetivos: en primer lugar, determinar si las personas utilizan la deseabilidad social del conocimiento sobre un tema como clave para estimar su habilidad para explicarlo. Segundo, verificar si este uso inferencial de la deseabilidad social en la estimación de la comprensión causal es facilitado, inhibido o no está relacionado con la capacidad para realizar procesamientos analíticos/controlados (tipo 2).

Confirmando la hipótesis 1 (ver Introducción), se encontró que el efecto de IPE ocurrió solamente en los temas de alta deseabilidad en los dominios histórico, económico y de dispositivos (Experimento 1). Esto indica que la deseabilidad tiene un rol importante en la estimación de la comprensión causal, lo cual es consistente con la literatura previa que sugiere que la estimación de juicios metacognitivos se basa en claves heurísticas de diverso tipo, algunas de ellas asociadas a información social (e.g., De Carvalho Filho & Yuzawa, 2001; Karabenick 1996; Zeveney & Marsh, 2016). Este resultado también es consistente con los hallazgos de Fisher y Keil (2016), los cuales muestran que la magnitud de la IPE se asocia a las expectativas que las personas tienen sobre lo que deberían conocer de acuerdo con su área de experticia. En síntesis, la IPE no es sólo un producto del contenido esquemático de las teorías intuitivas de nivel cognitivo, sino también del proceso metacognitivo en virtud del cual las personas estiman su comprensión sobre mecanismos específicos. Específicamente, la atención a claves inferenciales como la deseabilidad social puede ser un “atajo metacognitivo” usado para solventar la complejidad inherente al

proceso de monitoreo del conocimiento disponible y relevante para la explicación de un mecanismo concreto.

En el Experimento 2, se encontró que los participantes exhibieron una mayor IPE sobre los temas a los cuales asignaron mayores valores en la escala de deseabilidad social. No obstante, cuando este factor se determinó a partir la asignación a grupos de alta y baja deseabilidad, no se asoció de manera significativa con la IPE. Este hallazgo pone de manifiesto la necesidad de construir medidas sensibles a las diferencias individuales en la deseabilidad social atribuida o en otras claves inferenciales asociadas a los juicios metacognitivos, especialmente en los casos en los que dichas variables no pueden ser manipuladas experimentalmente con los métodos tradicionales de la investigación sobre juicios metacognitivos (e.g., Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992).

Adicionalmente, los resultados del Experimento 2 mostraron que las personas que en la prueba de *overclaiming* (Paulhus et al., 2003) afirmaban estar informadas sobre hechos del conflicto que nunca ocurrieron, sobrestimaron su comprensión sobre los temas más y menos deseables en dicho dominio. En contraste, quienes reportaron no estar al tanto de estos acontecimientos ficticios tendieron a sobrestimar en mayor medida los temas más deseables. Este hallazgo es novedoso, ya que relaciona dos métodos alternativos para medir la precisión en la estimación del propio conocimiento: el *overclaiming* y la IPE. Si las personas con la habilidad para distinguir entre sucesos reales y ficticios tienden a fijarse más en claves heurísticas como la deseabilidad social al estimar su comprensión causal, es razonable suponer que el efecto de la deseabilidad sobre la IPE no se debe a un esfuerzo deliberado para mostrar un atributo positivo del cual se carece. Por el contrario, dicho efecto se deriva del uso de información social en los procesos de estimación subjetiva de

conocimiento. Ante la correlación hallada entre los juicios de conocimiento propios de las pruebas de *overclaiming* (e.g., "Estoy muy bien informado sobre el proceso de negociación y desmovilización de grupos paramilitares en Colombia que tuvo lugar entre 2003 y 2006.") y los juicios de comprensión causal del paradigma de IPE (e.g., "Comprendo de manera detallada las causas de la creación de grupos paramilitares durante la década de 1980."), surge el interrogante de si ambos tipos de juicios dependen de claves heurísticas similares o distintas. Según la hipótesis de mecanismo único, una misma clave inferencial disponible en el ambiente de tarea o en la memoria de trabajo puede afectar distintos juicios metacognitivos (Dunlosky & Tauber, 2014). Algunos resultados experimentales parecen respaldar esta predicción. Por ejemplo, se ha encontrado que la precisión de los juicios de comprensión depende de la accesibilidad de la información (Dunlosky, Rawson, & Middleton, 2005) y que el efecto de dicha clave inferencial está condicionado a la familiaridad con los contenidos presentados (Koriat & Levy-Sadot, 2001). En este sentido, es razonable suponer que los juicios de familiaridad y comprensión causal estén relacionados. No obstante, solo la investigación empírica sistemática permitirá establecer de qué manera se articulan estos mecanismos.

El Experimento 2a replicó el efecto de la deseabilidad sobre la IPE en un conjunto más amplio de temas históricos. Sin embargo, no se encontraron efectos de los puntajes en el test de *overclaiming* ni interacciones entre esta variable y la deseabilidad social. Este resultado pudo deberse a que, en el Experimento 2a, los participantes respondieron la prueba de *overclaiming* después de realizar el paradigma de IPE. Esta decisión metodológica se tomó con el fin de que la mención de hechos históricos recientes en dicha prueba no provocara efectos de facilitación en las estimaciones iniciales de comprensión

causal. No obstante, la exposición previa al paradigma de IPE pudo haber producido estimaciones más conservadoras de la familiaridad sobre temas relacionados con el conflicto en la prueba de *overclaiming*, a causa de experiencias metacognitivas de baja accesibilidad o baja fluidez de procesamiento generadas durante la elaboración de las explicaciones.

Corroborando la hipótesis 2 (ver Introducción), en el Experimento 3 se encontró que las personas que pueden realizar procesamientos tipo 2 estimaron de manera más precisa su conocimiento causal sobre los temas más deseables socialmente. Este resultado aporta nueva evidencia de la existencia de rutas de procesamiento automático y controlado en el nivel metacognitivo y extiende el trabajo de Fernbach, Sloman et al. (2013), al demostrar que el tipo de procesamiento modifica la influencia de claves heurísticas como la deseabilidad social en la estimación de la comprensión causal. Una pregunta de interés para trabajos posteriores es si la capacidad para realizar procesamientos tipo 2 aumenta la precisión en los juicios metacognitivos en general, tal y como sugieren algunos trabajos previos (Mata et al., 2013), o si esta mejora solo ocurre en aquellos juicios que implican inhibir la activación de alguna clave heurística saliente (e.g., deseabilidad social, accesibilidad, fluidez de procesamiento, deseo de obtener buenos resultados en una prueba de conocimiento sobre un dominio particular, etc.).

En el Experimento 4, se manipuló la sobrecarga ejecutiva en cada condición experimental con el fin de determinar el nivel de procesamiento metacognitivo (alto o bajo) en el cual la deseabilidad social afecta la estimación de la comprensión causal. Aunque se encontró un efecto débil de la deseabilidad social, esta variable no interactúo con el nivel asignado de sobrecarga en la predicción de la IPE. En otras palabras, el nivel de sobrecarga

inducido no se asoció a una mayor o menor dependencia de la deseabilidad social en la estimación de la comprensión causal. Por esta razón, aunque el Experimento 4 fue diseñado con el propósito de contrastar las predicciones de los tres modelos de procesamiento metacognitivo postulados en la introducción (ver Figura 2), sus resultados no proporcionan evidencia concluyente a favor de alguno de ellos. Esto se debe a que dichos modelos asumen la existencia del efecto de la deseabilidad social sobre la IPE, el cual no fue significativo al incluirse la variable indicadora de sobrecarga (i.e., el porcentaje de aciertos en las tareas de memoria). Una posible explicación de este resultado negativo es que los participantes fueron inducidos a realizar procesamientos analíticos incluso en el nivel de sobrecarga baja, debido al contexto evaluativo generado por las tareas de memoria que se introdujeron siguiendo a Johnson et al. (2014). Así, la manipulación experimental podría haber creado una cantidad marginal de sobrecarga ejecutiva que restringió el uso de la deseabilidad social como clave informativa en ambos grupos. En el Experimento 4, se manipuló el contenido de la tarea (i.e., la complejidad de los estímulos y la memorización de información visual y verbal) en las dos condiciones de sobrecarga, con el fin de garantizar que ambos grupos de participantes realizaran tareas de similar duración y así controlar posibles efectos de fatiga. Una alternativa para verificar si la inclusión de tareas de memoria activó el procesamiento analítico consiste en realizar un estudio adicional con dos grupos experimentales: uno de sobrecarga nula, en el que los participantes no realicen tareas de memoria durante la estimación inicial, y otro de sobrecarga alta, en el que los participantes resuelvan tareas iguales o similares a las incluidas en el Experimento 4. De esta forma, se aumentaría el contraste entre las condiciones experimentales y los resultados en la condición de sobrecarga nula podrían ser comparados con los de experimentos anteriores.

Un hallazgo clave del Experimento 4 es que las manipulaciones de sobrecarga ejecutiva no tienen efectos uniformes sobre las estimaciones de comprensión, sino que estos están mediados por la forma en que los participantes reaccionan a las tareas concurrentes. En otras palabras, las manipulaciones de carga ejecutiva tienen efectos a través de los desempeños diferenciales en tareas de memoria, de modo que sujetos con diferentes niveles de funciones ejecutivas pueden reaccionar de formas distintas a las manipulaciones experimentales. Estas relaciones se pueden explorar en futuros trabajos usando técnicas de estimación causal avanzada (e.g., variables instrumentales) para determinar el efecto de tratamientos con efectos incompletos sobre las variables independientes (Gennetian, Morris, Bos, & Bloom, 2005).

Por otra parte, el hecho de que el desempeño en las tareas de memoria de las dos condiciones estuviese asociado a menores niveles de IPE sugiere que la sobrecarga ejecutiva puede afectar los procesos metacognitivos en general (Sannomiya & Ohtani, 2015) y las estimaciones de comprensión causal, en particular. Una posible explicación de este efecto es que las explicaciones no son recuperadas como piezas (*chunks*) acabadas de información de la memoria a largo plazo, sino que son construidas "sobre la marcha", utilizando la información más saliente o accesible al momento de su elaboración (Cimpian, 2015; Hussak & Cimpian, 2018). Desde esta perspectiva, las personas solo podrían estimar su conocimiento causal sobre un tema determinado al intentar recordar información relacionada con él. Así, cuando una tarea concurrente interfiere con este proceso de recuerdo, la estimación de la comprensión causal puede ser menos precisa dependiendo de la accesibilidad de la información que se intenta recordar. El efecto de la sobrecarga ejecutiva en los juicios metacognitivos también tiene una implicación práctica para el

diseño instruccional: la precisión de la autoevaluación de la comprensión puede depender de que esta se lleve a cabo en contextos de baja sobrecarga. En este sentido, las intervenciones dirigidas a mejorar la precisión de los juicios metacognitivos deberían minimizar las demandas del ambiente de tarea que ocupan recursos atencionales centrales.

Finalmente, las personas que se encontraban cursando (o habían cursado) asignaturas sobre historia de Colombia estimaron con mayor precisión su comprensión de los temas asignados. Este hecho resalta la necesidad de controlar de manera más rigurosa el nivel de conocimiento previo de los participantes en el paradigma de IPE. Aunque se sabe que las personas con más educación formal sobrestiman más su conocimiento causal sobre sus áreas de especialidad (Fisher & Keil, 2016), es posible que quienes recientemente hayan tenido que elaborar explicaciones sobre temas similares como parte de sus actividades académicas o profesionales puedan realizar estimaciones más precisas de su comprensión en dicho dominio. En este sentido, intercambiar explicaciones y recibir retroalimentación frecuente sobre la calidad de dichas explicaciones puede ser una estrategia efectiva para reducir la IPE en el mediano y largo plazo (Lachner & Neuburg, 2019; Rozenblit & Keil, 2002).

Limitaciones y futuros estudios

En el Experimento 1, se encontró un efecto inesperado de comparación entre los temas asignados. En particular, cuando los participantes explicaron inicialmente los temas menos deseables, las estimaciones de comprensión realizadas antes y después de explicar los temas más deseables fueron similares. Para controlar este posible efecto de orden en los experimentos posteriores, se decidió asignar los temas más y menos deseables socialmente a grupos distintos de participantes. Este resultado sugiere que las experiencias recientes de

estimación de la comprensión causal pueden afectar las estimaciones posteriores sobre temas diferentes. En línea con este hallazgo, Roeder y Nelson (2015) reportaron que el paradigma de IPE afecta las estimaciones de la comprensión *entre* dominios (*domain-agnostic*). Específicamente, estos investigadores encontraron que explicar el mecanismo causal de un artefacto o fenómeno perteneciente a un dominio particular (e.g., el funcionamiento de un helicóptero), disminuye la comprensión percibida de objetos y procesos relativos a otros dominios (e.g., las elecciones legislativas en Nigeria). De acuerdo con Roeder y Nelson, esto puede ocurrir porque los participantes cambian su interpretación de la escala de comprensión después de explicar un tema específico, porque utilizan dicha explicación como punto de contraste o "anclaje" (Tversky & Kahneman, 1974) en sus estimaciones posteriores. En este sentido, en futuras investigaciones será necesario controlar los efectos de la exposición repetida al paradigma de IPE y a otros arreglos experimentales utilizados tradicionalmente en el estudio de los juicios metacognitivos como el FOK o el JOL. La exploración de este tipo de efectos no solo es importante en virtud de sus implicaciones metodológicas, sino que además puede ayudar a esclarecer en qué medida los procesos de monitoreo y control asociados a los juicios metacognitivos son de dominio general o de dominio específico (Alexander, Winters, Dinsmore, & Parkinson, 2011; Morales, Lau, & Fleming, 2018).

Una explicación alternativa de los hallazgos del Experimento 1 es que el efecto de la deseabilidad social sobre la IPE ocurriese debido al total desconocimiento del tema menos deseable o que este sesgo metacognitivo fuese producto de un intento deliberado de los participantes por exagerar su competencia para explicar ciertos temas. El propósito del Experimento 2 fue verificar estas dos posibilidades. Sin embargo, no fue posible incluir las

estimaciones individuales de familiaridad como factores predictores en los modelos construidos, teniendo en cuenta que esta variable correlacionó con la deseabilidad social, lo que hubiera producido multicolinealidad (ver Resultados, Experimento 2). Este hallazgo plantea interrogantes sobre el origen y la organización de las claves heurísticas utilizadas en la evaluación metacognitiva subyacente a la IPE. Desde el punto de vista metodológico, es evidente la dificultad para distinguir los efectos de claves metacognitivas que pueden operar de manera simultánea y en la misma dirección, tales como la deseabilidad social y la familiaridad. En trabajos posteriores, será necesario determinar cuál es el conjunto de claves que las personas utilizan a la hora de estimar su comprensión causal de un mecanismo y hacer correcciones ante tareas de explicación, evaluando cómo estas claves se organizan e interactúan entre ellas. En concreto, un reto importante en futuras investigaciones consistirá en manipular de manera independiente la deseabilidad social del conocimiento y la familiaridad de los temas asignados para determinar si sus efectos sobre la IPE son aditivos o multiplicativos.

En investigaciones previas, se ha reportado que las personas no solo disminuyen sus estimaciones iniciales de comprensión de políticas controversiales después de tratar de explicarlas, sino que además reducen el extremismo de sus actitudes frente a ellas (Fernbach, Rogers et al., 2013; Johnson et al., 2016, Experimento 9). Por otra parte, Kahan (2013) encontró que las personas con la habilidad de realizar procesamientos tipo 2 son más proclives a cuestionar la validez de una afirmación cuando esta pone en duda las posiciones de su grupo ideológico sobre temas controversiales (e.g., el cambio climático). Por lo tanto, es posible que las personas que pueden estimar de manera más precisa su comprensión sobre un tema determinado (i.e., quienes tienen la capacidad de realizar inferencias

analíticas) sean también más renuentes a modificar sus actitudes sobre dicho tema después de explicarlo. Por ejemplo, si una persona se identifica como miembro de un movimiento político con actitudes abiertamente negativas sobre el matrimonio gay y la despenalización del aborto, es poco probable que sus actitudes cambien al darse cuenta de que desconoce los casos en los cuales la Corte Constitucional permite la interrupción del embarazo o los derechos legales que ya son reconocidos a las parejas del mismo sexo. En este caso, la motivación personal de identificarse como miembro de cierto partido o posición ideológica puede atenuar el efecto moderador de las actitudes asociado a la formulación de explicaciones. Recientemente se ha encontrado que las personas con posiciones políticas más extremas realizan estimaciones menos precisas de la corrección de sus respuestas en tareas perceptuales, de toma de decisiones y actualización de creencias (Rollwage, Dolan, & Fleming, 2018). Sin embargo, el mecanismo específico que subyace a esta relación aún no ha sido explorado en detalle. Así, es posible que el extremismo político sea o bien consecuencia de la aplicación sesgada de procesamientos metacognitivos de alto nivel (Arango-Muñoz, 2011), o el resultado de un déficit de habilidades metacognitivas más básicas. El estudio de las conexiones entre desempeño metacognitivo y procesamientos analíticos/controlados puede aportar información clave para entender las causas de la radicalización política en las sociedades democráticas contemporáneas.

La alta tasa de abandono de participantes entre la primera y la segunda sesión fue la limitación más importante del Experimento 3. Esto pudo ser ocasionado por un sesgo de auto-selección, en el cual las personas que creían estar menos informadas sobre las causas del conflicto (o tenían más dificultades para realizar inferencias analíticas) tuvieron una probabilidad más alta de abandonar el estudio al concluir la primera sesión. Una forma de

detectar este posible sesgo consistiría en comparar las estimaciones de comprensión previa de las personas que no completaron el estudio con las de aquellas que participaron en ambas sesiones. No obstante, los datos para realizar este análisis no están disponibles, debido a que dichas estimaciones fueron realizadas al inicio de la segunda sesión. Considerando que el Experimento 3 fue el único en el que participaron personas de distintos niveles de escolaridad y de varias regiones del país, la tasa de abandono pudo estar relacionada con la falta de incentivos adecuados para motivar la participación en las dos sesiones. En este contexto, Kahan (2016) destaca la importancia de contar con muestras representativas en los estudios sobre los procesos cognitivos asociados a la comprensión y formación de opiniones sobre temas de interés público. En estudios posteriores, será fundamental desarrollar estrategias de reclutamiento de participantes que permitan verificar la validez externa de los resultados en muestras estratificadas que capturen la diversidad de niveles de comprensión y actitudes hacia temas históricos, políticos y sociales de relevancia nacional.

Una diferencia adicional entre los experimentos previos y el Experimento 4 radica en la forma en la cual fueron presentados los temas históricos durante la estimación inicial de la comprensión: mientras que en los primeros cuatro experimentos los participantes estimaron su comprensión sobre los temas asignados en una matriz única, en el último experimento lo hicieron de manera secuencial. Esta modificación se implementó en el Experimento 4 para igualar la cantidad de sobrecarga ejecutiva a la que fueron expuestos los participantes durante la estimación inicial de comprensión sobre cada tema. Si los temas se hubiesen presentado en una sola matriz, los primeros en la lista habrían estado expuestos a una sobrecarga mayor. Es posible que estas dos formas de presentar los temas generen

diferencias en los juicios de comprensión, en la medida en que el contraste entre las estimaciones es más difícil de realizar cuando los temas se muestran de manera secuencial. La comparación experimental entre estas formas alternativas de presentación puede ayudar a dilucidar de qué manera los participantes realizan sus estimaciones de comprensión causal inicial.

En el contexto de la discusión sobre la hipótesis de mecanismo único de los juicios metacognitivos, Dunlosky y Tauber (2014) destacan "el valor de desarrollar nuevos métodos para explorar el grado en el cual las inferencias analíticas y no analíticas contribuyen a la influencia de cualquier clave dada (o la combinación de claves)" (p. 456). El diseño del Experimento 4 constituye un aporte metodológico en esta dirección, en la medida en que se utiliza la manipulación de la sobrecarga ejecutiva para determinar el nivel de procesamiento metacognitivo (alto o bajo) en el cual una clave inferencial particular (i.e., la deseabilidad social) afecta cierto tipo de juicios metacognitivos (i.e., juicios de comprensión causal). Examinar los mecanismos específicos a través de los cuales la sobrecarga ejecutiva interfiere con el desempeño metacognitivo permitirá refinar los modelos que integran las teorías de procesamiento dual con los procesos de monitoreo y control de la cognición.

Conclusión

En la vida cotidiana, las personas toman decisiones importantes que requieren algún tipo de conocimiento causal (e.g., elección de estilos de vida saludables, tratamientos médicos, prácticas de consumo ambientalmente responsables, régimen de ahorro pensional, esquema de crédito, o cómo votar una propuesta legislativa que se somete a referendo o plebiscito). La razonabilidad de estas decisiones depende críticamente de que las personas

sean capaces de reconocer si necesitan o no adquirir más información sobre un tema específico antes de tomar este tipo de decisiones. Esta capacidad es especialmente relevante si se tiene en cuenta que las personas deciden realizar búsquedas de información o repasar contenidos específicos en función de lo que ya creen saber o haber aprendido (Radecki & Jaccard, 1995; De Bruin & Van Gog, 2012). Los resultados presentados en esta disertación muestran que las estimaciones subjetivas de comprensión causal pueden ser influenciadas por claves que no constituyen indicadores fiables de la precisión y completitud del conocimiento disponible, esto es, claves cuya validez ecológica es incierta (Brunswik, 1956). Este efecto ocurre debido a las personas no poseen un sistema de monitoreo exhaustivo de su conocimiento causal, sino que realizan sus estimaciones de comprensión de manera indirecta, a partir de claves inferenciales como la deseabilidad social del conocimiento. Además, los sesgos asociados a estas estimaciones metacognitivas pueden aumentar cuando la sobrecarga de los recursos atencionales impide la puesta en marcha de procesamientos analíticos y controlados. Así, la capacidad para realizar los procesamientos tipo 2 no solo tiene efectos sobre los mecanismos inferenciales en el nivel cognitivo, sino también en los propios del nivel metacognitivo (Proust, 2013).

En un nivel global, esta disertación enlaza la investigación sobre metamemoria e IPE. En los estudios originales del paradigma de IPE, Rozenblit y Keil (2002) mencionan dos características que diferencian la IPE de otros tipos de paradigmas experimentales metacognitivos. En primer lugar, los juicios solicitados a los participantes en la IPE indagan sobre la "profundidad" o nivel de detalle de la comprensión causal, más que en el hecho de reconocer o no cierto ítem de información (como ocurre en el FOK). Por otro lado, Rozenblit y Keil (2002, Experimento 12a) no encontraron una asociación entre la

familiaridad de los temas y la magnitud de la IPE, análoga a la reportada previamente con respecto al FOK (Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992). No obstante, los hallazgos reportados en este trabajo hacen evidentes las conexiones teóricas, metodológicas y conceptuales que existen entre estos tipos de juicios metacognitivos. Los resultados muestran que la IPE se puede interpretar en términos de conceptos centrales en la investigación en metacognición (e.g., el uso de claves heurísticas) y no solo como un producto de la confusión entre la comprensión de los principios causales generales de un dominio con la comprensión de mecanismos y fenómenos específicos, tal y como tradicionalmente se ha interpretado la IPE (Keil, 2003; 2012; Keil et al, 2004).

En este sentido, los resultados aquí presentados tienen el potencial heurístico para abrir nuevos caminos en la investigación en IPE relativos, por ejemplo, a la forma en que los factores asociados al FOK o a otros juicios metacognitivos pueden intervenir en la generación de la IPE. Por lo anterior, esta disertación hace un aporte fundamental a la comprensión de los mecanismos que subyacen a la IPE, no desde conjeturas generales como en el caso de las teorías intuitivas o legas (*folk theories*), sino desde la manipulación sistemática de las variables que pueden estar detrás de la sobrestimación inicial de la comprensión causal y de su corrección posterior. En otras palabras, los resultados de esta tesis orientan la investigación de la IPE hacia explicaciones de proceso que den cuenta de la arquitectura y los mecanismos que subyacen a la estimación de la comprensión causal, más allá de la descripción o caracterización estadística del fenómeno (Larreamendy-Joerns, 2010). Consecuentemente, los estudios reportados constituyen un primer paso para posicionar la investigación sobre la IPE dentro del amplio espectro de juicios metacognitivos (e.g., FOK, EOL, JOL, JOC, FOR, etc.), lo que en el largo plazo puede

permitir caracterizar la IPE como un fenómeno articulado en una arquitectura o modelo metacognitivo más general, más allá de sus consistentes efectos comportamentales.

Los efectos encontrados de la deseabilidad sobre la IPE destacan el rol de lo social en la estimación metacognitiva, lo que puede servir para trazar puentes entre la investigación puramente psicológica y la investigación más amplia en las ciencias sociales que se enfoca en las dinámicas macro (e.g., agendas mediáticas, educación, construcción de capital humano). Más allá de ser un resultado trivial, lo que esto indica es que la IPE no solo depende del procesamiento y monitoreo individual de la comprensión en diferentes niveles dentro de cada dominio (Keil, 2012), sino también que es un fenómeno contextual que depende de lo que se espera del sujeto y de lo que él espera de sí mismo (Ottati, Price, Wilson, & Sumaktoyo, 2015), de lo que es valorado en su contexto social, y de la forma en que las interacciones posicionan o no algunos temas como relevantes. Por ejemplo, en un estudio realizado con adolescentes de Irlanda del Norte, Dunne, McCann, Millen, Wilson y Macdonald (2015) reportaron que la presión de pares para tener relaciones sexuales se asocia a bajos desempeños en una prueba de conocimientos sobre salud sexual. Así, la deseabilidad social del conocimiento sobre sexualidad en la población adolescente puede crear la ilusión de comprensión, aumentando la probabilidad de que ellos incurran en conductas de riesgo. En este sentido, es importante explorar las posibles variaciones culturales del fenómeno de IPE y cómo diferentes contextos, sociales, culturales y políticos pueden exacerbar o mitigar sus efectos.

La identificación consistente de la IPE en fenómenos de tipo histórico es otro elemento novedoso que merece la pena resaltar. Si bien la investigación previa ha identificado este fenómeno en el dominio social, los temas explicados podrían

caracterizarse como de tipo institucional o económico (Fernbach, Rogers et al., 2013; Johnson et al., 2016; Roeder & Nelson, 2015). La historia como dominio disciplinar tiene características particulares que la diferencian de otros dominios dentro de las ciencias sociales, entre ellos su carácter multicausal y narrativo, y la interacción entre elementos intencionales y estructurales dentro de la explicación del cambio (Hilton & Liu, 2017; Navarro-Roldán & Corredor, 2018). La distinción cognitiva entre el procesamiento en dominios expositivos/argumentativos, por una parte, y dominios narrativos, por otra, es importante en la caracterización de la resolución de problemas, la comprensión y la toma de decisiones (Corredor & Rojas, 2016; Graesser, Singer, & Trabasso, 1994; Weaver & Bryan, 1995). En este sentido, mostrar que la IPE se produce también en el dominio histórico es un hallazgo importante, porque indica que esta ilusión no depende de los elementos que cognitivamente diferencian narraciones históricas de mecanismos o dominios argumentativos o expositivos. El hecho de que en estos dominios se encuentre un efecto de la deseabilidad sobre la IPE, es, además, importante porque reafirma la idea de que la atención a claves heurísticas puede ser un mecanismo metacognitivo de carácter general y transversal a diferentes modos de representación del conocimiento. Los efectos de la deseabilidad sobre la comprensión de fenómenos históricos, más allá de lo puramente teórico, plantean importantes retos para la educación y para la comprensión de la política contemporánea, ya que muestra que, en condiciones de agitación política o radicalización ideológica, los ciudadanos pueden realizar monitoreos inadecuados de los procesos a través de los cuales toman decisiones de carácter político con implicaciones históricas (Anson, 2018).

En general, la influencia del procesamiento controlado y los efectos que sobre este tienen las restricciones de recursos cognitivos centrales tienen importantes consecuencias para el diseño de intervenciones sociales y educativas. En este sentido, el diseño de estrategias que minimicen la sobrecarga ejecutiva y promuevan el procesamiento metacognitivo de alto nivel durante la estimación de la comprensión causal puede contribuir a que los ciudadanos sean capaces de distinguir entre lo que saben y lo que desconocen (pero desean saber) sobre temas actuales que tienen implicaciones profundas para sus modos de vida y los de las sociedades en las que habitan.

Referencias

Ackerman, R., & Beller, Y. (2017). Shared and distinct cue utilization for metacognitive judgments during reasoning and memorization. *Thinking & Reasoning*, *23*, 376-408. https://doi.org/10.1080/13546783.2017.1328373

Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory. En A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 164-182). Hove, UK: Psychology Press.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*, 607-617. https://doi.org/10.1016/j.tics.2017.05.004

Aguiar, N. R., Stoess, C. J., & Taylor, M. (2012). The development of children's ability to fill the gaps in their knowledge by consulting experts. *Child Development, 83*, 1368-1381. https://doi.org/10.1111/j.1467-8624.2012.01782.x

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research*, *27*, 123-156. https://doi.org/10.1086/314317

Alexander, P. A., Winters, F., Dinsmore, D. L., & Parkinson, M. (2011). The role of domain knowledge in self-regulated learning. En B. Zimmerman & D. Schunk (Ed.), *Handbook of self-regulation of learning and performance* (pp. 393-407). New York: Routledge.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*, 1621-1630. https://doi.org/10.1037/0022-3514.49.6.1621

Alter, A. L., Oppenheimer, D. M., & Zemla, J. C. (2010). Missing the trees for the forest: a construal level account of the illusion of explanatory depth. *Journal of Personality and Social Psychology*, *99*, 436-451. https://doi.org/10.1037/a0020218

Anson, I. G. (2018). Partisanship, political knowledge, and the Dunning-Kruger effect. *Political Psychology*, *39*, 1173-1192. https://doi.org/10.1111/pops.12490

Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia*, *39*, 71-82. https://doi.org/10.1007/s11406-010-9279-0

Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, *26*, 1295-1303. https://doi.org/10.1177/0956797615588195

Baddeley, A. (2007). *Working memory, thought and action*. Oxford, UK: Oxford University Press.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90-109. https://doi.org/10.1016/j.cognition.2016.10.014

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*, 241-254. https://doi.org/10.1017/S0140525X07001653

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review*, *31*, 79-85. https://doi.org/10.1016/j.dr.2011.07.002

Bates, D., et al. (2019). lme4: Linear mixed-effects models using 'Eigen' and S4. [Software]. Recuperado de https://cran.r-project.org/web/packages/lme4

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2017). Teasing apart overclaiming, overconfidence and socially desirable responding. *Assessment*, 1-13. https://doi.org/10.1177/1073191117700268

Benson, H. H. (2011). Socratic method. En D. R. Morrison (Ed.), *The Cambridge companion to Socrates*. New York: Cambridge University Press. (pp. 179-200). https://doi.org/10.1017/CCOL9780521833424.008

Bialek, M., & Pennycook, G. (2017). The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods*, *50*, 1553-1559. https://doi.org/10.3758/s13428-017-0963-x

Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, *50*, 240-250. https://doi.org/10.1086/268978

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. En J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.

Bjork, R.A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417-444. https://doi.org/10.1146/annurev-psych-113011-143823

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology*, *3*, 1-6. https://doi.org/10.3389/fpsyg.2012.00229

Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology*, *41*, 491-505. https://doi.org/10.1037/0012-1649.41.3.491

Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society*, *17*, 82-84. https://doi.org/10.3758/BF03333674

Brennan, J. (2016). *Against Democracy*. Princeton, NJ: Princeton University Press.

Bromme, R., Thomm, E., & Ratermann, K. (2015). Who knows? Explaining impacts on the assessment of our own knowledge and of the knowledge of experts. *Zeitschrift für Pädagogische Psychologie*, *30*, 97-108. https://doi.org/10.1024/1010-0652/a000175

Brunswik, E. (1956). *Perception and representative design of psychological experiments*. Berkeley, CA: University of California Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542-579. https://doi.org/10.1016/0749-596X(91)90026-G

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in- versus out-group members in 14-month-old infants. *Child Development*, *84*, 422–428. https://doi.org/10.1111/j.1467-8624.2012.01860.x

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modelling approach. *Memory and Cognition*, *42*, 434-447. https://doi.org/10.3758/s13421-013-0367-9

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*, 121-138. https://doi.org/10.1017/S0140525X09000545

Carruthers, P. (2017). Are epistemic emotions metacognitive? *Philosophical Psychology*, *30*, 58-78. https://doi.org/10.1080/09515089.2016.1262536

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: evidence from judgments of learning. *Psychonomic Bulletin Review*, *14*, 107–111. https://doi.org/10.3758/BF03194036

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752-766. https://doi.org/10.1037/0022-3514.39.5.752

Cimpian, A. (2015). The inherence heuristic: Generating everyday explanations. En R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–15). Hoboken, NJ: John Wiley and Sons.

Clariana, M., Castelló, A., & Cladellas, R. (2016). Feeling of knowing and overclaiming in students from secondary school to university. *Learning and Individual Differences*, *49*, 421-427. https://doi.org/10.1016/j.lindif.2016.05.008

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589-608. http://dx.doi.org/10.1037/0033-295X.100.4.589

Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 544-554. https://doi.org/10.1037/0278-7393.18.3.544

Corredor, J. A., & Rojas, L. (2016). Narrative and conceptual expertise in massively multiplayer online role playing games. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *8*, 44-67. https://doi.org/10.4018/IJGCMS.2016010104

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 142–150. https://doi.org/10.1037/0278-7393.18.1.142

Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97, 931-945. https://doi.org/10.1037/a0017452

De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*, 245-252. https://doi.org/10.1016/j.learninstruc.2012.01.003

De Carvalho Filho, M. K., & Yuzawa, M. (2001). The effect of social influences and general metacognitive knowledge on metamemory judgments. *Contemporary Educational Psychology*, *26*, 571-587. https://doi.org/10.1006/ceps.2000.1077

De Neys, W. (2006a) Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*, 1070-1100. https://doi.org/10.1080/02724980543000123

De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428-433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2017). Bias, conflict, and fast logic: Towards a hybrid dual process future? En W. De Neys (Ed.), *Dual process theory 2.0* (pp. 47-65). London, UK: Routledge.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18. https://doi.org/10.1037/0022-3514.56.1.5

Dropp, K., Kertzer, J. D., & Zeltzoff, T. (2014). The less Americans know about Ukraine's location, the more they want U.S. to intervene. [Blog post]. Recuperado de: https://www.washingtonpost.com/news/monkey-cage/wp/2014/04/07/the-less-americans-know-about-ukraines-location-the-more-they-want-u-s-to-intervene/

Dunne, L., McCann, M., Millen, S., Wilson, J., & Macdonald, G. (2015). Overconfidence, sexual health awareness and sexual health risk among young female users of sexual health clinics. *International Journal of Educational Research*, 71, 26-32. http://dx.doi.org/10.1016/j.ijer.2015.02.007

Dunlosky, J., Hertzog, C., Kennedy, M. R. F., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10*, 4-11.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, 52, 551–565. http://dx.doi.org/10.1016/j.jml.2005.01.011

Dunlosky, J., & Tauber, S. K. (2014). Understanding people's metacognitive judgments: An isomechanism framework and its implications for applied and theoretical research. En T. Perfect & D. S. Lindsay (Eds.), *Handbook of applied memory* (pp. 444-464). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781446294703.n25

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding, students' judgments can improve their achievement. *Learning and Instruction*, *24*, 58-61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*, 526-534. https://doi.org/10.1016/j.jml.2005.01.009

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*, 277-287. https://doi.org/10.1027/1016-9040.13.4.277

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*, 5-17. https://doi.org/10.1037/0022-3514.84.1.5

Eitam, B., Hassin, R.R., & Schul, Y. (2008). Non-conscious goal pursuit in novel environments: The case of implicit learning. *Psychological Science*, *19*, 261-267. https://doi.org/10.1111/j.1467-9280.2008.02078.x.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709-724. https://doi.org/10.1037/0003-066X.49.8.709

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology, 71*, 390-405. https://doi.org/10.1037/0022-3514.71.2.390

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, *59*, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223-241. https://doi.org/10.1177/1745691612460685

Evans, N., & Stanwyck, D. J. (1997). Illusion of knowing in adult readers: Effects of reading skill and passage length. *Contemporary Educational Psychology*, *22*, 39-52. https://doi.org/10.1006/ceps.1997.0925

Fabricius, G., & Büttgen, M. (2013). Influence of knowledge on overconfidence: Consequences for management and project planning. *International Journal of Business and Management*, *8*, 1-12. https://doi.org/10.5539/ijbm.v8n11p1

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nature Human Behavior*. https://doi.org/10.1038/s41562-018-0520-3

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*, 939-946. https://doi.org/10.1177/0956797612464058

Fernbach, P. M., Sloman, S. A., St. Louis, R., & Shube, J. N. (2013). Explanation fiends and foes: How mechanistic detail determines understanding and preference. *Journal of Consumer Research*, *39*, 1115-1131. https://doi.org/10.1086/667782

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*, 813-821. https://doi.org/10.3758/MC.36.4.813

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564. https://doi.org/10.1037/0096-1523.3.4.552

Fisher, M., & Keil, F. C. (2014). The illusion of argument justification. *Journal of Experimental Psychology: General, 143,* 425–433. https://doi.org/10.1037/a0032234

Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science, 40,* 1251-1269. https://doi.org/10.1111/cogs.12280

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Frankish, K., & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. En J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1-30). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.003.0001

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning, 15,* 105-128. https://doi.org/10.1080/13546780802711185

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25-42. https://doi.org/10.1257/089533005775196732.

Gana, K., & Broc, G. (2019). Structural Equation Modeling with lavaan. London, UK: Wiley ISTE.

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual Review of Anthropology*, *40*, 379-398. https://doi.org/10.1146/annurev-anthro-081309-145822

Gennetian, L. A., Morris, P. A., Bos, J. M., & Bloom, H. S. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. En H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 75-114). New York, NY, US: Russell Sage Foundation.

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. En M. Waldmannn (Ed.), *The Oxford handbook of causal reasoning* (pp. 515-548). New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.28

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*, 597-602. https://doi.org/10.3758/BF03202442

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology*. *Learning, Memory, and Cognition*, *11*, 702–718. https://doi.org/10.1037/0278-7393.11.1-4.702

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*, 84-93. https://doi.org/10.3758/BF03197714

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 311-325. https://doi.org/10.1037/0278-7393.7.5.311

Graeff, T. R. (2003). Exploring consumers' answers to survey questions: Are uninformed responses truly uninformed? *Psychology & Marketing, 20*, 643-667. https://doi.org/10.1002/mar.10090

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371-395. http://dx.doi.org/10.1037/0033-295X.101.3.371

Graf, L. K., Mayer, S., & Landwher, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, 28, 393-411. https://doi.org/10.1002/jcpy.1021

Gruneberg, M. M., Monks, J., & Sykes, R. N. (1977). Some methodological problems with feeling of knowing studies. *Acta Psychologica*, *41*, 365-371. https://doi.org/10.1016/0001-6918(77)90014-2

Hall, M. P., & Raimi, K. T. (2018). Is belief superiority justified by superior knowledge? *Journal of Experimental Social Psychology*, *76*, 290-306. https://doi.org/10.1016/j.jesp.2018.03.001

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208-216. https://doi.org/10.1037/h0022263

Hart, J. T. (1967). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology*, *58*, 193-197. https://doi.org/10.1037/h0024908

Hilton, D. J., & Liu, J. H. (2017). History as the narrative of a people: From function to structure and content. *Memory Studies*, *10*, 297-309. https://doi.org/10.1177/1750698017701612

Hosey, L. A., Peynircioğlu, Z. F., & Rabinovitz, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica, 130,* 214-224. https://doi.org/10.1016/j.actpsy.2008.12.007

Hülür, G., Wilhelm, O., & Schipolowski, S. (2011). Prediction of self-reported knowledge with overclaiming, fluid and crystallized intelligence and typical intellectual engagement. *Learning and Individual Differences, 21*, 742-746. https://doi.org/10.1016/j.lindif.2011.09.006

Hussak, L. J., & Cimpian, A. (2018). Memory accessibility shapes explanation: Testing key claims of the inherence heuristic account**.** *Memory & Cognition, 46,* 68–88. https://doi.org/10.3758/s13421-017-0746-8

Ipsos MORI (2016, Junio 8). The perils of perception and the EU. Recuperado de: https://www.ipsos.com/ipsos-mori/en-uk/perils-perception-and-eu

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: Capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*, 25-49. https://doi.org/10.1007/s11409-013-9110-y

Jackson S. A., Kleitman S, Howie P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology*, *7*, 1559. https://doi.org/10.3389/fpsyg.2016.01559

Jaeger, A. J., & Wiley, J. (2015). Reading an analogy can cause the illusion of comprehension. *Discourse Processes: A Multidisciplinary Journal, 52*, 376-405. https://doi.org/10.1080/0163853X.2015.1026679

Johnson, D. R., Murphy, M. P., & Messer, R. M. (2016). Reflecting on explanatory ability: A mechanism for detecting gaps in causal knowledge. *Journal of Experimental Psychology: General*, *145*, 573-588. https://doi.org/10.1037/xge0000161

Johnson, E. D., Tubau, E., & De Neys, W. (2014). The unbearable burden of executive load on cognitive reflection: A validation of dual process theory. En P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting Cognitive Science Society* (pp. 2441-2446). Austin, TX: Cognitive Science Society.

Johnson, D. R., Tynan, M. E., Cuthbert, A. S., & O'Quinn, J. K. (2017). Metacognition in argument generation: The misperceived relationship between emotional investment and argument quality. *Cognition and Emotion, 32,* 566-578. https://doi.org/10.1080/02699931.2017.1330743

Johnson, S. G. B., & Ahn, W. (2017). Causal mechanisms. En M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 127-146). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.12

Kahan, D. M. (2013). Ideology, motivation and cognitive reflection: An experimental study. *Judgement and Decision Making*, 8, 407-424. Recuperado de: http://journal.sjdm.org/13/13313/jdm13313.pdf

Kahan, D. M. (2016). The Politically Motivated Cognition Paradigm, Part 1: What politically motivated cognition is and how to measure it. Recuperado de: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703011

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review, 93,* 1449-1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. En K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267-293). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251. https://doi.org/10.1037/h0034747

Karabenick, S. A. (1996). Social influences on metacognition: Effects of colearner questioning on comprehension monitoring. *Journal of Educational Psychology, 88,* 689-703. https://doi.org/10.1037/0022-0663.88.4.689

Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science, 29,* 521-536. https://doi.org/10.1177/0956797617740646

Keil, F.C. (1995). The growth of causal understandings of natural kinds: Modes of construal and the emergence of biological thought. En D. Sperber, D. Premack & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 234-262). New York: Clarendon Press/Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524021.003.0009

Keil, F.C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Science, 7,* 368-373. https://doi.org/10.1016/S1364-6613(03)00158-X

Keil, F. C. (2010). The feasibility of folk science. *Cognitive Science, 34,* 826-862. https://doi.org/10.1111/j.1551-6709.2010.01108.x

Keil, F. C. (2012). Running on empty? How folk science gets by with less. *Current Directions in Psychological Science, 21,* 329-334. https://doi.org/10.1177/0963721412453721

Keil, F. C., & Kominsky, J. (2015). Grounding concepts. En E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 677-692). Cambridge, MA: MIT Press.

Keil, F.C., Stein, C., Webb, L., Billings, V.D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science, 32,* 259-300. https://doi.org/10.1080/03640210701863339

Keil, F.C., Rozenblit, L.R., & Mills, C. (2004). What lies beneath? Understanding the limits of understanding. En D.T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 227-249). Cambridge, MA: MIT Press.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing—Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1–24. https://doi.org/10.1006/jmla.1993.1001

Keren, G., & Schul, Y. (2009). Two is not always better than one. A critical evaluation of two-system theory. *Perspectives on Psychological Science*, *4*, 533-550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition*, *4*, 553-558. https://doi.org/10.3758/BF03213218

Kominsky, J., & Keil, F. C. (2014). Overestimation of knowledge about word meanings: The "misplaced meaning" effect. *Cognitive Science*, *38*, 1604-1633. https://doi.org/10.1111/cogs.12122

Kominsky, J. F., Zamm, A. P., & Keil, F. C. (2017). Knowing when help is needed: A developing sense of causal complexity. *Cognitive Science*, *42*, 491-523. https://doi.org/10.1111/cogs.12509

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology*: *General*, *124*, 311-333. https://doi.org/10.1037/0096-3445.124.3.311

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. En P. D. Zelazo, M. Moscovitch & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). New York: Cambridge University Press.

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 945–959. https://doi.org/10.1037/0278-7393.34.4.945

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119,* 80-113. https://doi.org/10.1037/a0025648

Koriat, A. (2015). Metacognition: Decision-making processes in self-monitoring and self-regulation. En G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of*

metamemory (pp. 127-147). New York: Oxford University Press. https://doi.org/10.1002/9781118468333.ch12

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 187-194. https://doi.org/10.1037/0278-7393.31.2.187

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643-656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. En S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483-502). New York: Guilford Press.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 34-53. https://doi.org/10.1037/0278-7393.27.1.34

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118. https://doi.org/10.1037/0278-7393.6.2.107

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation

between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36-69. https://doi.org/10.1037/0096-3445.135.1.36

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. En J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 117-135). New York, NY: Psychology Press.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449-468. https://doi.org/10.1037/a0017350

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology, 65*, 1132–1151. https://doi.org/10.1037/0022-3514.65.6.1132

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

LaBorde, P. M., Mottner, S., & Whalley, P. (2015). Personal financial literacy: Perceptions of knowledge, actual knowledge and behavior of college students. *Journal of Financial Education*, *39*, 1-30. Recuperado de: http://www.jstor.org/stable/23608645

Lachner, A., & Neuburg, C. (2019). Learning by writing explanations: computer-based feedback about the explanatory cohesion enhances students' transfer. *Instructional Science*, *47*,19-37. https://doi.org/10.1007/s11251-018-9470-4

Larreamendy-Joerns, J. (2010). Proceso, coherencia, cotidianidad y *qualia* como desafíos explicativos para la psicología. *CS, 5*, 17-41. http://dx.doi.org/10.18046/recs.i5.450

Lawson, R. (2006). The science of cycology: Failures to understand how everyday objects work. *Memory & Cognition, 34*, 1667-1675. https://doi.org/10.3758/BF03195929

Leiser, D., & Shemesh, Y. (2018). *How we misunderstand economics and why it matters: The psychology of bias, distortion and conspiracy*. New York: Routledge.

Leonesio, R. J., & Nelson. T. O. (1990). Do different metamemory judgments tap the same underlying tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 464-470. https://doi.org/10.1037/0278-7393.16.3.464

Lichtenstein, S., & Fischoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior & Human Decision Processes, 20*, 159-183. https://doi.org/10.1016/0030-5073(77)90001-0

Lin-Miao, L., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications. *Contemporary Education Psychology, 23*, 345-391. https://doi.org/10.1006/ceps.1998.0972

Liu, Y., Su, Y., Xu, G., & Chan, R. C. K. (2007). Two dissociable aspects of feeling-of-knowing: Knowing that you know and knowing that you do not know. *Quarterly Journal of Experimental Psychology, 60*, 672-680. https://doi.org/10.1080/17470210601184039

Lombrozo, T. (2009). Explanation and categorization: How ‘‘why?” informs ‘‘what?”. *Cognition, 110*, 248–253. https://doi.org/10.1016/j.cognition.2008.10.007

Lombrozo, T. (2010). Causal-explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*, 303-332. https://doi.org/10.1016/j.cogpsych.2010.05.002

Lupia, A. (2016). *Uninformed: Why people seem to know so little about politics and what we can do about it*. New York: Oxford University Press.

Lusk, A. (2016). Metacognitive strategies in the introduction to political science classroom. *Journal of Political Science Education*, *12*, 141-150. https://doi.org/10.1080/15512169.2015.1060888

Lutz, D. J., & Keil, F. C. (2002). Early understandings of the division of cognitive labor. *Child Development*, *73*, 1073–1084. https://doi.org/10.1111/1467-8624.00458

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*, 723–731. https://doi.org/10.1037/0022-0663.97.4.723

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*, 353-373. https://doi.org/10.1037/a0033640

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Science*, *22*, 280-293. https://doi.org/10.1016/j.tics.2018.02.001

Metcalfe, J. (2008). Evolution of metacognition. En J. Dunlosky, & R. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 29-46). New York: Psychology Press.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18,* 159-163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15,* 174-179. https://doi.org/10.1016/j.jml.2007.03.006

Metcalfe, J., & Son, L. K. (2012). Anoetic, noetic, and autonoetic metacognition. En M. Beran, J. Brandl, J. Perner & J. Proust (Eds.), *The foundations of metacognition* (pp. 289-301). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646739.003.0019

Mills, C., & Keil, F. C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology, 87,* 1-32. https://doi.org/10.1016/j.jecp.2003.09.003

Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621-640. https://doi.org/10.1037//0096-3445.130.4.621

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502-517. https://doi.org/10.1037/0033-295X.115.2.502

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *The Journal of Neuroscience, 38,* 3534-3546. https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Motta, M., Callaghan, T., Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science and Medicine, 211,* 274-281. https://doi.org/10.1016/j.socscimed.2018.06.032

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93,* 245-258. https://doi.org/10.1016/j.jml.2016.10.008

Mueller, M. L, Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20,* 378-384. https://doi.org/10.3758/s13423-012-0343-6

Navarro-Roldan, C. P., & Corredor, J. A. (2018). Comunalidades entre las narrativas históricas de estudiantes y textos escolares: un análisis cualitativo. *Revista Colombiana de Educación, 75,* 119-138. https://doi.org/10.17227/rce.num75-8104

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109-133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metacognition: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26,* 125-173. https://doi.org/10.1016/S0079-7421(08)60053-5

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1154-1170. https://doi.org/10.1037/xlm0000372

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Noyes, A., & Dunham, Y. (2017). Mutual intentions as a causal framework for social groups. *Cognition*, *162*, 133-142. https://doi.org/10.1016/j.cognition.2017.02.007

Noyes, A., Keil, F. C., & Dunham, Y. (2018). The emerging causal understanding of institutional objects. *Cognition*, *170*, 83-87. https://doi.org/10.1016/j.cognition.2017.09.008

Ojalehto, B. L., & Medin, D. L. (2015). Perspectives on culture and concepts. *Annual Review of Psychology*, *66*, 249-275. https://doi.org/10.1146/annurev-psych-010814-015120

Oppenheimer, D. M., Meyvis T., & Davidenko N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Ortoleva, P., & Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, *105*, 504-535. https://doi.org/10.1257/aer.20130921

Ottati, V., Price, E. D., Wilson, C., & Sumaktoyo, N. (2015). When self-perceptions of expertise increase close-minded cognition: The earned dogmatism effect. *Journal of Experimental Social Psychology*, *61*, 131-138. https://doi.org/10.1016/j.jesp.2015.08.003

Park, C. W., Gardner, M. P., & Thukral, V. K. (1988). Self-perceived knowledge: Some effects on information processing for a choice task. *The American Journal of Psychology*, *101*, 401-424. https://doi.org/10.2307/1423087.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. En H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum.

Paulhus, D. L. (2012). Overclaiming on personality questionnaires. En M. Ziegler, C. McCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessments* (pp. 151-164). New York: Oxford University Press.

Paulhus D. L. (2017). Socially desirable responding on self-reports. En V. Zeigler-Hill, & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-28099-8

Paulhus, D. L., & Dubois, P. J. (2014). Application of the overclaiming technique to scholastic assessment. *Educational and Psychological Measurement*, *74*, 975-990. https://doi.org/10.1177/0013164414536184

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The overclaiming technique: Measuring bias independent of accuracy. *Journal of Personality and Social Psychology, 84*, 681-693. https://doi.org/10.1037/0022-3514.84.4.890

Paulhus, D. L., & Trapnell, P. D. (2008). Self-presentation of personality: An agency-communion framework. En O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 492-517). New York: The Guilford Press.

Pennycook, G. (2017). A perspective on the theoretical foundation of dual-process models. En W. De Neys (Ed.), *Dual process theory 2.0* (pp. 5-27). New York, NY: Psychology Press.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D.J., & Fugelsang, J.A. (2014). The role of analytic thinking in moral judgments and values. *Thinking & Reasoning*, *20*, 188-214. https://doi.org/10.1080/13546783.2013.865000

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D.J., & Fugelsang, J.A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *213*, 335-346. https://doi.org/10.1016/j.cognition.2012.03.003

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24, 425-43. https://doi.org/10.1177/0963721415604610

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What make us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34-72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G. & Rand, D. G. (2017). Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. Recuperado de https://ssrn.com/abstract=3023545. https://doi.org/10.2139/ssrn.3023545

Pennycook, G., & Rand, D. G. (2018). Cognitive reflection and the 2016 US presidential election. Recuperado de https://ssrn.com/abstract=3110929. https://doi.org/10.2139/ssrn.3110929

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*, 1774-1784. https://doi.org/10.3758/s13423-017-1242-7

Pennycook, G., & Thompson, V.A. (2016). Base-rate neglect. En R. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory*, 2a. Ed., (pp. 44-61). Hove, UK: Psychology Press.

Peña-Ayala, A., & Cárdenas, L. (2015). A conceptual model of the metacognitive activity. En A. Peña-Ayala (Ed.), *Metacognition: Fundaments, applications and trends* (pp. 39-72). Berlin: Springer. https://doi.org/10.1007/978-3-319-11062-2_3

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. En L. Berkowitz (Ed.), *Advances in Experimental Social Psychology, 19*, (pp. 123-205). New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

Pew Research Center (2012, Julio 2). Division, uncertainty over Court's Health Care ruling. Recuperado de: http://www.people-press.org/2012/07/02/division-uncertainty-over-courts-health-care-ruling/

Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning, 8*, 213-234. https://doi.org/10.1007/s11409-013-9104-9

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences, 110*, E3965-E3972. https://doi.org/10.1073/pnas.1304326110

Proust, J. (2012). Metacognition and mindreading: one or two functions? En M. Beran, J. Brandl, J. Perner & J. Proust (Eds.), *The foundations of metacognition* (pp. 234-251).

Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646739.003.0015

Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness.* Oxford, UK: Oxford University Press.

Qualtrics (2005). [Software]. Provo, UT: Qualtrics.

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication*, 43, 103-121. https://doi.org/10.1016/j.specom.2004.02.004

Qvortrup, M. (2018). Two hundred years of referendums. En M. Qvortrup (Ed.), *Referendums around the world* (pp. 263-272). Cham, Switzerland: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-57798-2_9

R Core Team (2017). R: A language and environment for statistical computing. Viena, Austria: R Foundation for Statistical Computing. Recuperado de: http://www.r-project.org/

Radecki, C. M., & Jaccard, J. (1995). Perceptions of knowledge, actual knowledge and information search behavior. *Journal of Experimental Social Psychology*, *31,* 107–138. https://doi.org/10.1006/jesp.1995.1006

Raimi, K. T., & Leary, M. R. (2014). Belief superiority in the environmental domain: Attitude extremity and reactions to fracking. *Journal of Environmental Psychology*, *40*, 76-85. https://doi.org/10.1016/j.jenvp.2014.05.005

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Green, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. https://doi.org/10.1038/ncomms4677

Rawson, K. A., Dunlosky, J., & McDonald, S. L. (2002). Influences of metamemory on performance predictions for text. *Quarterly Journal of Experimental Psychology*, *55*, 505-524. https://doi.org/10.1080/02724980143000352

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory*, *1*, 217-229.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with questions terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 435-451. https://doi.org/10.1037/0278-7393.18.3.435

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615-625. https://doi.org/10.1037/a0013684

Rips, L., & Hespos, S. J. (2015). Divisions of the physical world: Concepts of objects and substances. *Psychological Bulletin*, *141*, 786-811. https://doi.org/10.1037/bul0000011

Robey, A. M., Dougherty, M. R., & Buttaccio, D. R. (2017). Making retrospective confidence judgments improves learners' ability to decide what not to study. *Psychological Science*, *28*, 1683-1693. https://doi.org/10.1177/0956797617718800

Roeder, S. S., & Nelson, L. D. (2015). Folk theories are corrupted by cross-domain explanations. Recuperado de: https://ssrn.com/abstract=2622301.

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. Current Biology, 28, 4014-4021.e8 https://doi.org/10.1016/j.cub.2018.10.053

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1-36. https://doi.org/10.18637/jss.v048.i02

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521-562. https://doi.org/10.1207/s15516709cog2605_1

Sannomiya, M., & Ohtani, K. (2015). Does a dual-task selectively inhibit the metacognitive activities in text revision? *Thinking Skills and Creativity*, *17*, 25-32. https://doi.org/10.1016/j.tsc.2015.04.002

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66. https://doi.org/10.1037/0033-295X.84.1.1

Schuman, H., & Presser, S. (1980). Public opinion and public ignorance: The fine line between attitudes and nonattitudes. *The American Journal of Sociology*, *85*, 1214–1225. https://doi.org/10.1086/227131

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, *6*, 132-137. https://doi.org/10.1111/1467-8721.ep10772899

Schwartz, B. L., Boduroglu, A., & Tekcan, A. I. (2015). Methodological concerns: The feeling-of-knowing task affects resolution. *Metacognition and Learning*, *11*, 305-316. https://doi.org/.1007/s11409-015-9152-4

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. En J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93-113). Cambridge, MA: MIT Press.

Schwarz, N. (2015). Metacognition. En M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbooks in psychology. APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 203-229). Washington, D. C.: APA. http://dx.doi.org/10.1037/14341-006

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition*, *44*, 1127-1137. https://doi.org/10.3758/s13421-016-0624-9

Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. En D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition and education* (pp. 278-298). New York: Routledge/Taylor & Francis Group.

Shea, N., Bold, A., Bang, D. Yeung, N., Heyes, C., & Frith, C. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*, 186-193. https://doi.org/10.1016/j.tics.2014.01.006.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. En P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2937–2942). Austin, TX: Cognitive Science Society.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S., & Fernbach, P. M. (2017). *The knowledge illusion: Why we never think alone*. New York: Riverhead Books.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda based-monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*, 1236-1242. https://doi.org/10.1037/a0023548

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 299-314. https://doi.org/10.1037/0278-7393.30.2.299

Somin, I. (2016). *Democracy and political ignorance: Why smaller government is smarter* (2a ed). Palo Alto, CA: Stanford University Press.

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. En D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44-78). New York: Clarendon Press/Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524021.003.0003

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645-665. https://doi.org/10.1017/S0140525X00003435

Strack, F., & Deutsch, R. (2015). The duality of everyday life: Dual-process and dual system models in social psychology. En M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbooks in psychology. APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 891-927). Washington, D. C.: APA. http://dx.doi.org/10.1037/14341-028

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, *112*, 159-192. https://doi.org/10.1037/0033-295X.112.1.159

Tauber, S. K., & Rhodes, M. G. (2010). Metacognitive errors contribute to the difficulty in remembering proper names. *Memory*, *18*, 522-532. https://doi.org/10.1080/09658211.2010.481818

Thomas, A. K., Lee, M., & Hughes, G. (2016). Introspecting the elusive: The uncanny state of the feeling of knowing. En J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 81-94). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.16

Thompson, V. A. (2009): Dual-process theories: A metacognitive perspective. En J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195).

Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.003.0008

Thompson, V. A., Evans, J. St. B. T., & Campbell, J. I. C. (2013) Matching bias on the selection task: It's fast and it feels good. *Thinking & Reasoning*, *19*, 431-452. https://doi.org/10.1080/13546783.2013.820220

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking and Reasoning*, *20*, 215-244. https://doi.org/10.1080/13546783.2013.869763

Thompson, V. A., Prowse-Turner, J. A., Pennycook, J. (2011). Intuition, reason and metacognition. *Cognitive Psychology*, *63*, 107-140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristic-and-biases tasks. *Memory & Cognition*, *39*, 1275-1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*, 147-168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A, & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3-14. https://doi.org/10.1007/s11409-006-6893-0

Vitriol, J. A., & Marsh, J. K. (2018). The illusion of explanatory depth and endorsement of conspiracy beliefs. *European Journal of Social Psychology, 48*, 955-969. https://doi.org/10.1002/ejsp.2504

Voelkel, J. G., Brandt, M. J., & Colombo, M. (2018). I know that I know nothing. Can puncturing the illusion of explanatory depth overcome the relationship between attitudinal dissimilarity and prejudice? *Comprehensive Results in Social Psychology*. https://doi.org/10.1080/23743603.2018.1464881

Wason, P. C. (1966). Reasoning. En B. M. Foss (Ed.), *New horizons in psychology*: *I* (pp. 106–137). Harmandsworth, UK: Penguin.

Wason, P. C., & Evans, J. (1975). Dual processes in reasoning? *Cognition*, *3*, 141-154. https://doi.org/10.1016/0010-0277(74)90017-1

Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*, 12-22. https://doi.org/10.3758/BF03210553

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132*, 408–428. https://doi.org/10.3200/GENP.132.4.408-428

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring student's calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology, 27*, 551–572. https://doi.org/10.1016/S0361-476X(02)00006-1

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential biases for incubation effects in problem solving. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *13*, 187-205. https://doi.org/10.1037/0278-7393.13.2.187

Zabrucky, K. M. (2010). Knowing what we know and do not know: Educational and real world implications. *Procedia Social and Behavioral Sciences*, *2*, 1266-1269. https://doi.org/10.1016/j.sbspro.2010.03.185

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*, 111-125. https://doi.org/10.1177/1745691613518075

Zeveney, A. S., & Marsh, J. K. (2016). The illusion of explanatory depth in a misunderstood field: The IOED in mental disorders. En A. Papafragou, D. Grodner, D. Mirman, & J.C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting Cognitive Science Society* (pp. 1020-1025). Austin, TX: Cognitive Science Society.

Zhao, Q., & Linderholm, T. (2010). Anchoring effects on prospective and retrospective metacomprehension judgments as a function of peer performance information. *Metacognition and Learning*, *6*, 25-43. https://doi.org/10.1007/s11409-010-9065-1

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*, 493-504. https://doi.org/10.1037/pspa0000056

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*, 75-98. https://doi.org/10.1007/s10683-009-9230-z

Zohar, A., & Ben David, A. (2009). Paving a clear path in a thick forest: A conceptual analysis of a metacognitive component. *Metacognition and Learning*, *4*, 177-195. https://doi.org/10.1007/s11409-009-9044-6

Anexo A. Escala de deseabilidad social y temas del estudio preliminar

(Experimento 1)

Por favor califique los temas en una escala de 7 puntos como se muestra abajo:

Para nada avergonzado						Muy avergonzado
< 1	2	3	4	5	6	7 >

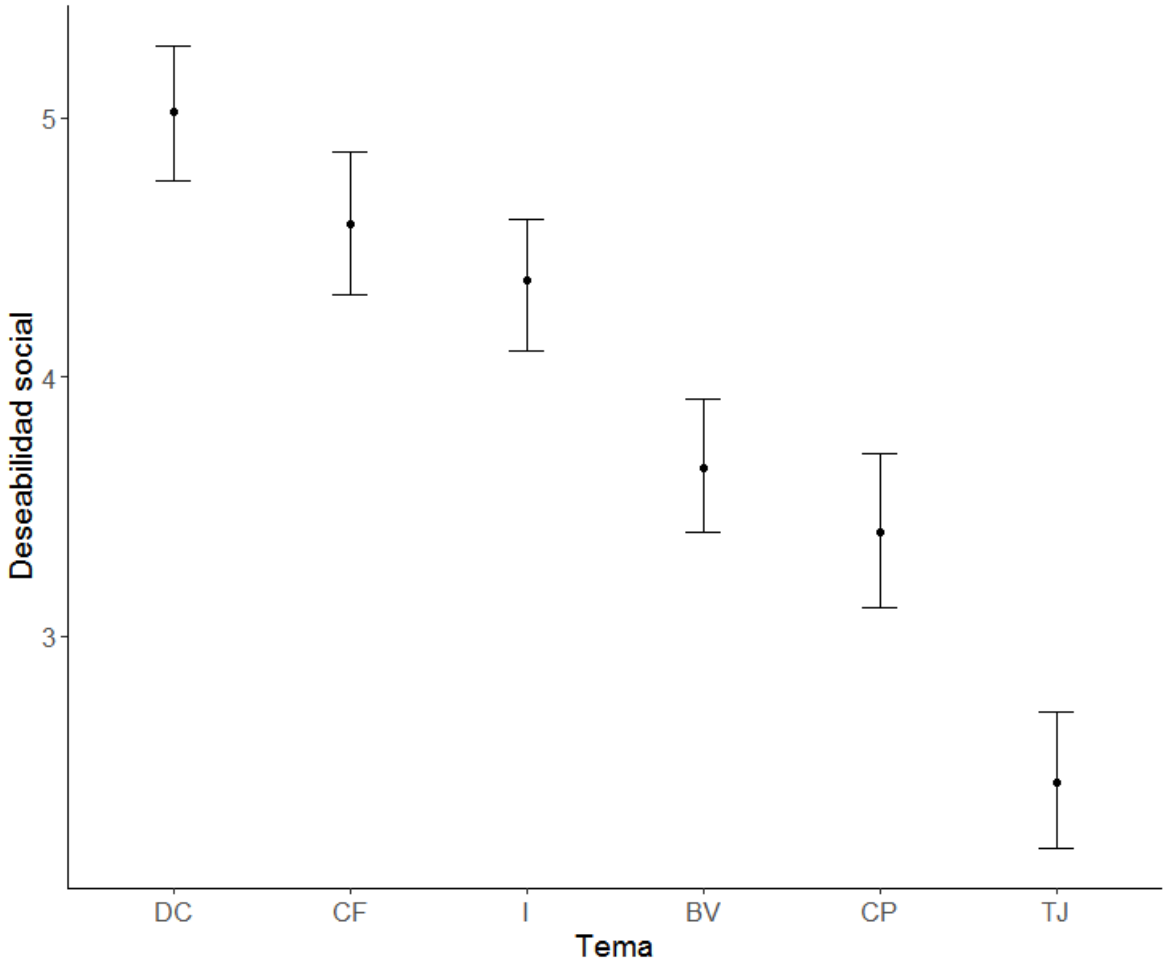
"1" en esta escala significa "Si yo NO tuviera una buena comprensión o conocimiento de este tema, NO me sentiría avergonzado en lo más mínimo".

"7" en esta escala significa "Si yo NO tuviera una buena comprensión o conocimiento de este tema, me sentiría MUY avergonzado".

Tema	**¿Cuán avergonzado?** 1: Para Nada						7: Mucho
Cómo funciona una turbina de propulsión jet	< 1	2	3	4	5	6	7>
Por qué surgieron las FARC en la década de 1960	< 1	2	3	4	5	6	7>
Cómo funciona el lavado de activos	< 1	2	3	4	5	6	7>
Cómo funciona un mouse óptico	< 1	2	3	4	5	6	7>
Por qué surgieron los grupos paramilitares en Colombia en la década de 1980	< 1	2	3	4	5	6	7>
Cómo se calcula la inflación anual del país	< 1	2	3	4	5	6	7>
Por qué fracasaron los diálogos de paz con las FARC en 2002	< 1	2	3	4	5	6	7>
Cómo funciona un control remoto	< 1	2	3	4	5	6	7>
Cómo funciona una burbuja económica	< 1	2	3	4	5	6	7>

Por qué se dieron las alianzas entre grupos paramilitares y políticos durante las décadas de 1990-2000	< 1	2	3	4	5	6	7>
Cómo funciona un reactor nuclear	< 1	2	3	4	5	6	7>
Por qué el conflicto interno colombiano ha sido de larga duración	< 1	2	3	4	5	6	7>
Cómo funcionan las bolsas de valores	< 1	2	3	4	5	6	7>
Cómo funciona un reloj de péndulo	< 1	2	3	4	5	6	7>
Por qué el conflicto interno colombiano ha sido más intenso en las zonas rurales	< 1	2	3	4	5	6	7>
Por qué se genera el déficit fiscal	< 1	2	3	4	5	6	7>
Cómo los bancos calculan las tasas de interés de un crédito para compra de vivienda	< 1	2	3	4	5	6	7>
Cómo funciona un encendedor	< 1	2	3	4	5	6	7>
Por qué se produjo el exterminio de los miembros de la Unión Patriótica en las décadas de 1980-1990	< 1	2	3	4	5	6	7>
Cómo se establece la tasa de cambio entre el peso y las monedas extranjeras	< 1	2	3	4	5	6	7>
Cómo funciona una caña de pescar	< 1	2	3	4	5	6	7>

Anexo B. Deseabilidad social de los temas seleccionados para el estudio principal (Experimento 1).



Nota: DC=Prolongación del conflicto armado interno; CF=Creación de las FARC-EP; I=Aumento de la inflación; BV=Bolsa de valores; CP= Caña de pescar; TJ=Turbina jet. Las barras de error representan los intervalos de confianza de la media al 95%.

Diferencias de deseabilidad entre los 6 temas: F (5,965) = 59.92, p < .001

Pruebas *post hoc* de deseabilidad social entre los temas seleccionados.

Dominio	Ítem	*M* (*EE*)	*F*	η^2_p
Histórico	Prolongación del conflicto armado	5.02 (.13)	16.34**	.078
	Creación de las FARC-EP	4.58 (.13)		
Económico	Aumento de la inflación	4.37 (.13)	38.31**	.16
	Bolsa de valores	3.65 (.13)		
Dispositivos	Caña de pescar	3.40 (.15)	28.78**	.13
	Turbina jet	2.44 (.13)		

Nota: ***p* <.001; *M*= Media; *EE* = Error estándar

Anexo C. Test de verificación de lectura de las instrucciones de la tarea

La mayoría de teorías modernas sobre toma de decisiones reconocen que estas no tienen lugar aisladamente. La preferencia individual y el conocimiento previo, junto con otras variables situacionales, pueden afectar en gran medida la decisión que se tome. Con el fin de facilitar nuestra investigación y asegurar la validez de los datos obtenidos en ella, nos interesa verificar que usted se toma el tiempo necesario para leer las instrucciones. Si no es así, los resultados de nuestro estudio pueden verse afectados o llegar a ser inútiles, ya que dependen de que se sigan las instrucciones correctamente. Con el fin de demostrar que ha leído las instrucciones hasta este punto, por favor ignore las alternativas de deportes presentadas a continuación y seleccione la opción que se encuentra vacía a la derecha para poder continuar con el experimento. Muchas gracias por leer atentamente las instrucciones y ayudarnos a asegurar la calidad de la información recolectada en este estudio.

¿Cuál es su deporte favorito?

Fútbol	Baloncesto	Béisbol	Natación	Ninguno	
○	○	○	○	○	○	○

Anexo D. Instrucciones de la escala de comprensión (Experimento 1)

En el siguiente apartado, le pediremos que califique su comprensión del funcionamiento de diferentes fenómenos o mecanismos en una escala de 7 puntos, donde 1 indica "comprensión vaga o pobre" y 7 "comprensión detallada".

Antes de empezar, le explicaremos el significado de las puntuaciones en esta escala a través de un ejemplo.

A continuación, usted encontrará tres ejemplos de explicaciones que se ubican en diferentes niveles de comprensión acerca de **la Rebelión de los comuneros/una burbuja económica/una ballesta**. En general, mientras más información se suministre sobre las causas del fenómeno, más alto será el puntaje en la escala.

Por favor lea atentamente cada explicación con el fin de comprender cómo usar la escala. Como podrá notar, 7 implica una amplia y detallada comprensión de las causas de **la Rebelión de los comuneros/una burbuja económica/una ballesta**; 1 significa un entendimiento vago y poco detallado y 4 una comprensión intermedia.

a) Dominio histórico

Nivel 7 de comprensión: *"La rebelión se desencadenó a raíz del alza en los impuestos a víveres esenciales (sal, tabaco, algodón) y a las transacciones comerciales, establecida por la Corona Española para financiar la guerra contra Inglaterra. Entre otras cosas, estas medidas buscaban expropiar y vender tierras de los resguardos indígenas para aumentar los ingresos de la Corona y reducir el poder económico de las élites criollas. A lo anterior se sumó el malestar social ocasionado por la violencia con la que se hicieron los recaudos y el resentimiento de los criollos al no poder acceder a cargos administrativos dentro de las instituciones coloniales. Todo esto facilitó la unión de las élites criollas con los campesinos e indígenas productores en contra de la forma de gobierno colonial, aunque tuviesen intereses económicos y sociales diferentes. El levantamiento inicial ocurrió en la localidad del Socorro, aunque el movimiento se extendió posteriormente a otras regiones del país."*

Nivel 4 de comprensión: *"La rebelión se produjo por el alza en impuestos que ordenó la Corona Española y los procedimientos abusivos que se emplearon para cobrarlos a los comerciantes criollos. Esto produjo la unión de varias provincias en contra el mal gobierno y el alzamiento de campesinos e indígenas, liderados por los criollos."*

Nivel 1 de comprensión: *"La causa de la revuelta fue alza de impuestos a los comerciantes criollos, impuesta por la Corona española."*

b) Dominio económico

Nivel 7 de comprensión: *"Una burbuja económica se produce cuando hay un aumento rápido y sostenido en el tiempo del precio de un bien o una acción por encima de su valor real. Esta alza normalmente se debe a que algunos especuladores empiezan a acumular el*

bien o la acción, aumentando la demanda y elevando el precio. A medida que aparecen nuevos compradores, los inversionistas iniciales venden a precios altos y la demanda continúa aumentando. Cuando el precio se eleva al punto en el que ya no aparecen nuevos compradores, todos los inversionistas tratan de vender simultáneamente para evitar pérdidas. Al aumentar súbitamente la oferta, la burbuja explota y los precios empiezan a bajar rápidamente, generando pérdidas a quienes compraron el bien o la acción a los precios más altos."

Nivel 4 de comprensión: *"Una burbuja económica se genera cuando hay un incremento progresivo en el precio de algún producto o acción, que se aleja de su precio real. Los precios suben rápidamente y luego bajan repentinamente cuando la burbuja explota. Esto genera pérdidas a quienes esperaban vender a precios más altos que el precio al que compraron."*

Nivel 1 de comprensión*: "Una burbuja económica se da cuando los precios de un producto suben cada vez más hasta que en algún punto la burbuja explota."*

c) Dominio de dispositivos

Nivel 7 de comprensión: *"La ballesta consta de un arco flexible. un soporte rígido de madera o metal con una ranura en la mitad, una cuerda o banda elástica y un mecanismo de liberación. El arco está montado de forma fija en el soporte. La cuerda se puede tensar hacia atrás por medio de una palanca, un pequeño bloque de poleas o por una manivela enrollada alrededor de un carrete que tira de otro cable conectado a la parte delantera del soporte fijo. Antes de cada disparo, la cuerda se tensa mediante dicho mecanismo y se asegura su posición con la ayuda de un pasador. A menudo, el pasador se bifurca de modo que el extremo posterior de la flecha pueda reposar directamente sobre la superficie de la cuerda. El pasador está conectado directamente con el gatillo a través de un perno, de manera que cuando se tira del gatillo, el perno gira hacia abajo junto con el pasador y libera la cuerda conectada al arco. Cuando el pasador libera la cuerda, el arco se destensa instantáneamente, transmitiendo toda la energía almacenada en el arco flexionado a la flecha.*"

Nivel 4 de comprensión: *"La ballesta es un arco montado sobre un soporte fijo que impulsa una flecha con una potencia mayor que un arco normal, gracias a un mecanismo que le permite tensar una cuerda muy fuerte y lanzar con fuerza la flecha. También permite disparar flechas con mayor rapidez, arrastrando de nuevo la cuerda y disparando la siguiente flecha por medio de un gatillo."*

Nivel 1 de comprensión: *"Una ballesta es un aparato que sirve para disparar flechas y funciona de manera parecida a un arco ubicado horizontalmente."*

Instrucciones para la formulación de explicaciones

Ahora queremos conocer su explicación sobre algunos de estos temas. Por favor describa con todos los detalles posibles las causas de _____.

El objetivo de esta explicación es mostrar claramente cómo cada paso o momento ocasiona el siguiente y ubicarlos en un orden secuencial desde la aparición de las causas hasta el momento en que el fenómeno ocurre. En otras palabras, trate de contar una historia tan completa como pueda (sin vacíos o lagunas) y que pueda ser entendida por cualquier persona.

Por favor, tómese el tiempo que considere necesario. Nosotros esperamos contar con su mejor explicación.

Anexo E. Escala de deseabilidad social del conocimiento (Experimento 2 y 2a)

1. ¿Qué tan avergonzado se sentiría si alguien le pide una explicación del tema y usted NO tuviera una buena comprensión o conocimiento de…

2. ¿Qué tan importante es para usted tener una buena comprensión o conocimiento de…

3. ¿Qué tan positiva sería la opinión de las demás personas sobre usted si demuestra tener una buena comprensión de…

4. ¿Qué tan negativa sería la opinión de las demás personas sobre usted si demuestra NO tener una buena comprensión de…

Anexo F. Deseabilidad social de los temas relacionados con la historia del conflicto armado colombiano (Experimento 2)

Tema	*M*	*DE*
Creación de las FARC	4.65	1.33
Creación de grupos paramilitares	4.55	1.29
Fracaso del proceso de paz de El Caguán	4.22	1.25
Alianzas entre políticos y paramilitares	4.03	1.23
Prolongación del conflicto armado interno	5.24	1.27

Intensidad del conflicto en zonas rurales	5.06	1.26
Financiación del Plan Colombia	3.85	1.31
Exterminio de la Unión Patriótica	4.03	1.43
Creación del ELN	4.29	1.39
Creación del movimiento a favor de la Constituyente de 1991	4.03	1.36
Éxito de la desmovilización del M-19	4.33	1.27
Finalización de la dictadura de Rojas Pinilla	3.92	1.39
Creación del Quintín Lame	3.49	1.46
Cese al fuego de 1984 entre el gobierno y las guerrillas	3.57	1.24
Reforma agraria de 1960	3.35	1.32
Estatuto de seguridad de 1982	3.12	1.30

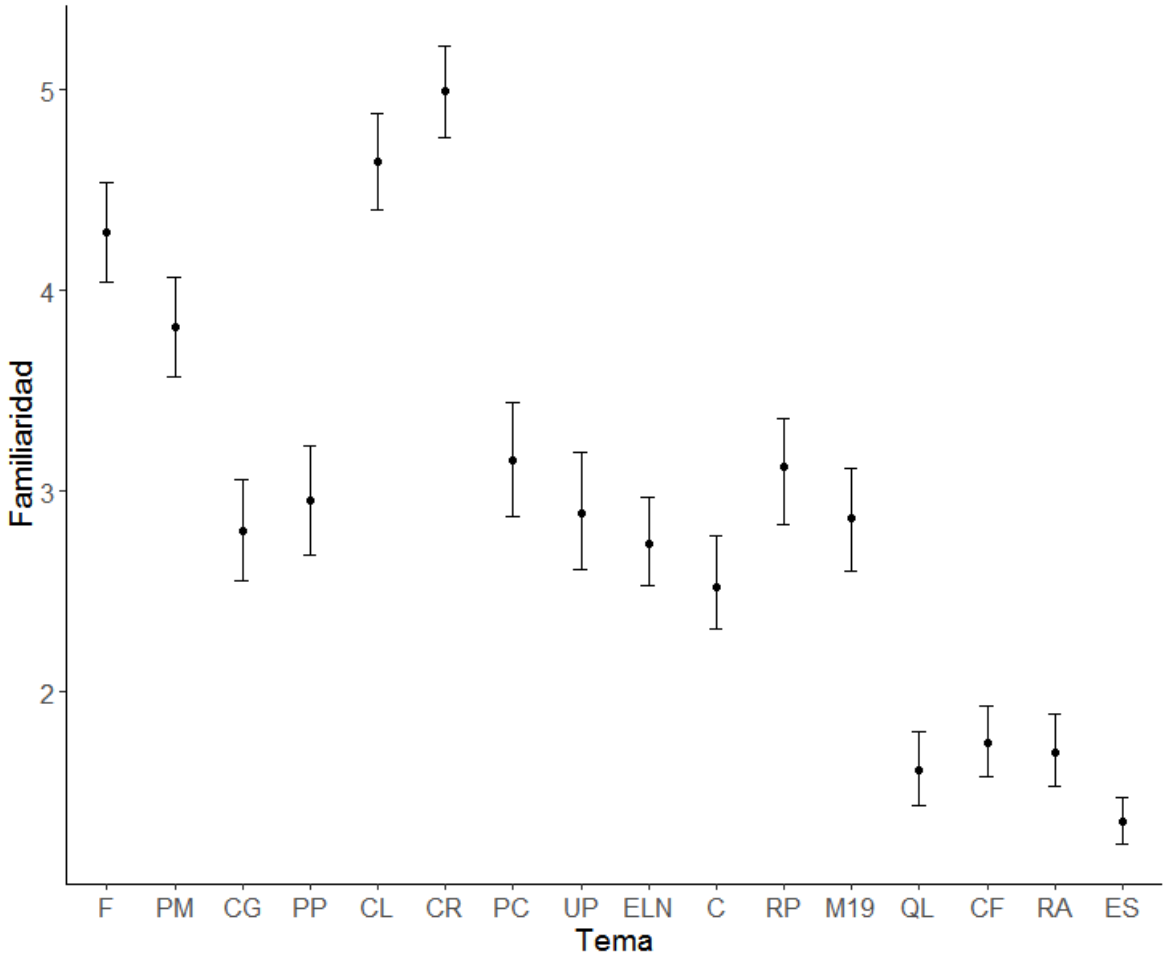
Nota: *M*=Media; *DE*=Desviación estándar

Comparación de medias de deseabilidad social del conocimiento de los cuatro temas seleccionados para el Experimento 2.

Tema	Tema	*t* (177)	*d* de Cohen
Prolongación del conflicto	Conflicto zonas rurales	3.33**	0.14
Prolongación del conflicto	Plan Colombia	14.34**	1.07
Prolongación del conflicto	Rojas Pinilla	13.97**	.99
Conflicto zonas rurales	Plan Colombia	12.44**	0.94
Conflicto zonas rurales	Rojas Pinilla	11.35**	0.86
Plan Colombia	Rojas Pinilla	0.69	0.05

Nota: **= p <.01

Anexo G. Familiaridad con los temas presentados en el Experimento 2.



Nota: F=Creación de las FARC; PM=Creación de grupos paramilitares; CG=Fracaso del proceso de paz de El Caguán; PP=Alianzas entre políticos y paramilitares; CL=Prolongación del conflicto armado interno; CR=Intensidad del conflicto en zonas rurales; PC=Financiación del Plan Colombia; UP=Exterminio de la Unión Patriótica; ELN=Creación del ELN; C=Creación del movimiento a favor de la Constituyente de 1991; RP=Finalización de la dictadura de Rojas Pinilla; M19=Éxito de la desmovilización del M-19; QL=Creación del Quintín Lame; CF=Cese al fuego de 1984 entre el gobierno y las guerrillas; RA=Reforma agraria de 1960; ES=Estatuto de seguridad de 1982. Las barras de error representan los intervalos de confianza de la media al 95%.

Anexo H. Medias y confiabilidad de la escala de deseabilidad social de los temas seleccionados en el Experimento 2.

Tema	*M*	*DE*	I. C. 95 %	α de Cronbach (*n*=178)
Prolongación del conflicto	4.50	1.59	[4.14, 4.85]	.74
Intensidad del conflicto en zonas rurales	4.93	1.47	[4.60,5.26]	.73
Financiación del Plan Colombia	2.93	1.77	[2.53, 3.33]	.77
Finalización de la dictadura de Rojas Pinilla	2.92	1.68	[2.54, 3.30]	.78

Nota: *M*= Media; *DE*=Desviación estándar

Anexo I. Medias y confiabilidad de la escala de deseabilidad social de los temas seleccionados en el Experimento 2a.

Tema	*M*	*EE*	I.C. (95%)	α de Cronbach (*n*=189)
Creación de las FARC-EP	4.83	.09	[4.66-5.02]	.74
Creación de los grupos paramilitares	4.68	.10	[4.50-4.88]	.78
Prolongación del conflicto armado	5.21	.08	[5.04-5.38]	.72
Intensidad del conflicto armado en zonas rurales	5.10	.09	[4.93-5.26]	.74
Alianzas entre paramilitares y políticos	4.04	.10	[4.21-4.60]	.77
Movimiento Constituyente de 1990	4.38	.11	[4.17-4.59]	.80
Genocidio de la Unión Patriótica	4.27	.01	[4.10-4.47]	.77
Plan Colombia	3.97	.09	[3.79-4.14]	.72

Nota: *M*= Media; *EE*=Error Estándar; I.C.= Intervalo de confianza de la media

Anexo J. Prueba de *overclaiming* sobre conocimiento de la historia del conflicto armado colombiano.

A continuación, usted encontrará una lista de hechos históricos. Por favor indique qué tan bien informado está usted sobre cada uno de estos acontecimientos. Por favor califique cada tema en una escala de 7 puntos, donde 1 en esta escala significa “Para NADA estoy informado sobre este hecho histórico." y 7 “Estoy MUY bien informado sobre este hecho histórico".

Ítems reales

El secuestro de la excandidata presidencial Ingrid Betancourt en el departamento de Caquetá en 2002.

El proceso de negociación y desmovilización de grupos paramilitares en Colombia que tuvo lugar entre 2003 y 2006.

La Asamblea Nacional Constituyente del año 1991 celebrada durante el mandato del expresidente César Gaviria.

Las ejecuciones extrajudiciales o "falsos positivos" de jóvenes provenientes del municipio de Soacha en 2008.

Los acontecimientos del 9 de Abril de 1948 tras el asesinato del candidato presidencial Jorge Eliécer Gaitán.

La muerte del cura guerrillero Camilo Torres en combates con el ejército nacional en 1966.

El ataque al campamento de alias "Raúl Reyes" por parte del ejército colombiano en territorio ecuatoriano en 2008.

La instalación de los diálogos de paz entre el gobierno nacional y las FARC en Oslo en 2012.

La" Operación Jaque", que permitió el rescate de 15 secuestrados en poder de las FARC en el año 2008.

El asesinato del comediante, político, actor y presentador Jaime Garzón en Bogotá en el año 1999.

El atentado terrorista al Club El Nogal en Bogotá por parte de la guerrilla de las FARC en 2006.

La toma del palacio de justicia realizada por integrantes del grupo guerrillero M-19 en el año 1985.

El inicio de la fase pública de diálogos de paz entre el ELN y el gobierno nacional en 2017.

La firma del acuerdo de paz entre las FARC y el gobierno colombiano en la ciudad de Cartagena en 2016.

La masacre en el municipio de Bojayá a causa de fuego cruzado entre las FARC y grupos paramilitares en 2002.

El cese al fuego bilateral entre el gobierno nacional y las FARC en 2014, producto de los diálogos de paz.

Ítems falsos

El asesinato del ministro de gobierno del expresidente Álvaro Uribe en 2003 por parte de las FARC.

El asesinato de líderes estudiantes en la ciudad de Manizales por parte de paramilitares entre 2000 y 2005.

El secuestro y posterior liberación de tres periodistas estadounidenses perpetrado por el M-19 en 1985.

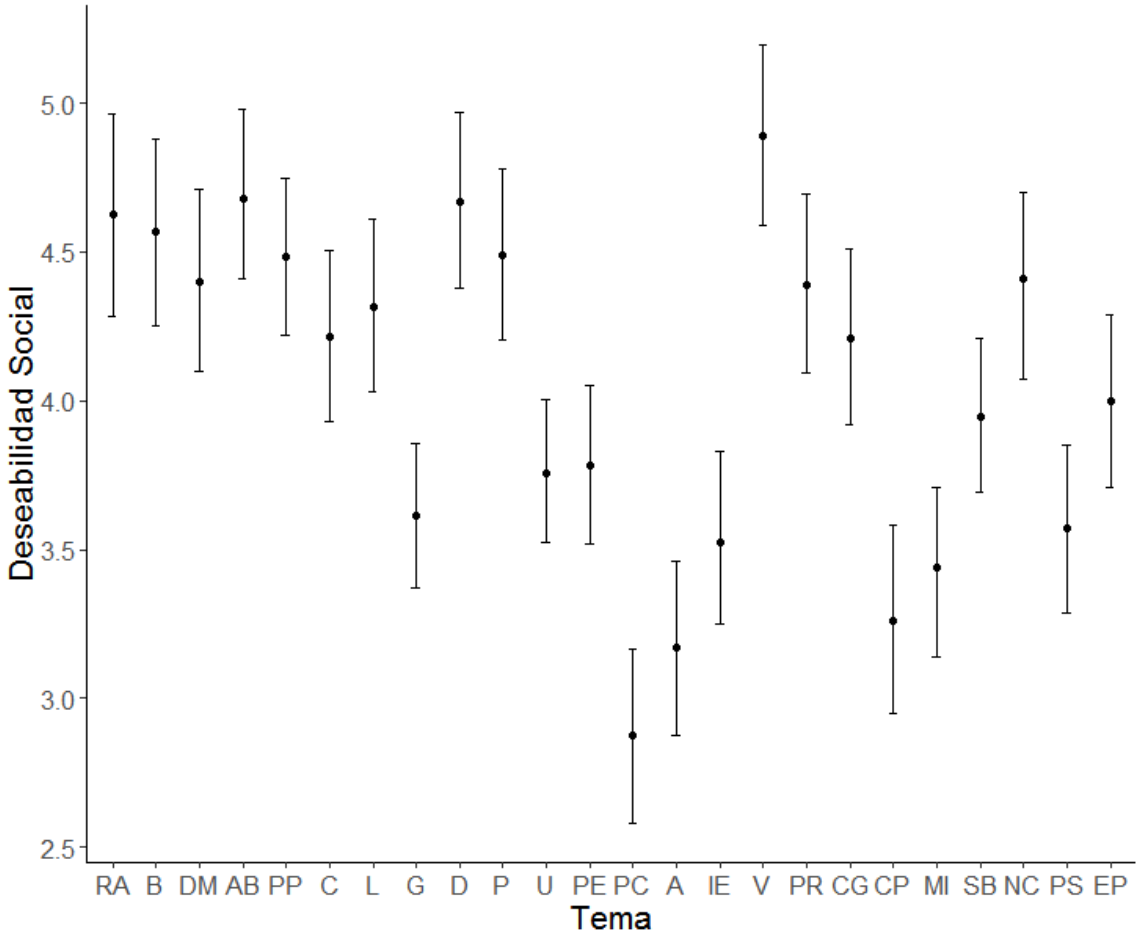
La toma guerrillera a la capital del departamento de Arauca por parte del ELN en 1999.

Anexo K. Confiabilidad de la escala de deseabilidad social de los temas del estudio preliminar (Experimento 3)

Tema	α de Cronbach (*n*=70)
Acuerdo de reforma agraria integral	.782
Distribución de baldíos entre campesinos y víctimas del conflicto	.73
Apoyo socioeconómico a desmovilizados	.778
Sentencia de la Corte Constitucional sobre el aborto	.711
Programa “Ser pilo paga”	.712
Modelo de medición de grupos de investigación de Colciencias	.754
Reforma a los programas universitarios de licenciaturas	.711
Fomento del consumo informado de bebidas azucaradas	.571
Acuerdo sobre solución al problema de las drogas ilícitas	.738
Nuevo código de policía	.72
Reglamentación del servicio Uber	.620
Estímulos a pago electrónico	.698
Obras públicas en honor a los próceres	.671
Ley de derechos laborales para actores y actrices	.767

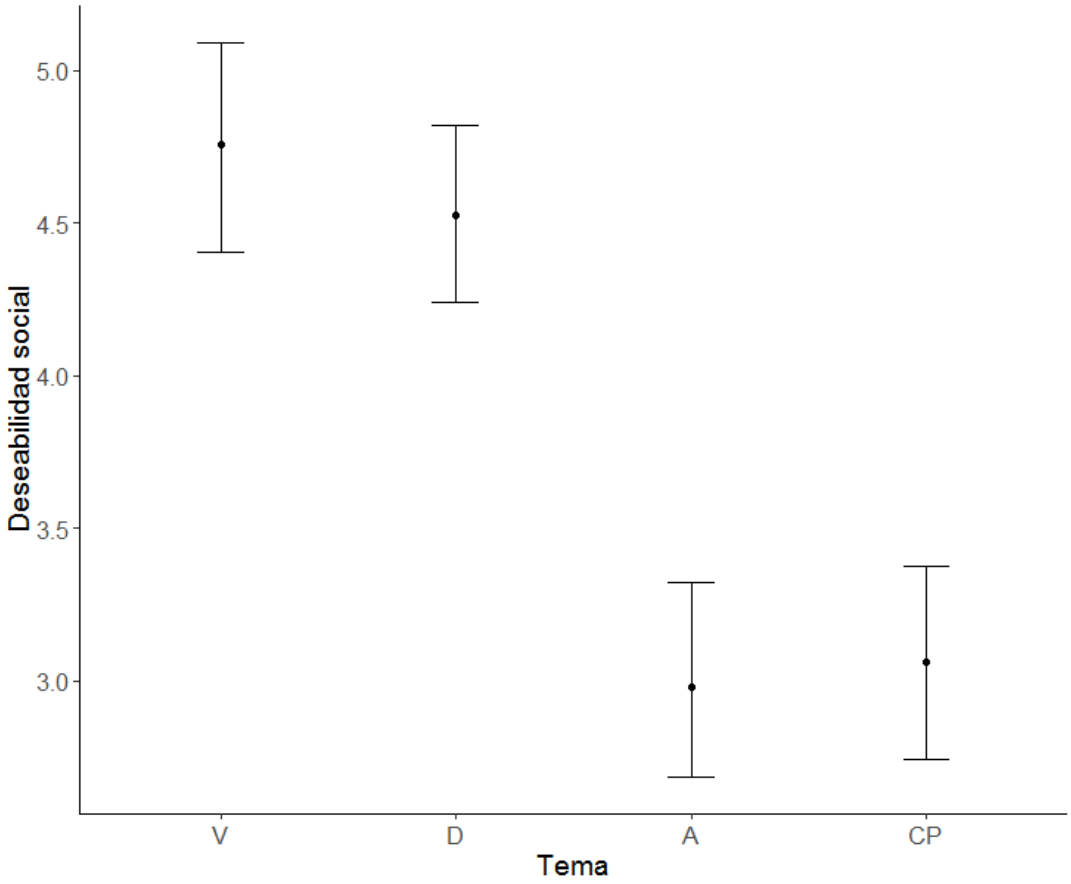
Descuento para pago electrónico de impuestos a pequeñas empresas	.736
Acuerdo sobre víctimas y jurisdicción especial para la paz	.785
Facultades especiales del presidente para la implementación de los acuerdos de paz	.707
Ley de cuotas para la equidad de género en las listas para corporaciones públicas	.763
Requisitos para la convalidación de títulos de cirugía plástica obtenidos en el exterior	.779
Normas que regulan el ejercicio de la medicina interna	.749
Destinación de saldos de cuentas abandonadas al interés social	.664
Reducción del número de congresistas	.758
Ley de protección de pescadores artesanales	.668
Reforma al equilibrio de poderes	.674

Anexo L. Deseabilidad social de los ítems seleccionados en el Experimento 3.



Nota: RA=Acuerdo de reforma rural integral; B=Distribución de baldíos entre campesinos y víctimas del conflicto; DM=Apoyo socioeconómico a desmovilizados; AB=Sentencia de la Corte Constitucional sobre el aborto; PP=Programa "Ser pilo paga"; C=Modelo de medición de grupos de investigación de Colciencias; L= Reforma a los programas universitarios de licenciaturas; G=Fomento del consumo informado de bebidas azucaradas; D=Acuerdo sobre solución al problema de las drogas ilícitas; P=Nuevo código de policía; U=Reglamentación del servicio Uber; PE=Estímulos a pago electrónico; PC=Obras públicas en honor a los próceres; A= Ley de derechos laborales para actores y actrices; IE= Descuento para pago electrónico de impuestos a pequeñas empresas; V=Acuerdo sobre

víctimas y jurisdicción especial para la paz; PR=Facultades especiales del presidente para la implementación de los acuerdos de paz; CG=Ley de cuotas para la equidad de género en las listas para corporaciones públicas; CP=Requisitos para la convalidación de títulos de cirugía plástica obtenidos en el exterior; MI=Normas que regulan el ejercicio de la medicina interna; SB=Destinación de saldos de cuentas abandonadas al interés social; NC=Reducción del número de congresistas; PS=Ley de protección de pescadores artesanales; EP=Reforma al equilibrio de poderes.



Medias de deseabilidad social del conocimiento de los cuatro temas seleccionados para el experimento 3. Nota: V= Acuerdo de víctimas; D=Acuerdo de solución al problema de las drogas; A=Ley del actor; CP= Regulación de títulos para ejercer la cirugía plástica. Las barras de error representan los intervalos de confianza de la media al 95%.

Diferencias de deseabilidad entre los 4 temas: F (3,207) = 55.89, p < .001

Medias marginales estimadas de la deseabilidad social del conocimiento de los temas utilizados en el experimento 3.

Tema	*M*	*EE*	I.C. 95%
Acuerdo de víctimas	4.75	.17	[3.98 - 5.00]
Acuerdo de solución al problema de las drogas	4.52	.16	[3.70 - 4.71]
Ley del actor	2.98	.16	[2.63 - 3.61]
Regulación de títulos para ejercer la cirugía plástica	3.06	.16	[2.75 - 3.63]

Nota: *EE*: Error estándar; IC 95%=Intervalo de confianza al 95%

Anexo M. Instrucciones de la escala de comprensión (Experimento 3)

En el siguiente apartado, le pediremos que califique su comprensión del funcionamiento de diferentes fenómenos o mecanismos en una escala de 7 puntos, donde 1 indica "comprensión vaga o pobre" y 7 "comprensión detallada".

A continuación, usted encontrará tres ejemplos de explicaciones que se ubican en diferentes niveles de comprensión de la Ley 100 de 1993.

Por favor lea atentamente cada explicación con el fin de comprender cómo usar la escala. Como podrá notar, 7 implica una amplia y detallada comprensión; 1 significa un entendimiento vago y poco detallado y 4 una comprensión intermedia.

Nivel 7 de comprensión: "La Ley 100 regula el acceso y la prestación de servicios de seguridad social para todos los ciudadanos. Esta ley crea el Sistema de Seguridad Social Integral, el cual tiene cuatro componentes: salud, pensiones, riesgos profesionales y servicios complementarios.

La ley 100 autoriza la creación de empresas privadas que administren servicios en salud (EPS y ARS). Los ciudadanos pueden escoger libremente la empresa a la cual desean afiliarse junto con su núcleo familiar, al tiempo que el Estado debe vigilarlas a través de la Superintendencia de la Salud y el Ministerio de la Protección Social. Quienes tengan ingresos fijos deben aportar al financiamiento del sistema (régimen contributivo), mientras que aquellos sin capacidad de pago son financiados principalmente por fondos estatales (régimen subsidiado). El plan obligatorio de salud (POS) garantiza que todos los afiliados reciban los mismos servicios básicos, independientemente de sus ingresos.

La ley 100 también permite la creación de fondos privados de pensiones, un fondo estatal de solidaridad pensional para subsidiar los aportes de los trabajadores de bajos ingresos, seguros de riesgos profesionales pagados por los empleadores y otros beneficios como la afiliación a las cajas de compensación familiar. Todo esto permite la participación de la empresa privada en la oferta de servicios de seguridad social, los cuales eran administrados por el Estado antes de la aprobación de esta ley."

Nivel 4 de comprensión: "La ley 100 reforma el sistema de seguridad social en Colombia, permitiéndole a empresas privadas manejar los recursos para pensión, salud y riesgos profesionales. Con esto se buscaba mejorar la calidad de los servicios y garantizar una cobertura universal en seguridad social, independientemente de los ingresos de cada persona.

Esta ley crea los regímenes subsidiado y contributivo en salud, que permite que los ciudadanos aporten al sistema de acuerdo con sus ingresos. También autoriza la creación de administradoras de riesgos profesionales y otros servicios. Los usuarios pueden elegir las empresas a las cuales desean afiliarse y estas empresas están bajo la supervisión de los organismos de control del Estado."

Nivel 1 de comprensión: "La ley 100 es una reforma a la salud que creó a las EPS y a los fondos de pensiones."

Instrucciones para la formulación de explicaciones

Ahora queremos conocer su explicación sobre algunas de estas políticas.

Por favor describa con todos los detalles posibles *el acuerdo sobre víctimas y jurisdicción especial para la paz firmado en la mesa de diálogos de la Habana*, de tal manera que una persona que no esté informada sobre esta política pueda tener una comprensión completa de la misma después de leer su explicación.

El propósito de esta explicación es mostrar claramente los elementos que componen esta política y la manera en que se relacionan entre sí. En otras palabras, trate de hacer una exposición tan completa como pueda (sin vacíos o lagunas) y que pueda ser entendida por cualquier persona.

Para nosotros es muy importante que su respuesta esté basada únicamente en su conocimiento sobre el tema. Por lo tanto, le pedimos abstenerse de consultar páginas de internet o cualquier otra fuente adicional de información (libros, revistas, apuntes, conversaciones, etc.) mientras elabora su explicación.

Por favor, tómese el tiempo que considere necesario. Nosotros esperamos contar con su mejor explicación.

Anexo N. *Cognitive Reflection Test* (CRT)

Un bate y una pelota cuestan 1.10 dólares en total. El bate cuesta 1 dólar más que la pelota. ¿Cuánto cuesta la pelota?

Si a 5 máquinas les toma 5 minutos fabricar 5 dispositivos, ¿Cuántos minutos le tomaría a 100 máquinas fabricar 100 dispositivos?

En un lago hay una capa de algas. Cada día, la capa dobla su tamaño. Si en 48 días las algas cubren toda la superficie del lago, ¿Cuántos días tardarán en cubrir la mitad del lago?

Anexo O. Información enviada a los participantes al concluir el Experimento 3

¿Cómo funciona un encendedor? ¿Por qué se producen los terremotos? ¿A qué se deben las burbujas económicas? La familiaridad que tenemos con fenómenos como estos y la posibilidad de obtener respuesta a estas preguntas con sólo hacer un click, nos produce la sensación de que tenemos una comprensión suficiente sobre ellos y de que en cualquier momento podemos explicar en qué consisten y cuál es el proceso a través del cual se producen. Pero una vez lo intentamos, nos damos cuenta de que nuestra comprensión abarca apenas la punta del iceberg, que el resto, del que nos creíamos conocedores, pareció derretirse de pronto en cuanto tuvimos que dar muestra de nuestra capacidad explicativa frente a situaciones presentes en nuestra vida cotidiana. Esta diferencia entre lo que

creemos saber y lo que somos capaces de explicar, fue denominada Ilusión de Profundidad Explicativa por Leonid Rozenblit y Frank Keil, profesores de la Universidad de Yale.

Las personas formamos opiniones que creemos están respaldadas por información completa. Sin embargo, esto sería solo producto de un efecto ilusorio, pues al no ser capaces de dar explicaciones sobre el tema nos damos cuenta que no sabíamos tanto como pensábamos, disminuye el convencimiento que teníamos. Pero no sólo las opiniones se ven influenciadas por la Ilusión de Profundidad Explicativa sino también la intensidad con que sostenemos nuestra posición, e incluso este efecto puede ser la base de las posturas políticas más extremas puesto que confundimos la familiaridad que tenemos sobre algunos temas con un conocimiento real de ellos.

En este sentido, este estudio del cual usted hizo parte buscaba indagar por su comprensión sobre varias políticas que ya fueron aprobadas por el Congreso o que están siendo tramitadas por el gobierno nacional. De manera que si usted creía comprender sobre la política que se le presentó, pero luego cuando tuvo que explicar el funcionamiento de la misma modificó su nivel de comprensión, entonces usted experimentó la Ilusión de Profundidad Explicativa.

Como seguramente habrá notado, hubo una considerable variedad en la lista de temas en la primera parte encuesta, de los cuales algunos pudieron parecerle importantes y otros no tanto; esta gama temática no fue colocada de manera fortuita; más bien corresponde con el interés del estudio. Nuestra hipótesis es que los temas más atrayentes suelen producirnos mayor Ilusión de Profundidad Explicativa, es decir que cuanto más importante consideremos un tema, más creeremos saber de él. De esta forma, con la primera parte de la encuesta queríamos averiguar cuáles eran los temas más deseables para los participantes, mientras que con la segunda parte buscamos verificar si el efecto se producía en mayor proporción con estos temas.

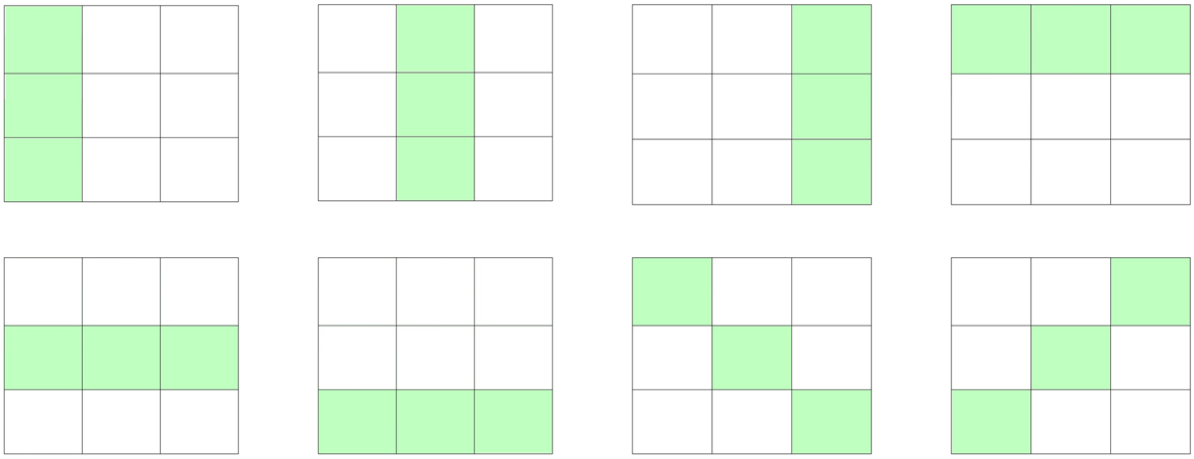
Por último, esperamos que esta información haya sido satisfactoria para dar sentido a la colaboración que usted nos brindó. Si desea conocer más resultados del estudio, no dude en comunicarse a través de este correo.

Agradecemos nuevamente su paciencia y su compromiso con este estudio. Sus respuestas constituyen un insumo indispensable en la construcción de ese trabajo y el diseño de futuras investigaciones.

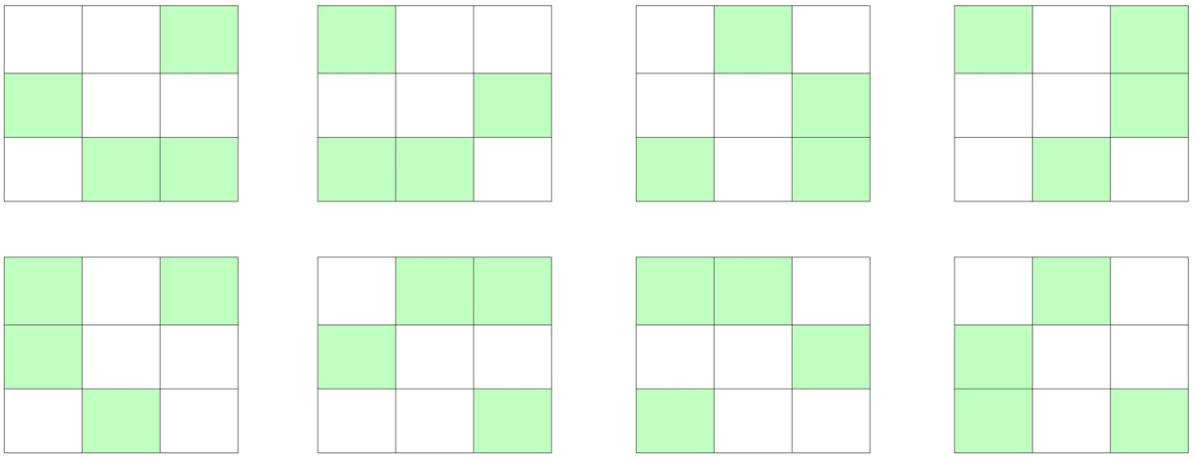
Anexo P. Materiales utilizados en las tareas de memoria del Experimento 4

Matrices utilizadas en la tarea de memoria visoespacial.

Sobrecarga baja



Sobrecarga alta



Pares de sílabas utilizadas en la tarea de memorización de sílabas.

BLA	PRI	MIN	MAX	COM	NAL	REL	SEN
COR	NUM	SAR	DIF	VAR	MED	FAM	TRA

Anexo Q. Medias y confiabilidad de la escala de deseabilidad social de los temas utilizados en el Experimento 4.

Tema	*M*	*EE*	I.C. (95%)	α de Cronbach (*n*=250)
Creación de las FARC-EP	4.76	.08	[4.60, 4.92]	.77
Creación de los grupos paramilitares	4.60	.08	[4.45, 4.76]	.75
Prolongación del conflicto armado	4.98	.08	[4.8, 5.13]	.70
Intensidad del conflicto armado en zonas rurales	4.99	.08	[4.84, 5.15]	.72
Alianzas entre paramilitares y políticos	4.38	.08	[4.28, 4.60]	.75
Movimiento Constituyente de 1990	4.23	.08	[4.06, 4.4]	.74
Genocidio de la Unión Patriótica	4.1	.09	[3.93, 4.28]	.76
Plan Colombia	4.01	.08	[3.85, 4.17]	.72

Nota: *M*= Media; *EE*=Error Estándar; I.C.= Intervalo de confianza de la media

Anexo R. Implicaciones Éticas

Los estudios incluidos en esta disertación se ajustan a la Resolución No. 8430 de 1993 del Ministerio de Salud y los artículos 50 y 51 de la Ley 1090 de 2006. En cumplimiento de esta norma, todos los participantes fueron mayores de edad que expresaron su consentimiento informado antes de participar en la investigación. Cuando los participantes fueron menores de edad, sus padres o acudientes enviaron el consentimiento diligenciado por escrito. La información personal suministrada no fue utilizada para la identificación posterior de los participantes. Adicionalmente, se informó que la participación en los estudios no intentaba realizar alguna modificación permanente de factores biológicos, psicológicos o sociales del individuo, ni trataba aspectos sensibles de su identidad o comportamiento. Los participantes tenían la posibilidad de abandonar el estudio en cualquier momento sin que esto implicase para ellos alguna consecuencia negativa. Los incentivos académicos o en especie asociados a la participación en estos estudios no dependieron en modo alguno de las respuestas dadas. Al final de las sesiones, los participantes que así lo solicitaron recibieron información sobre los propósitos de la investigación.