UNIVERSIDAD DE GRANADA

Tesis Doctoral Internacional – International PhD Thesis

Programa de Doctorado en Psicología

ERRORES DEDUCTIVOS EN ESCOLARES Y ADULTOS: RELACIÓN ENTRE SESGOS EGOCÉNTRICOS Y PENSAMIENTO SOBRE LO CONTRARIO A LA REALIDAD

DEDUCTIVE ERRORS IN SCHOOL-AGE CHILDREN AND ADULTS: THE RELATION BETWEEN EGOCENTRIC BIASES AND COUNTERFACTUAL THOUGHT

Presentada por

Cristina Gordo Gordo

Director

Sergio Moreno Ríos

Departamento de Psicología Evolutiva y de la Educación

Granada, 14 de junio de 2021

Editor: Universidad de Granada. Tesis Doctorales
Autor: Cristina Gordo Gordo
ISBN: 978-84-1306-992-0
URI: http://hdl.handle.net/10481/70148

ERRORES DEDUCTIVOS EN ESCOLARES Y ADULTOS: RELACIÓN ENTRE SESGOS EGOCÉNTRICOS Y PENSAMIENTO SOBRE LO CONTRARIO A LA REALIDAD

DEDUCTIVE ERRORS IN SCHOOL-AGE CHILDREN AND ADULTS: THE RELATION BETWEEN EGOCENTRIC BIASES AND COUNTERFACTUAL THOUGHT

Tesis doctoral presentada por Crsitina Gordo Gordo en el Departamento de Psicología Evolutiva y de la Educación de la Universidad de Granada, dentro del Programa Oficial de Doctorado en Psicología, para aspirar al grado de "Doctor en Psicología" con mención Internacional.

Doctoral dissertation presented by Cristina Gordo Gordo in the Department of Developmental and Educational Psychology in the University of Granada (Programa Oficial de Doctorado en Psicología) to obtain the degree of "PhD in Psychology" with International distinction.

La tesis ha sido realizada bajo la dirección del Catedrático Sergio Moreno Ríos quién avala la calidad de la misma, así como la formación de la doctoranda para aspirar al grado de doctor.

The doctoral dissertation has been supervised by Dr. Sergio Moreno Ríos, who is acting as guarantor for the quality of the thesis and the aspirant's qualification to obtain the PhD degree.

En Granada, a 14 de Junio de 2020

Fdo:

Cristina Gordo Gordo
(La doctoranda)

Sergio Moreno Ríos
(El director)

RECONOCIMIENTOS

El presente trabajo de investigación ha sido financiado a través del Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i (FPI 2016; Ministerio de Ciencia e Innovación), al que ha estado vinculada la doctoranda desde abril de 2017 hasta julio de 2021 (Ref. BES-2016-O78358); así como por los proyectos "*Errores y antídotos al deducir con información ambigua y falsa: desarrollo en escolares y juicios sociales de adultos*" (Ref. PSI2015-63505-P) y "E*l desarrollo deductivo de conclusiones sesgadas: implicaciones en aprendizaje escolar y en contextos sociales*" (Ref. PGC2018-095868-B-I00), concedidos por el Ministerio de Economía, Industria y Competitividad.

A mi “chumi”, mi flor del ébano,
mi abuela…

AGRADECIMIENTOS

Mentiría si digo que no he pensado en este momento, en las cosas que diría, las personas a las que agradecería su apoyo, su cariño, comprensión, etc. Sin embargo, llegado el momento no tengo palabras suficientes para expresar lo agradecida que os estoy. No obstante, lo voy a intentar… GRACIAS:

A mi **madre y a mi padre**, por estar junto a mí en cada momento importante de mi vida, tanto bueno como malo. Por apoyarme siempre, por confiar siempre en mí, por no dejar nunca que me rindiese (ni en esto, ni en nada), por vuestra paciencia, por vuestros consejos, por vuestro cariño,… Necesitaría otras 200 páginas para poder simplemente enumerar todas las cosas que he de agradeceros (y, seguro, alguna me dejaría); sin duda, os lo debo todo en la vida. ¡Os quiero con locura!

A mi **hermano y mi Eli** (mi cuñada) por vuestro apoyo y cariño y, por supuesto, por el "regalo" que mi hicisteis aquel 27 de noviembre de 2019: mi bebé, mi Oliver.

A **mi familia**: a mi tita Raquel, a mi tito Nono, a mi tito Fer, a mi tito Javi, a mi tita Nieves, a mi Tita Sara, a mi Nieves ("la niña de mis ojos"), a mi Ángela (siempre presente), a mi Sara, a mi Helena, a mi Luna, a mi Lucía y a mi Teo; por vuestro cariño y apoyo, por ser unos "segundos padres" y unos "segundos hermanos". ¡Os adoro!

Por supuesto, al artífice de todo esto: **Sergio Moreno Ríos**, mi director de tesis y mi amigo. Nuestra amistad podría decirse que surgió rápidamente; tan solo necesite 2 años y 245 cafés para comenzar a tutearte… Podría escribir 3512 "versiones de agradecimientos" (y sabes que sería capaz jejejeje), y en ninguna de ellas lograría expresar con suficiente claridad lo agradecida y afortunada que me siento de haberte tenido como director. Gracias por tu apoyo, tu paciencia, tu comprensión, tus consejos, por esos cafés/desayunos, por transmitirme ese amor por la investigación (recuerdo salir de tu despacho pensando: "¡Buah tengo el mejor trabajo del mundo!") y, por supuesto, por todo lo que me has enseñado (y me ensañarás). De corazón, ¡gracias!

A mis **amig@s** por esos momentos "que alargan la vida" tan necesarios en muchas ocasiones: a mis nenas (en especial a mi Brígida, mi "mami"), a mi mini-ONU (en especial a "ese corazón con piernas", mi Carmen Laguna), a mis Girasoles (en especial, como no, a mi "marió", mi Antonio), a los de Graná (en especial, a mi Evita) y a todas aquellas personas que han llegado (o han vuelto) en los últimos tiempos a mi vida y que me han hecho en muchos momentos recordar que "la vida puede ser maravillosa".

A mi “english supervisor”, **Hartmut Blank** por sus enseñanzas y esa calurosa acogida; a Beatriz (mi “spanish angel”), a Daphne y a Chris por hacerme sentir “como en casa”.

A **Paqui Serrano**, por su cariño, apoyo y consejos.

Por supuesto, a esos seres tan especiales que ha ocupado algunos de los, sin duda, mejores momentos de esta experiencia: **mi niñ@s** (mis participantes) por compartir conmigo “lo que piensan” y por el cariño recibido de ell@s. Y como no, a los **colegios** que han accedido a participar en mis investigaciones, en especial, al C.E.I.P Federico García Lorca de Tocón, allí empezó todo (en todos los sentidos).

Y a todas aquellas personas que me encontrado por el camino y de las que tanto he podido aprender; de corazón, ¡GRACIAS A TOD@S!

ÍNDICE DE CONTENIDOS

TABLE OF CONTENTS

SECCIÓN FINAL

'*': indicates that the chapter is written in English

Resumen

Overview

Gran parte de la investigación evolutiva se ha focalizado en la niñez temprana. Sin embargo, poco se conoce sobre qué ocurre en los años posteriores. En la presente tesis doctoral se investigan los cambios cognitivos en el desarrollo de dos aspectos centrales de la cognición humana: la atribución de conocimiento y la atribución de responsabilidad, en un período del ciclo vital especialmente inexplorado, la infancia media y tardía. Para ello, se analizan ambos fenómenos en escolares de entre 8 y 13 años y adultos.

La atribución de conocimiento se investiga a través de dos de los denominados sesgos egocéntricos: el sesgo de transparencia ilusoria en la intención (Keysar, 1994),y el sesgo retrospectivo (Fischhooff, 1975). El primero esta relacionado con la atribución de conocimiento de la intención comunicativa, mientras que el segundo se relaciona con el efecto que el conocimiento de un resultado tiene en la atribución de conocimiento a otras personas. En el presente trabajo se examinan ambos sesgos con el fin de conocer (o conocer más, en el caso del sesgo retrospectivo) si están presentes en la infancia, su curso evolutivo y los procesos psicológicos que subyacen a cada sesgo.

En el capítulo 2 se examina el sesgo de transparencia ilusoria en niñas/os haciendo uso de la tarea tradicional, la cual requiere de la comprensión del sarcasmo. Además, se ponen a prueba dos de las explicaciones ofrecidas al sesgo: la explicación pragmática (Gerrig y col., 2000) y la explicación de la *construal* (Keysar, 1994, 2000). Los resultados muestran que el sesgo esta presente en escolares y que la magnitud de este permanece estable durante estos años. En lo que respecta a las explicaciones ofrecidas al sesgo, los resultados sugieren que tanto principios pragmáticos como operaciones de *construal* estarían actuando en el la infancia.

El sesgo de transparencia ilusoria es de nuevo investigado en el Capítulo 3 pero en este caso en adultos. Basándonos en la literatura sobre *creeping determinism* (una de las manifestaciones o componentes del sesgo retrospectivo) evaluamos una explicación alternativa al sesgo de transparencia según la cual en el sesgo estarían actuando procesos de razonamiento causal. Entre otras, la literatura sobre *creeping determinism* establece que el sesgo sólo aparece cuando las descripciones de los eventos incluyen información causal relevante, de lo contrario, este desaparece (Yopnick y Kim, 2012). Los resultados muestran que el *creeping determinism* aparece cuando los escenarios de transparencia ilusoria incluyen información sobre un resultado (Experimento 1). Además, al igual que el *creeping determinsm*, la magnitud del sesgo de transparencia ilusoria disminuye cuando es difícil construir un modelo coherente sobre cómo ocurrieron las cosas (Experimento 2), así como cuando se debilita el nexo causal entre antecedente y

consecuente (Experimento 3). En cambio, cuando se fortalece el nexo causal, la magnitud del sesgo incrementa (Experimento 3).

El sesgo retrospectivo se investiga en el Capítulo 4. En dos experimentos, se examina el sesgo retrospectivo auditivo en niños/as. En el Experimento 1, se investiga a estas edades los escolares exhiben sesgo retrospectivo auditivo y, de ser así, si este es susceptible de ser eliminado. Además, en este experimento se explora si la magnitud y curso evolutivo del sesgo difieren en función del tipo de estímulos empleados para su evaluación (auditivos y visuales). En el Experimento 2 se evalúa si los procesos psicológicos que subyacen al sesgo retrospectivo auditivo en la infancia difieren en base al tipo de diseño experimental (hipotéticos o de memoria) utilizado para su evaluación. Los resultados muestran que las/os niñas/os exhiben sesgo retrospectivo auditivo y que no disminuye a pesar de favorecerles las condiciones para ello. Alrededor de los 9 años, el sesgo comienza a disminuir. Los resultados también sugieren que no hay diferencias en magnitud ni curso evolutivo relacionados con el tipo de estímulos utilizados -auditivos y visuales- (Experimento 1). Sin embargo, los resultados sugieren que diferentes procesos psicológicos subyacen al sesgo retrospectivo auditivo en los diseños hipotéticos y de memoria.

En lo que respecta a la atribución de responsabilidad, esta es explorada en el marco del Modelo Estructural de Atribución de Responsabilidad (Laganado y col., 2015), basado en la atribución de causalidad. La idea central del modelo es que los juicios de responsabilidad de las personas son producto de dos factores: criticidad (importancia percibida de la contribución de un agente para lograr un resultado conjunto) y pivotalidad (grado en el que el resultado depende de la acción del agente). Éste último factor, de acuerdo al modelo, depende exclusivamente del pensamiento contrafáctico.

En el Capítulo 5 se evalúa si el modelo puede explicar las atribuciones de responsabilidad causal que realizan niñas/os y adultos a un componente de un dispositivo electrónico. Además, se evalúa si existe una relación entre las habilidades de razonamiento contrafáctico de los/as niños/as y sus juicios de pivotalidad. Los resultados muestran que tanto niñas/os como adultos son sensibles a los valores de criticidad y pivotalidad y, lo que es más importante, que sus juicios de responsabilidad son producto de ambos factores. En lo que respecta al curso evolutivo, sólo los juicios de pivotalidad muestran cambios evolutivos. En concreto, los resultados muestran que las/os niñas/os más mayores son más sensibles al valor de pivotalidad que los pequeños. El mismo curso evolutivo se encuentra en la habilidad para razonar contrafácticamente. En consistencia

con la propuesta original, se constató la existencia de correlación positiva entre la sensibilidad al factor de pivotalidad de los escolares y su precisión en las inferencias semifácticas.

Por último, en el capítulo 6, se detallan los principales resultados encontrados en los diferentes experimentos, resaltando las similitudes y diferencias entre los diferentes fenómenos abordados, así como sus implicaciones teóricas.

The vast majority of developmental research has focused on early childhood. However, little is known about what happens latter in development. This doctoral dissertation investigates cognitive developmental changes in two central aspects of human cognition: knowledge attribution and responsibility attribution, in a particularly unexplored period in development, middle and late childhood. To this end, both phenomena are examined in children aged from 8 to 13 years old, and in adults.

Knowledge attribution is investigated thought two of the so-called egocentric biases: illusory transparency of intention bias (Keysar, 1994) and hindsight bias (Fischhoff, 1975). The former is related to communicative intent knowledge attribution, whereas the latter is linked to the effect that knowing an outcome has on people's knowledge attributions to others. In the present work both biases are examined in order to find out (or further know, in the case of hindsight bias) whether they are present in childhood, their developmental trend and the underlying psychological processes to each bias.

In Chapter 2, illusory transparency bias is examined in children using the traditional task, which requires sarcasm comprehension. Furthermore, two bias explanations are tested: the pragmatic account (Gerrig et al., 2000) and the construal account (Keysar, 1994, 2000). Results show that children exhibit illusory transparency bias and also, that the bias magnitude remains stable during this age range. With regards to the bias accounts, results suggest that both pragmatic principles and construal operations are active during childhood.

Illusory transparency bias is again investigated in Chapter 3 but, in this case, in adults. On the basis of creeping determinism (one of the hindsight bias manifestations or components) literature, an alternative explanation for illusory transparency bias involving the action of causal reasoning processes is assessed. Among others, creeping determinism literature establishes that the bias only appears when the event descriptions include relevant causal information; otherwise, it disappears (Yopnick y Kim, 2012). Results show that creeping determinism appears when illusory transparency scenarios include outcome information (Experiment 1). Furthermore, like creeping determinism, the magnitude of illusory transparency bias diminishes when participants' readiness to set up a coherent model on how things came about is hindered (Experiment 2), as well as when the causal link between antecedent and consequent is weakened (Experiment 3). However, when the causal link is strengthened, the bias magnitude increases (Experiment 3).

Hindsight bias is investigated in Chapter 4. Through two experiments, auditory hindsight bias is examined in children. In Experiment 1, it is investigated whether they show auditory hindsight bias at this age and, if so, it the bias can be avoided. This experiment also explores whether hindsight bias magnitude and development trend differ depending on the kind of stimuli used to assess them (auditory and visual). In Experiment 2, it is assessed whether different psychological processes underlie auditory hindsight bias in each experimental design (hypothetical and memory designs) in childhood. Results show that children exhibit auditory hindsight bias and also that the bias persists in spite of providing them with opportunities to avoid it. Around the age of 9, the bias magnitude starts to decline. Results also suggest that there are not differences in the bias magnitude or its developmental trend related to the sort of stimuli used – auditory and visual (Experiment 1). However, they suggest that different processes underlie auditory hindsight bias in hypothetical and memory designs.

With respect to responsibility attribution, it is explored within the framework of the Structural Model of Responsibility Attribution (Lagando et al., 2015), which is based on causality attribution. The core idea of this model is that people's responsibility judgements are the product of two factors: criticality (the perceived importance of an agent's contribution to achieve a joint outcome) and pivotality (the extent to which the outcome depends on the agent's action). According to the model, this latter factor depends solely on counterfactual thinking.

In Chapter 5, it is assessed whether the model can explain children and adults' causal responsibility attributions to a component of a mechanical device. Furthermore, it is assessed whether there is a relationship between children's counterfactual reasoning abilities and their pivotality judgements. Results show that both children and adults are sensitive to criticality and pivotality values and, what is more important, that their responsibility judgements are the product of both factors. With regards to the developmental trend, only pivotality judgements change with age. In particular, results show that older children are more sensitive to the pivotality value than younger ones. The same developmental trend is found in children's ability to reason counterfactually. In accordance with the original idea, a positive correlation was found between children's sensitivity to pivotality and their accuracy in semifactual inferences.

Finally, in Chapter 6, the main results obtained in the different experiments are described, by highlighting the similarities and differences found among the different phenomena addressed, as well as their theoretical implications.

Capítulo 1

Introducción

1. EGOCENTRISMO

"Al igual que el aire que respiramos está contaminado por gases industriales, nuestra mente está contaminada por nuestro conocimiento…" (Wilson y Brekke, 1994)

Bien porque no somos consciente de ello, o bien por que no podemos evitarlo, lo cierto es que muchas de las emociones, comportamientos y/o juicios que emitimos en nuestra vida diaria están influenciados por nuestro conocimiento. Como si de unas gafas se tratase, nuestro conocimiento nos otorga una forma particular de percibir e interpretar el mundo que nos rodea y los eventos que en él acontecen. Esta visión propia o egocéntrica de la realidad es especialmente evidente durante los primeros años de vida. Tal y como señaló Piaget (1958), los/as niños/as interpretan la realidad a partir de su propia visión de ésta y no conciben que otras personas puedan interpretar o percibir esa misma realidad de modo diferente. Piaget mostró esta limitación en la toma de perspectiva de los niños a través de la tarea de las tres montañas. En ella, se mostraba a los/as niños/as una maqueta con tres montañas y se les pedía que identificasen la perspectiva visual de la escena que tendría un muñeco que había sido colocado en uno de los laterales de la maqueta. Para ello, se les mostraba un conjunto de fotografías de la maqueta tomadas desde diferentes ángulos, entre las que se encontraba una fotografía tomada desde la posición de la niña (frente a la maqueta) y otra que coincidía con la percepción visual de la escena que tendría el muñeco (desde uno de los laterales). De manera sistemática, los preescolares confundían su propia perspectiva visual de la escena (frente a la maqueta) con la perspectiva visual del muñeco, mostrando así su imposibilidad para apreciar otras perspectivas de la realidad.

Otro ejemplo del egocentrismo exhibido por los/as niños/as podemos encontrarlo en los estudios sobre Teoría de la Mente (TOM). Al igual que la tarea de las tres montañas de Piaget, estos estudios muestran la incapacidad de los/as niños/as para comprender que otras personas pueden tener una representación de la realidad diferente a la suya. Por ejemplo, en una de las múltiples tareas empleadas para el estudio de la TOM, se les presenta a los/as niños/as una caja de caramelos. El experimentador abre la caja frente a la niña y le muestra el contenido de la caja. Esta caja, en lugar de caramelos, contiene rotuladores. A continuación, el experimentador cierra la caja frente a la niña y le pide que anticipe la respuesta de un/a compañero/a al que sólo se le muestra la parte externa de la caja ("¿Qué pensará tu compañero/a que contiene la caja?"). Hasta los 4 años, los/as

niños/as no comprenden que la representación de la realidad de su compañero/a ("una caja de caramelos") sería distinta a la representación que ellos/as poseen de esa misma realidad ("una caja de caramelos que contiene rotuladores"), lo cual les lleva a afirmar que su compañero/a pensará que la caja contiene rotuladores.

De acuerdo con Piaget, esta visión egocéntrica de la realidad es propia del estadio preoperacional (entre los 2 y los 7 años) y desaparecerá en torno a los 7 años, con la llegada de las operaciones concretas (Piaget, 1958). Sin embargo, aunque con el desarrollo los/as niños/as van adquiriendo la capacidad de comprender que otras personas pueden representar la realidad de un modo diferente al suyo (p.ej., Baron-Cohen y col., 1985; Gopnick y Slaughter, 1991; Laggatutta y col., 2014; Lagatutta y col., 2010), lo cierto es los seres humanos nunca dejamos de ser egocéntricos (p. ej., Bernstein y col., 2011; Bayen y col., 2007; Pohl y col., 2018). De hecho, niñas/os y adultos compartirían esa tendencia inicial a interpretar la realidad de modo egocéntrico, difiriendo únicamente en su capacidad para corregir esa primera, y preponderante, respuesta egocéntrica (Epley y col., 2004b). Así, el egocentrismo mostrado por los adultos sería el mismo que el exhibido por los/as niños/as, aunque de modo atenuado.

1.1.Egocentrismo en adultos: el fenómeno de la contaminación mental

En adultos, cientos de investigaciones han mostrado como los juicios, emociones y/o comportamientos de los individuos están influenciados, de modo inconsciente o involuntario, por su conocimiento (p. ej., Arkes, 2013; Arkes y col., 1981; Dawson y col., 1988; Harley, 2007; Kamin y Rachlinski, 1995; Hawkins y Hastie, 1990; Mante-Estacio y Bernardo, 2014; Savitsky y col., 2011). A lo largo de la historia y dependiendo de la disciplina esta influencia del propio conocimiento ha recibido diferentes etiquetas tales como "curso de conocimiento" (Camerer y cols., 1989), "sesgo de realidad" (Mitchell y Taylor, 1999), "egocentrismo adulto" (p.ej., Kelley y Jacoby, 1996), "sesgo retrostpectivo" (p.ej., Fischhoff, 1975), "la maldición de la experiencia" (p. ej., Hinds, 1999), entre otros.

En 1994 Wilson y Brekke acuñaron el término "*contaminación mental*" para referirse al "*proceso a través del cual una persona tiene un juicio, emoción o comportamiento indeseado porque su procesamiento mental es inconsciente y/o incontrolado"* (Wilson y Brekke, 1994, p. 117). Como estos autores señalan en la propia definición del fenómeno, la contaminación mental es indeseada: la persona preferiría evitarla; sin embargo, bien porque no es consciente de ella o bien porque no puede

evitarla, esta contaminación acontece (Wilson y Brekke, 1994; Bernstein y col., 2012; Harley y col., 2004).

El fenómeno de la contaminación mental ha sido ampliamente estudiado por la Psicología de la Cognición Social a través del estudio de los denominados sesgos cognitivos. En uno de estos estudios, Bernstein, y colaboradores (2012) presentaron a los/as participantes un conjunto de palabras distorsionadas. En la mitad de los ensayos, antes de escuchar la palabra distorsionada, los/as participantes escuchaban una versión clara (no distorsionada) de la palabra. En cada ensayo, se les pidió que estimaran el número de personas (de un grupo formado por 100) que identificarían la palabra únicamente escuchando la versión distorsionada de ésta. Los/as participantes afirmaron que un mayor número de personas identificaría la palabra distorsionada en aquellos ensayos en los que ellos habían sido informados previamente de la identidad de la misma. Siguiendo este mismo procedimiento, estos autores evaluaron si los/as participantes podrían evitar esta contaminación o influencia del propio conocimiento. Para ello, instruyeron a los/as participantes sobre el tipo de error que habitualmente cometen las personas en estas tareas (sobreestimar el número de personas que identificarían la palabra cuando reciben información sobre su identidad) y les pidieron que tratasen de evitarlo. Sin embargo, de nuevo los/as participantes proporcionaron estimaciones más altas en los ensayos en los que recibían información sobre la identidad de la palabra que en los que únicamente escuchaban la versión distorsionada de ésta. Así, tal y como Wilson y Brekke (1994) señalaron, la contaminación mental parece ocurrir de manera inconsciente y/o incontrolada.

1.2.Egocentrismo Epistémico

Las tareas empleadas por la Psicología de la Cognición Social para el estudio de los sesgos y las tareas clásicas de TOM guardan ciertas similitudes. En ambos casos, existe un conocimiento que ha de ser ignorado (p.ej. el conocimiento sobre la identidad de la palabra y el contenido real de la caja) a fin de emitir un juicio "correcto" (no sesgado) sobre la perspectiva de otra u otras personas (Royzman y col., 2003; Birch y Bernstein, 2007; Ghrear y col., 2016). En base a las similitudes entre ambas tareas, y en un intento por sistematizar el fenómeno de la contaminación mental en el estudio de la toma de perspectiva, Royzman y colaboradores (2003) acuñaron el término "*Egocentrismo Epistémico*" (EE) para denominar a todos los sesgos (o errores) en la toma de perspectiva causados por "contaminar" con el propio conocimiento los juicios que se emiten sobre la perspectiva de otras personas. En concreto, el EE engloba a todos aquellos casos en los

que los individuos, pese a ser conocedores de que las otras personas no tienen acceso a una determinada pieza de conocimiento (p. ej., identidad de la palabra distorsionada o contenido real de la caja), no pueden evitar que dicho conocimiento influencie los juicios que emiten sobre la perspectiva de esas personas. Así, el EE incluye a sesgos (o errores) en la toma de perspectiva como, por ejemplo, el sesgo retrospectivo (Fischooff, 1975), el sesgo de transparencia ilusoria en la intención (Keysar, 1994) o el sesgo de realidad (Mitchell y col., 1996); así como a todos los errores en la toma de perspectiva documentados por la Psicología Evolutiva en su estudio de la TOM.

Epley y colaboradores (2004a) propusieron que todos los sesgos en la toma de perspectiva englobados dentro del EE son producto de un mecanismo común: el uso del Heurístico de Anclaje y Ajuste (Tversky y Kahneman, 1974). De acuerdo con estos autores, cuando los individuos han de adoptar la perspectiva de otra persona se basan en sus propias creencias con el fin de establecer un punto de referencia o "ancla" sobre el cual posteriormente realizar los mínimos ajustes necesarios para obtener una predicción precisa de la perspectiva de otra persona (Nickerson, 1999; Epley y col., 2006; Epley, 2014; Epley y col., 2004a). Sin embargo, el "ajuste" que se realiza es insuficiente, lo cual conduce a predicciones erróneas, influenciadas por el propio conocimiento, de la perspectiva de otra persona. Pese a que la propuesta de Epley y colaboradores (2004a) no ha sido contrastada empíricamente, algunas investigaciones han tratado a los sesgos incluidos en el EE como fenómenos intercambiables a partir los cuales investigar la toma de perspectiva (p.ej., Todd y col., 2011).

El EE cuenta con un paradigma experimental propio. Las tareas empleadas para su estudio requieren dotar a un individuo de una información de tipo privilegiado, para que, posteriormente, éste anticipe el juicio, emoción o respuesta comportamental de otro sujeto que carece de dicha información (Royzman y col., 2003). En todos los casos, los/as participantes son informados directa o indirectamente de que la persona a juicio no tiene acceso a la información privilegiada que ellos poseen. Pese a ello, los/as participantes le atribuyen ese conocimiento, dando esto lugar a juicios sesgados sobre la perspectiva de la otra persona (p. ej., Bernstein y Harley, 2007; Epley y col., 2004a; Mante-Estacio y Bernardo, 2014; Todd y col., 2011;).

Aunque el EE epistémico ha sido ampliamente estudiado en adultos, lo cierto es que son pocas las investigaciones que han abordado el fenómeno del EE en niñas/os (Royzman y col., 2003), y los resultados que de ellas se desprenden son contradictorios. Mientras algunos resultados parecen indicar una similitud en estos sesgos egocéntricos

entre niñas/os y adultos (Tversky y Kahneman, 1983; Mossler y col., 1976; Harris y col., 1989; Epley y col., 2004b), otros apuntan a una mayor proclividad en el caso de los/as niños/as para sucumbir a este tipo de sesgos (Bernstein y col., 2004, 2011; Pohl y col., 2018). Incluso hay un estudio que sugiere los/as niños/as podrían ser menos propensos que los adultos a caer en este tipo de errores (Mitchell y col., 1996). El principal argumento esgrimido para explicar esta disparidad de resultados ha sido la escasa sensibilidad de las pruebas empleadas por la Psicología del Desarrollo para el estudio de la toma de perspectiva en niñas/os. En los estudios con niñas/os, a diferencia de en las investigaciones con adultos, se han empleado variables de tipo cualitativo o dicotómico -supera o no supera- (Ghrear y col., 2016; Royzman y col., 2003). Esto ha hecho imposible establecer diferencias evolutivas en el desarrollo de la toma de perspectiva, así como comparar la ejecución de niñas/os y adultos en este tipo de tareas. En este sentido, algunos/as autores/as (Ghrear y col., 2016; Royzman y col., 2003), han señalado la necesidad de desarrollar nuevas tareas para el estudio de la toma de perspectiva en general, y de los sesgos egocéntricos en particular, que puedan ser aplicadas tanto a niñas/os como a adultos. De acuerdo con Ghrear y colaboradores (2016), estas tareas, a diferencia de las tareas clásicas de TOM, han de incluir medidas cuantitativas de la ejecución, así como suprimir las demandas lingüísticas que tradicionalmente han acompañado a dichas tareas. En la presente tesis doctoral, pondremos a prueba la hipótesis del mecanismo común propuesta por Epley y colaboradores (2004), investigando en niñas/os dos de los sesgos incluidos en el EE: el sesgo de transparencia ilusoria en la intención (Keysar, 1994) y el sesgo retrospectivo (Fischhoff, 1975). Recientemente, Turner y Schley (2016) han propuesto considerar los efectos de anclaje y ajuste en términos causales, es decir, como un antecedente (anclaje) y consecuente (ajuste). Por ello, en este trabajo también exploraremos la relación de éstos con el razonamiento causal. Para ello, desarrollaremos nuevas medidas que nos permitan superar las limitaciones anteriormente mencionadas de las tareas empleadas para el estudio de la toma de perspectiva en niñas/os.

2. EL SESGO DE TRANSPARENCIA ILUSORIA EN LA INTENCIÓN

¡Maravilloso, realmente maravilloso!

Tradicionalmente, el sesgo de *transparencia ilusoria en la intención* (Keysar, 1994) ha sido evaluado haciendo uso de relatos en los que un mensaje ambiguo (como,

por ejemplo, el que inicia este capítulo) puede ser interpretado como sincero o como sarcástico en función de la información de la que disponga el receptor del mensaje. En uno de estos relatos, David pide a su compañera de trabajo, Julia, que le recomiende un lugar donde ir a cenar esa noche. Julia recomienda a David un restaurante y éste acude a cenar esa noche. Al día siguiente, David se dirige al despacho de Julia para contarle cómo fue su experiencia en el restaurante; sin embargo, Julia no se encuentra en ese momento en su despacho, así que David decide dejarle una nota con el siguiente mensaje: "Querías saber cómo me fue en el restaurante ¡Fue maravilloso, realmente maravilloso!" (Keysar, 1994, p. 173). En una de las condiciones, la experiencia de David en el restaurante se describe como positiva ("la comida era deliciosa y el servicio impecable), mientras que en otra condición ésta es descrita como negativa ("la comida era horrible y el servicio mediocre"). A los/as participantes se les pregunta si David tiene la intención de ser sarcástico y si Julia, la destinataria, interpretará el mensaje de David como sarcástico. En ambas condiciones, Julia desconoce cómo fue la experiencia de David en el restaurante; por tanto, carece de la información necesaria para interpretar el mensaje como sarcástico cuando su experiencia es negativa. Así, en ambas condiciones, los/as participantes habrían de predecir una respuesta sincera por parte de Julia; sin embargo, cuando la experiencia es negativa algunos/as participantes señalan que Julia interpretará el mensaje como sarcástico, mostrando así el sesgo de *transparencia ilusoria en la intención* (p. ej. Keysar, 1994; Moreno-Ríos y col., 2011; Weingartner & Klin, 2005, 2009). La diferencia de puntuaciones (sarcasmo atribuido a la interpretación del receptor) entre la condición positiva y la negativa es la *magnitud del sesgo de transparencia ilusoria en la intención* (p. ej., Epley y col., 2004a; Moreno-Ríos y col., 2011).

En lo que respecta la intención del emisor del mensaje, la mayoría de los/as participantes coinciden en atribuir una intención sincera a David cuando la experiencia es positiva, y una intención sarcástica cuando la experiencia es negativa (p. ej. Epley y col., 2004a; Keysar, 1994; Moreno-Ríos y col., 2011). No obstante, cabe señalar que, en la condición negativa, la atribución de sarcasmo a la intención del emisor es mayor que el sarcasmo atribuido a la interpretación del receptor. O dicho de otro modo, son más los/as participantes que afirman que David tiene la intención de ser sarcástico que aquellos que predicen una interpretación sarcástica por parte de Julia.

2.1. Explicaciones al sesgo

Además de la explicación del Heurístico de Anclaje y Ajuste vista en la sección anterior, el sesgo de transparencia ilusoria ha recibido otras explicaciones. En el trabajo

original, Keysar (1994) defendió que el error se produce porque los lectores asumen que las intenciones comunicativas del emisor del mensaje (p. ej., David) son "transparentes" para el receptor (p. ej. Julia); de ahí el nombre del sesgo, *transparencia ilusoria en la intención*. En concreto, el sesgo se produciría porque los lectores utilizan su conocimiento privilegiado para detectar la intención sarcástica en el mensaje, y asumen que esta intención (la cual perciben gracias a su conocimiento privilegiado) es transparente para el receptor del mensaje: el receptor puede "ver" la intención en el mensaje (aunque desconozca la experiencia) y, por tanto, interpretar éste como sarcástico.

Además de la "*transparencia en la intención*", Keysar (1994) ofrece una explicación alternativa al sesgo (la explicación de la construal), la cual señala como su explicación preferida en trabajos posteriores (Keysar, 2000). En ella, defiende que, al igual que ocurre cuando se interpretan otros estímulos ambiguos (p. ej., Griffin y col., 1990; Keysar y Bly, 1995), cuando los individuos interpretan un mensaje ambiguo, éstos "reconstruyen" dicho mensaje de manera que éste es percibido como menos ambiguo de lo que realmente es. En concreto, el conocimiento sobre la experiencia permitiría a los /as participantes "desambiguar" el mensaje ambiguo, transformándolo en un mensaje claramente sarcástico. Así, el sesgo se produciría porque los/as participantes perciben el mensaje como un mensaje claramente sarcástico que ya informa *per se* de que la experiencia fue negativa.

El fenómeno de la transparencia ilusoria en la intención (Keysar, 1994) no ha estado exento de críticas, muy probablemente porque cuestiona una de las principales premisas de la Teoría Estándar del Uso del Lenguaje: los hablantes se basan en el conocimiento compartido entre emisor y receptor cuando éstos han de adoptar la perspectiva lingüística de otra persona, es decir, cuando han de inferir cómo otra persona interpretará un determinado mensaje (p. ej. Clark y Marshall, 1981; Greene y col., 1981). Gerrig y colaboradores (2010) ofrecieron una explicación al fenómeno desde la Teoría Estándar del Lenguaje según la cual el sesgo no sería producto de un conocimiento privilegiado, sino del deseo de los lectores de mantener los principios de cooperación comunicativa y racionalidad en los relatos de transparencia ilusoria. Más específicamente, estos autores defienden que el sesgo es producto de la resistencia de los lectores a aceptar que el emisor del mensaje ha violado los principios de cooperación en la comunicación (Grice, 1975); y en concreto, su máxima de calidad: "no decir aquello que es falso". Esta resistencia conduciría a los/as participantes asumir que el receptor del mensaje ha de

disponer de información adicional a la expresada en el relato, la cual le permitirá detectar la intención sarcástica del emisor del mensaje, e interpretar éste como sarcástico.

Pese a ser un fenómeno robusto y ampliamente estudiado (p. ej., Epley y col., 2004a; Gerrig y col., 2000; Keysar, 1994; Moreno-Ríos y col., 2011; Weingartner y Klin, 2005, 2009) aún se desconocen los mecanismos que subyacen al sesgo de transparencia ilusoria. Estudios previos han demostrado que el sesgo desaparece cuando se les pide a los/as participantes que pronostiquen cómo un lector accidental (p. ej. La secretaria de Julia) interpretaría el mensaje ambiguo (Moreno-Ríos y col., 2011). Lo mismo ocurre cuando se les proporciona a los/as participantes una razón por la que el emisor del mensaje deja un mensaje positivo tras una experiencia negativa (p. ej., "David no quiere herir los sentimientos de Julia"). Estos resultados sugieren que en el sesgo de transparencia ilusoria podrían estar actuando principios pragmáticos; sin embargo, resultan insuficientes para explicar el fenómeno de la transparencia ilusoria en su totalidad.

En niñas/os, el sesgo de transparencia ilusoria no ha sido evaluado, muy probablemente, porque la tarea tradicional requiere que los/as menores sean capaces de comprender el sarcasmo, es decir, que sean capaces de detectar la incongruencia entre lo que se dice ("todo fue maravilloso") y lo que realmente se quiere decir – "todo fue horrible" – (p. ej., Capelli y col., 1990; Ivanko y Pexman, 2003; Nilsen y col., 2011). Dicha habilidad aparece alrededor de los 5 o 6 años y continúa desarrollándose a lo largo de la infancia (p. ej., Burnett, 2015; Glenwright y Pexman, 2010; Harris y Pexman, 2003). Alrededor de los 8 años, los/as niños/as no sólo son capaces de comprender el significado "real" de un mensaje positivo tras una experiencia negativa, sino que además son capaces de ajustar sus predicciones sobre la interpretación que otra persona realizará de un determinado mensaje al conocimiento que esta posee (Nilsen y col., 2011). En la presente tesis doctoral evaluaremos si los/as niños/as de entre 8 y 13 años muestran el sesgo de transparencia ilusoria en la intención e investigaremos los que mecanismos que subyacen al mismo en niñas/os y adultos.

3. EL SESGO RETROSPECTIVO

¡Lo sabía!

Una experiencia común es pensar "¡lo sabía!" tras conocer la respuesta correcta a una de las preguntas del Trivial Pursuit, o tras comprobar que, un año más, uno/a no es el afortunado/a ganador/a del Gordo de Navidad. En estos casos la persona ha sido presa de

el sesgo retrospectivo (Fischooff, 1975). En 1975, Fischhoff informó de un fascinante fenómeno: tras ser informados de un resultado, los individuos tienden percibir este como más probable y/o predecible. En su estudio, Fischhoff (1975; Experimento 1) pidió a los/as participantes que leyeran un relato en el que se describía el conflicto armado que tuvo lugar a principios del siglo XIX entre británicos y un grupo de guerrilleros indios llamados Gurkhas. Tras esto, los/as participantes habían de estimar las probabilidades de ocurrencia de cuatro posibles resultados: victoria británica, victoria de los Gurkhas, estancamiento militar sin acuerdo de paz y estancamiento militar con acuerdo de paz. Los escenarios incluían antecedentes tanto a favor de la victoria británica (p. ej., el ejército Gurkhas contaba con tan sólo 12.000 soldados) como a favor de la victoria de los Gurkhas (p. ej., los soldados británicos no estaban acostumbrados a luchar en terrenos montañosos). En la mitad de los escenarios se incluyó una frase final en la que se informaba del resultado del conflicto (estancamiento militar sin acuerdo de paz). Los/as participantes a los que se les informó del desenlace del conflicto, juzgaron éste como más probable que aquellos que no habían recibido esta información.

Desde el estudio Fischhoff (1975), cientos de investigaciones han abordado el fenómeno del sesgo retrospectivo mostrando que se trata de un fenómeno universal (Pohl y col., 2002) que puede ser documentado en una gran variedad de situaciones como, por ejemplo, decisiones judiciales (p. ej., Harley, 2007), diagnósticos médicos (Arkes, 2013), eventos deportivos (Learly, 1981) o elecciones electorales (Blank y col., 2003); además de con un amplio abanico de estímulos, tales como, estímulos visuales (Bernstein y col., 2007; Harley y col., 2004), auditivos (Bernstein y col, 2012), verbales (p.ej. Nestler y col., 2008a; Roese y Olsen, 1996) o gustativos (Pohl y col., 2003).

3.1.Diseños para la evaluación del sesgo retrospectivo

La evaluación del sesgo retrospectivo se ha llevado a cabo, principalmente, a través de dos diseños experimentales: los diseños de memoria y los diseños hipotéticos (Pohl, 2007). En los diseños de memoria, se les pide a los/as participantes que respondan a una serie de preguntas de conocimiento general como, por ejemplo, cuántos dientes tiene un mosquito (juicios originales). A continuación, se les informa de la respuesta correcta a todas (o algunas) de las preguntas (p. ej., "los mosquitos tienen 42 dientes"). Tras esto, se les pide que recuerden la respuesta que proporcionaron anteriormente a cada una de las preguntas (juicios de recuerdo). Habitualmente, el recuerdo de los/as participantes de sus juicios originales se muestra sesgado por el conocimiento recibido (e.g., Bernstein y col., 2011; Pohl y col., 2010; Wood, 1978), así, por ejemplo, un/a participante cuya

respuesta original fue "los mosquitos tienen 24 dientes" en la fase de recuerdo afirmaría que su estimación original fue de 32 dientes (una estimación más próxima a la respuesta correcta, 42 dientes). En estos diseños, la magnitud del sesgo retrospectivo la representa la diferencia entre las estimaciones proporcionadas en la fase inicial (juicios iniciales -sin información-) y las proporcionadas en la fase final (juicios de recuerdo -con información-). La magnitud del sesgo incrementa a medida que incrementa el intervalo de retención, es decir, a medida que incrementa el tiempo que transcurre entre la fase inicial y la fase de recuerdo.

En cambio, en los diseños hipotéticos los/as participantes reciben en primer lugar información sobre, por ejemplo, el resultado de un evento -p. ej., "Los dos bandos acordaron un estancamiento militar, pero fueron incapaces de llegar a un acuerdo de paz"- (Fischhoff, 1975; Experimento 2) o la respuesta correcta a una pregunta (p. ej., "los mosquitos tienen 42 dientes"). Tras esto, se les pide que ignoren la información recibida e informen sobre cuál habría sido su respuesta de no haber recibido dicha información (p. ej., "Si no te hubiese dicho que los mosquitos tienen 42 dientes, ¿cuál habría sido tu respuesta?"). En una condición de control, los/as participantes responden a esas mismas preguntas, pero sin recibir información sobre la respuesta correcta. De nuevo aquí, los/as participantes que reciben información sobre el resultado proporcionan estimaciones más próximas a éste (o juzgan éste como más probable) que aquellos/as que no reciben información (p. ej., Fischhoff, 1975; Experimento 2). Algunos estudios han utilizado una versión "social" de los diseños hipotéticos en la que se les pide a los/as participantes que estimen el número de personas que, sin ser informados del resultado, lo conocerían (p.ej., Fischhoff, 1975, Experimento 3; Higham y col., 2017; Bernstein y col., 2018). En ambas versiones (tradicional y social), el tamaño del error de los participantes (magnitud del sesgo) es muy similar (p., Fischhoff, 1975). En cambio, la magnitud del sesgo retrospectivo sí difiere en función de si es evaluado a través de un diseño hipotético o un diseño de memoria.

Se ha visto que la magnitud del sesgo es mayor en los diseños hipotéticos que en los de memoria (p. ej., Campbell y Tesser, 1983; Fischhof, 1977; Higham y col., 2017; Wood, 1978). Higham y colaboradores (2017) argumentaron que estas diferencias podrían deberse a que los juicios requeridos en los diseños hipotéticos son más susceptibles al sesgo que los requeridos por los diseños de memoria. En los diseños de memoria, a diferencia de en los hipotéticos, los/as participantes tienen "algo que recordar"; por tanto, puede ocurrir que algunos/as participantes recuerden correctamente todos (o algunos) sus

juicios originales, es decir, que no muestren sesgo (o que lo muestren sólo en determinados ítems). En cambio, en los diseños hipotéticos, todos los juicios son susceptibles de mostrase sesgados dado que no hay estimación previa que recordar. Para superar esta limitación, algunos autores proponen que en los diseños de memoria se eliminen de la muestra a aquellos ítems (o participantes) en los que se produce un recuerdo correcto del juicio original para así obtener una medida "real", libre de artefactos, del sesgo comparable con el obtenido en los diseños hipotéticos (Erdfleder y Buchner, 1998; Pohl, 2007).

3.2.Manifestaciones del sesgo retrospectivo

Hasta la primera década del siglo XXI, el sesgo retrospectivo fue tratado como un fenómeno unitario, gobernado por un mismo mecanismo y con un "único efecto" o manifestación. Sin embargo, en la actualidad se asume que pueden distinguirse, al menos, tres efectos o manifestaciones diferentes del conocimiento de un resultado ("visión de los componentes separados"; Blank y col., 2008; Nestler y col., 2010); cada una de ellas gobernada por un proceso psicológico distinto (p. ej., Blank y Nestler, 2006; Nestler y col., 2010): 1) impresión de previsibilidad, 2) impresión de inevitabilidad y 3) distorsiones de memoria.

En el componente de distorsiones de memoria, el conocimiento de un resultado sesga el recuerdo del juicio original en dirección al resultado informado. Este componente es producto de procesos de memoria, más específicamente, de cómo la información sobre el juicio original (juicio sin información) es recuperada y reconstruida tras ser informados/as de un resultado (véase para más detalles Blank y Nestler, 2007; Pohl y col., 2003; Hoffrage y col., 2000).

Por su parte, el componente de impresión de inevitabilidad incluye los casos en los que el conocimiento de un resultado hace que éste sea percibido como más probable, inevitable o causalmente determinado (p. ej. Nario y Branscombe, 1995; Nestler y von Collani, 2008; Nestler y col., 2008a). Este componente ha sido evaluado principalmente a través de escenarios como los utilizados por Fischhoff (1975) y se ha visto que depende de procesos de razonamiento causal (p. ej., Nestler y col., 2010; Nestler y col., 2008a; Nestler y col. 2008b; Nestler y von Collani, 2008). La *Causal Model Theory for creeping determinism* – CMT- (Blank y Nestler, 2007) es la principal teoría utilizada para explicar los hallazgos de *creeping determinism*. De acuerdo a la CMT, cuando los individuos son informados de un resultado, éstos se involucran en un proceso de razonamiento causal con el fin de dar sentido al resultado informado. Este proceso de razonamiento causal

implicaría tanto la búsqueda de antecedentes que potencialmente puedan explicar el resultado, como su evaluación en términos de "cómo de bien" los antecedentes identificados pueden explicar el resultado. El *creeping determinism* aparece si tiene éxito el proceso de búsqueda de antecedentes, así como su evaluación e integración en un modelo causal. Así, el sesgo se produce al construirse un modelo (sesgado) que explica el resultado (Blank y Nestler, 2007). La implicación de procesos de razonamiento causal en el *creeping determinsm* es tal que tanto la emergencia del sesgo (Yopchick y Kim, 2012) como la magnitud del mismo dependen de la cantidad de información de naturaleza causal disponible (Nestler y col., 2008a; Oeberst y col., 2014; Wasserman y col., 1991). Así, cuanto mejor puedan explicar los antecedentes disponibles el resultado, mayor es el *creeping determinism* (p. e.j., Nario y Branscombe, 1995; Nestler y col., 2008a; Nestler y von Collani, 2008b; Roese y Olson, 1996; Oeberst y col., 2014; Yopnick y Kim, 2012). En cambio, cuando la información disponible no permite (o dificulta) crear un modelo causal coherente sobre el evento, el sesgo disminuye (p.ej., Ash, 2009; Calvillo y Gomes, 2011; Pezzo, 2003; Schkade y Kilbourne, 1991) o incluso desaparece (Yopchick y Kim, 2012; Wasserman et al., 1991). De manera similar, el sesgo incrementa cuando se establecen relaciones de tipo contrafáctico ("Si no hubiese ocurrido p, no habría ocurrido q") entre antecedentes y resultado (Nestler y von Collani, 2008a; Roese y Olson, 1996), mientras que disminuye cuando la relación establecida entre ambos es semifáctica -aunque hubiera ocurrido p, habría ocurrido q- (Roese y Olson, 1996).

Por último, en el componente de "impresión de previsibilidad", el conocimiento de un resultado genera en los/as participantes la creencia de que siempre conocieron el resultado y/o que habrían sido capaces de predecirlo de no haber sido informados de éste (Fischhoff, 1977). Este componente se ha relacionado con aspectos metacognitivos (p. ej., Nestler y col., 2010) y más específicamente, con procesos de fluencia - familiaridad con el estímulo- (Bernstein y col., 2018; Bernstein y Harley, 2007; Harley y col., 2004; Higham y col., 2017). Así, las impresiones de previsibilidad ocurrirían porque el conocimiento de un resultado como, por ejemplo, la identidad de un estímulo auditivo distorsionado, facilitaría el procesamiento de dicho estímulo. Esta facilidad de procesamiento, en lugar de ser atribuida al conocimiento (resultado) recibido, sería erróneamente atribuida a las características del propio estímulo ("el estímulo es fácil de identificar" – *fluency missatribution*-), generando en los/as participantes la creencia de que otras personas podrían identificarlo pese a no contar con su mismo conocimiento ("el estímulo es fácil de identificar").

3.3. El sesgo retrospectivo en niños/as

Son pocos los estudios que han explorado el sesgo retrospectivo en niñas/os. Para su evaluación se ha empleado tanto la versión social de los diseños hipotéticos como diseños de memoria. En los diseños de memoria, el procedimiento ha sido el mismo que el empleado en investigaciones con adultos (p. ej., Pohl y col., 2010). En cambio, en la versión social de los hipotéticos se ha incluido una leve variación en el procedimiento (Bernstein y col., 2004, 2007, 2011). Por ejemplo, Bernstein y colaboradores (2004) pidieron a los/as niños/as que identificarán, tan pronto como pudiesen, una serie de objetos que se presentaban completamente difuminados (p. ej., un avión) y progresivamente se iban clarificando. A continuación, se les presentaba a una marioneta llamada Ernie y se les decía que Ernie era un niño de su misma edad que no había visto esos objetos antes. La tarea de los/as niños/as consistió en indicar en qué momento de la secuencia de clarificación Ernie descubriría el objeto. De manera sistemática los/as niños/as afirmaron que Ernie sería capaz de identificar los objetos en un estado más difuminado del que ellos/as lo habían identificado previamente.

El curso evolutivo del sesgo parece ser el mismo en los diseños hipotéticos y en los de memoria. Resultados de investigaciones previas muestran que, en ambos casos, el sesgo esta presente, al menos, desde los 3 años de edad (Bernstein y col., 2004, 2007, 2011). La magnitud del sesgo retrospectivo disminuye entre los 3 y los 5 años (Bernstein et al., 2004, 2007, 2011) para luego permanecer estable hasta llegar a la senectud donde, de nuevo, incrementa (Bernstein y col., 2011). Es decir, a partir de los 6 años de edad los/as niños/as mostrarían el mismo sesgo que los adultos mayores (Bernstein y col., Pohl, Martin y Bayen, 2010). No obstante, estos resultados no coinciden con los encontrados en un estudio reciente por Pohl y colaboradores (2018) encontraron evidencias de una disminución del sesgo entre los 9 y 12 años de edad en los diseños de memoria.

En lo que respecta a las tareas utilizadas para evaluar el sesgo en niñas/os, los materiales que estas han empleado han sido exclusivamente visuales en el caso de los diseños hipotéticos (y verbales, en el caso de los diseños de memoria); por tanto, se desconoce si los/as niños/as mostrarían el sesgo si se emplean estímulos de otras modalidades perceptivas como, por ejemplo, estímulos auditivos; y de ser así, si el curso evolutivo de éste sería el mismo que el encontrado con tareas visuales. A nivel atencional, existen diferencias vinculada a la modalidad perceptiva de los estímulos. Los adultos muestran una preferencia por atender a estímulos visuales frente a estímulos auditivos (Dunifon y col., 2016). Estas preferencias atencionales cambian con el desarrollo. Hasta

los 6 años, los estímulos auditivos son los dominantes a nivel atencional (Nava y Pavani, 2013; Robinson y Sloutsky, 2004; Robinson col., 2010; Sloutsky y Napolitano, 2003; Sloutsky y Robinson, 2008). Sin embargo, entre los 9 y 12 años aparece esa preferencia "adulta" por atender a estímulos visuales -frente a auditivos- (Nava y Pavani, 2013). Entre los 6 y los 8 años, ambos estímulos tendrían el mismo impacto a nivel atencional (Nava y Pavani, 2013). Estas diferencias evolutivas podrían tener un efecto en la magnitud del sesgo retrospectivo en función de los estímulos empleados para su evaluación. Por otra parte, la "leve" variación en el procedimiento empleada podría hacer que la medida fuese menos sensible a los cambios que acontecen a lo largo del curso evolutivo dado que podría no involucrar, o hacerlo en menor medida, ciertos aspectos relacionados con la metacognición que sí estarían presentes en el resto de estudios con adultos que emplean la versión social de los diseños hipotéticos.

3.4.Sesgo retrospectivo auditivo

Como comentamos anteriormente, el sesgo retrospectivo ha sido obtenido haciendo uso de estímulos de diferentes modalidades perceptivas, entre ellos, estímulos auditivos (Bernstein y col., 2012; Epley y col., 2004a). En uno de los estudios de sesgo retrospectivo auditivo, los/as participantes escuchaban un fragmento de la canción "Another one bites the Dust" de la banda de rock Queen donde se podía escuchar el mensaje "es divertido fumar marihuana" cuando ésta era reproducida hacia atrás (Epley col., 2004a). A la mitad de los/as participantes se les informó del mensaje antes de reproducir la canción, mientras que a la otra mitad no; y se les pidió que estimasen cuántos de un grupo formado por 100 personas serían capaces de identificar el mensaje sin ser informados de este con antelación. Los/as participantes a los/as que se les informó del mensaje antes de reproducir la canción proporcionaron estimaciones superiores a aquellos/as que no fueron informados del mensaje.

Otros estudios han empleado palabras distorsionadas para evaluar el sesgo retrospectivo auditivo (p.ej., Higham y col., 2017) y han mostrado que el sesgo persiste a pese a instruir a los/as participantes sobre su naturaleza y pedirles que tratasen de evitarlo (Bernstein y col., 2012). No obstante, recientemente Bernstein y colaboradores (2018) han mostrado que el sesgo desaparece cuando se les pide a los/as participantes que identifiquen la palabra distorsionada antes de recibir información sobre su identidad y realizar sus estimaciones sobre el número de personas que serían capaces de identificarla sin conocer su identidad.

Estudios recientes muestran que los procesos implicados en el sesgo retrospectivo auditivo difieren en función del diseño experimental empleado para su evaluación (Higham y col., 2017; Bernstein y col., 2018). En los diseños hipotéticos, el sesgo se ha relacionado con procesos de fluencia ("fluency missatribution"). Estos estudios han hecho uso del priming para demostrar el papel de la fluencia en la emergencia del sesgo basándose en la siguiente lógica (Higham y col., 2017): tanto el priming como el conocimiento de un resultado tiene el mismo efecto en términos de fluencia (ambos facilitan el procesamiento de un estímulo distorsionado), así si la fluencia con la que se procesa un estímulo auditivo es incrementada a través de priming antes de que esta sea incrementada por el resultado (identidad del estímulo auditivo), el efecto del resultado, en términos de fluencia, debería de ser mucho menor que cuando la fluencia no ha sido incrementada previamente. Para poner a prueba esto, Higham y colaboradores (2017) presentaron a los/as participantes palabras semánticamente relacionadas (p. ej., enfermera), o no (p. ej., césped), con una palabra distorsionada (p.ej., "dtr" – doctor -). En la fase de no-información, el priming (relacionado o no) aparecía en una pantalla durante unos segundos e, inmediatamente después, la palabra distorsionada era reproducida. En la fase de información, además del priming, los/as participantes escuchaban una versión nítida de la palabra antes de que esta fuese reproducida en su versión distorsionada. En ambas fases, la tarea de los/as participantes consistió en estimar cuántos de un grupo formado por 100 personas identificaría la palabra escuchando su versión distorsionada. Los resultados mostraron cómo el priming moduló el efecto de sesgo retrospectivo. En concreto, la magnitud del sesgo fue mayor para en los ensayos de priming no relacionado que en los ensayos de priming relacionado. En este mismo estudio, Higham y colaboradores (2017) incluyeron un grupo de participantes que recibían instrucciones de memoria en la fase de información (recordar las estimaciones que proporcionaron en la fase no información). Aquí, a diferencia de en los diseños hipotéticos, el priming no moduló la magnitud del sesgo retrospectivo.

Más recientemente, Bernstein y colaboradores (2018) volvieron a examinar el papel de la fluencia en los diseños hipotéticos, pero en esta ocasión haciendo uso de priming de repetición. Estos autores manipularon el número de presentaciones (cero, uno, tres y seis) del priming. En la fase inicial, los/as participantes escuchaban una, tres o seis veces, versiones no distorsionadas de las palabras. A continuación, en la fase de evaluación, unos/as participantes volvían a escuchar la versión nítida de la palabra distorsionada antes de que esta fuese reproducida (condición nítida-distorsionada),

mientras que otros/as tan sólo escuchaban la versión distorsionada la palabra antes de realizar sus estimaciones (condición de solo distorsión). En ambas fases, y en ambas condiciones, la tarea de los/as participantes fue la misma: estimar el número de personas que identificarían la palabra distorsionada sin haber escuchado la versión nítida de la misma. De nuevo aquí, el sesgo retrospectivo auditivo, estuvo modulado por el priming: el efecto fue mayor para las palabras que no habían sido primadas (cero repeticiones) que para las que habían sido primadas (una, tres o seis repeticiones). Sin embargo, este efecto fue independiente del número de repeticiones. Es decir, no hubo diferencias en la magnitud del sesgo cuando el priming se presentaba una, tres o seis veces.

Los sesgos no sólo se han estudiado al comprender a los otros, sus intenciones o estimar lo que sabríamos u otros sabrían, también y, muy particularmente, al interpretar cómo se comportan las personas y cómo funcionan las cosas, así como para interpretar los eventos y determinar cuáles son sus causas y sus consecuencias.

4. CAUSALIDAD, PENSAMIENTO CONTRAFÁCTICO Y ATRIBUCIÓN DE RESPONSABILIDAD

Desde los primeros meses de vida nuestro cerebro muestra una preferencia por atender a estímulos causales (Moors y col., 2017) priorizando éstos aprendizajes frente a otros, quizás, menos enriquecedores para nosotros/as en esos momentos. Y es que, como señala Daniel Kahneman (2011), buena parte de la cognición humana se basa en el establecimiento de relaciones causales, entre causas y consecuencias. Damos sentido al mundo que nos rodea identificando e interpretando las relaciones causales que en él acontecen. Esto nos permite, por ejemplo, determinar que la causa por la que mi amiga resbaló fue el suelo mojado, que la causa por la que el mayor proyecto espacial español, el satélite SEOSAT-Ingenio, cayó al mar tan sólo 8 minutos después de su despegue fue la mala conexión de varios cables e incluso que la causa por la que las flores de mi abuela se secaron estaba en sus nietos, quienes las regaron con lejía.

A lo largo de la historia el análisis de la causalidad ha sido objeto de interés para disciplinas tales como la filosofía o la psicología. Estas disciplinas han tratado de definir la causalidad en términos de regularidades (Mackie, 1974), probabilidades (Suppes, 1970) o contrafactualidad (Lewis, 1986), entre otros. Sin embargo, todas ellas, en mayor o menor medida, han mostrado limitaciones para explicar fielmente cómo las personas atribuyen causas. A continuación, se describen algunos de los principales modelos de atribución de causalidad en los que el pensamiento contrafáctico tiene un papel importante.

4.1. El modelo de causalidad de Lewis y el problema de la "sobredeterminación"

En términos generales, la teoría estándar de atribución de causalidad (Lewis, 1973, 1986) establece dos requisitos para determinar que X (p. ej., el suelo mojado) causó Y (la caída de mi amigo). El primero de éstos requisitos es que X e Y han de ser ambos ciertos. El segundo, y más restrictivo, es que ha de existir una dependencia contrafáctica entre X e Y. Es decir, en un mundo hipotético donde X no hubiese ocurrido, Y tampoco debería de haber ocurrido. Así, ajustándonos a lo establecido por la teoría podríamos concluir que la causa de que mi amiga resbalase fue el suelo mojado: es cierto que mi amiga resbaló y que el suelo estaba mojado; además, existen una dependencia contrafáctica entre causa y efecto de modo que, si el suelo no hubiese estado mojado, mi amiga no habría resbalado. Sin embargo, ¿qué ocurría con el caso de las flores de mi abuela? ¿Podría la teoría desvelar quién o quiénes fueron los causantes de que éstas se secarán? Helena, Luna, Teo y Lucía, sus cuatro nietos, regaron sus macetas con lejía; por tanto, la acción de los cuatro nietos cumpliría el primero de los requisitos: es cierto que todos regaron sus flores con lejía, así como lo es que éstas se secaron. Sin embargo, la acción individual de cada uno de ellos no superaría el test contrafáctico (el segundo de los requisitos). Así, si sometemos, por ejemplo, la acción de Helena a dicho test podemos comprobar que no existe una relación de dependencia contrafáctica entre su acción ("regar las flores con lejía") y el resultado ("las flores se secaron") dado que, aunque Helena no hubiese regado sus flores con lejía, éstas habrían acabado igualmente secas gracias al "nutritivo" riego de sus otros nietos. Lo mismo ocurriría si sometemos la acción de cualquiera de los tres nietos restantes al test contrafáctico: no habría dependencia contrafáctica entre su acción particular y el resultado y, por tanto, de acuerdo a la Teoría de Lewis (1973, 1986), no sería posible establecer una relación de causa-efecto entre el riego con lejía de las macetas y su posterior secado. Este ejemplo ilustra uno, entre otros (veasé, Collins y col., 2004; Moore, 2009; Lagnado y Gerstenberg, 2016), de los principales problemas del modelo de atribución de causalidad de Lewis (1976, 1986): el problema de la sobredeterminación. Así, aunque el modelo de Lewis permite explicar la causalidad en situaciones en las que una única causa contribuye a un resultado, lo cierto es que ofrece una visión reduccionista de la causalidad pues no permite atribuir causalidad en aquellas situaciones en las que múltiples causas (o agentes) contribuyen a un mismo resultado. Aunque ha habido varios intentos por superar ésta y otras de las limitaciones que presenta el modelo estándar,

quizás, el modelo que en la actualidad ofrece una visión más amplía y dinámica de la causalidad sea el modelo estructural de causalidad (Halpern y Pearl, 2005).

4.2.Modelo Estructural de Causalidad

En el Modelo Estructural (Chockler y Halpern, 2004; Halpern y Pearl, 2005) la causalidad es evaluada en el marco de un evento causal especifico (p.ej. el secado de las flores de mi abuela) donde las múltiples causas (el riego de cada uno de sus nietos) que contribuyen a un resultado (su secado) son representadas como variables. Así, cada una de las variables contempladas en el modelo se relacionan entre sí y con el resultado en cuestión a través de un conjunto de ecuaciones estructurales. Estas ecuaciones estructurales capturan el efecto que cada variable *per se* posee en el resultado (p.ej., el efecto que el regado de cada nieto tuvo en el secado de sus flores) y permiten evaluar las posibles repercusiones que la intervención en una (o varias) de las variables tendría en el resultado. Al igual que el modelo de Lewis (1986), el modelo estructural asume que sólo se puede atribuir causalidad cuando existe una relación de dependencia contrafáctica entre la causa X y el resultado Y. Sin embargo, su concepción dinámica de causalidad le permite establecer grados de responsabilidad causal en base a cómo de cerca un resultado estuvo de depender contrafácticamente de una causa determinada (Chockler y Halpern, 2004). Así, el grado de responsabilidad causal de X en Y vendría dado por el número de cambios requeridos en el modelo causal del evento para que, si X no hubiese ocurrido, Y tampoco hubiese ocurrido.

4.3.Modelo Estructural de Atribución de Responsabilidad

Basándose en el Modelo Estructural (Chockler y Halpern, 2004; Halpern y Pearl, 2005) así como en teorías sobre atribución de responsabilidad legal (Cane, 2002; Hart, 2008; Honoré, 1999), Lagnado y Gerstenberg (2015) desarrollaron un modelo orientado a explicar los juicios de atribución de responsabilidad en situaciones en las que múltiples agentes (o causas) contribuyen a un mismo resultado. La idea central del modelo es que los juicios de responsabilidad son producto de factores prospectivos, denominados criticidad en el modelo, y factores retrospectivos, denominados pivotalidad. La criticidad es evaluada antes de que el resultado sea conocido y refleja la medida en que resultado positivo futuro depende de la acción individual de un agente. En concreto, el componente de criticidad es definido como "*la probabilidad de que la contribución x_i del agente A_i en la situación S sea necesaria para un resultado grupal*" (Lagnado y cols., 2015, p. 222):

$$\text{Criticidad } (A_i, S) = \frac{p\left(\frac{y}{x_i}\right) - p\left(\frac{y}{\neg x_i}\right)}{p\left(\frac{y}{x_i}\right)}$$

En el caso de mi abuela, la estructura causal de la situación (S) determina que sólo un riego con lejía es necesario para que las flores se sequen; por tanto, la criticidad de cada uno de sus nietos sería igual a ¼. En cambio, en una situación hipotética en la que los cuatro riegos con lejía fuesen necesarios para el secado de sus flores, la criticidad de cada uno de sus nietos sería igual a 1.

La pivotalidad, en cambio, es evaluada una vez se conoce el resultado y refleja cómo de cerca estuvo éste de ser contrafácticamente dependiente de la acción del agente. La pivotalidad implica la misma concepción de dependencia contrafáctica contemplada en el Modelo de Lewis (1986): el resultado depende contrafácticamente de la acción del agente o no (el agente es pivotal al resultado o no lo es). Sin embargo, al igual que el Modelo Estructural (Halpern y Pearl, 2005), incluye la noción de grados de dependencia contrafáctica o pivotalidad al resultado. Así, el grado de pivotalidad de un agente vendría determinado por la siguiente fórmula:

$$\text{Pivotality } (A, 0, S) = \frac{1}{(N+1)}$$

Extrapolando esta fórmula al caso de mi abuela (S), la pivotalidad (A, S, O) es el grado de responsabilidad causal de, por ejemplo Helena, en el secado de sus flores (O), y N el número mínimo de cambios requeridos para hacer que el secado de las flores de mi abuela dependiese contrafácticamente del riego de su nieta Helena (Lagnado y Gerstenber, 2015, p. 219). En este caso, serían necesarios tres cambios en la situación actual: se necesitaría que Luna, Teo y Lucía no hubiesen regado las flores con lejía para que el secado de éstas dependiese contrafácticamente de la acción de Helena; por tanto, la pivotalidad de Helena sería igual a 1/3. Sin embargo, en una situación hipotética en la que hubiesen sido necesarios cuatro riegos con lejía para el secado de las flores, Helena habría sido completamente pivotal al resultado (P = 1).

El modelo de Lagnado y colaboradores (2014, 2015) es fruto de los resultados obtenidos en diversas investigaciones con adultos sobre atribución de responsabilidad en grupos (Gerstenberg y Lagnado, 2010; Zultan, Gerstenber y Lagnado, 2012; Lagnado y cols., 2014). En uno de estos estudios Lagnado y colaboradores (2014) pidieron a los/as

participantes que trataran de pulsar el mayor número de veces posible durante un periodo de tiempo determinado un punto que cambiaba de posición en una pantalla. Cada participante jugaba junto a tres jugadores hipotéticos. La criticidad del agente se manipuló variando la estructura causal de los retos de cada equipo. En la condición de estructura conjuntiva, todos los miembros del equipo debían de realizar correctamente su tarea (pulsar un número de veces determinado el punto) para que el equipo ganase. En cambio, en la condición de estructura disyuntiva, el equipo ganaba si al menos uno de los jugadores realizaba correctamente su tarea. Por último, en la condición mixta, el equipo ganaba si dos de los jugadores, por ejemplo, el jugador A y B, y uno de los dos jugadores restantes (jugador C o jugador D) realizaban correctamente su tarea. La tarea de los/as participantes consistió en estimar cómo de importante era la actuación del jugador A en retos con diferentes estructuras causales. Los resultados mostraron como los/as participantes tasaron como más critica la actuación del jugador A en los retos de estructura causal conjuntiva, donde todos/as los/as jugadores/as habían de realizar correctamente su tarea para la victoria del equipo (C = 1), que en los retos de estructura causal disyuntiva donde sólo un/a jugador/a había de realizar correctamente la tarea para el triunfo del equipo (C = ½).

En la segunda fase del estudio, los/as participantes habían de determinar cómo de responsable era el jugador A del resultado del equipo. La pivotalidad del agente se manipuló variando el número de jugadores que realizaban correctamente su tarea. Así, había retos en los que el/la jugador/a A era completamente pivotal al resultado del equipo (P = 1), y otros retos en los que su pivotalidad debido al fallo de otros/as jugadores/as (p. ej., jugador/a A y otro/a jugador/a adicional: P = ½; ó jugador/a A y dos jugadores/as adicionales: P = 1/3). La tarea de los/as participantes consistió en estimar el grado de responsabilidad del jugador A en cada reto. Los resultados mostraron como los/as participantes juzgaron como más responsable al jugador A del resultado del equipo en los retos en los este/a era completamente pivotal al resultado (P = 1) que en aquellos en los que su pivotalidad era menor (P = ½ ó P = 1/3). Y lo que es más importante, estos juicios no sólo dependieron del grado de pivotalidad del agente si no también por su nivel de criticidad. Así, en los retos en los que el/la jugador/a A era completamente pivotal (P = 1) al resultado, los/as participantes juzgaron a éste como más responsable cuando su actuación era máximamente crítica para lograr un resultado positivo (C = 1) que en los retos en los que su criticidad era menor (C = ½). En concreto, estos autores encontraron que cuando se mantenía constante la pivotalidad (P = 1), la responsabilidad atribuida al

jugador/a A dependía de su nivel de criticidad (más crítico más responsable), mientras que cuando se mantenía constante la criticidad, la responsabilidad atribuida al agente variaba en función de su nivel de pivotalidad. Siguiendo, la misma lógica de este estudio, Lagnado y colaboradores (2015) evaluaron si las predicciones del modelo también se cumplen en el caso de atribuciones de responsabilidad a diferentes componentes de un aparato electrónico. Según estos autores, al igual que con agentes, las atribuciones de responsabilidad estuvieron determinadas por el nivel de criticidad y pivotalidad de cada componente; sin embargo, hasta la fecha los resultados de este estudio no han sido publicados.

4.4. Causalidad y pensamiento contrafáctico en niñas/os

En la actualidad se desconoce si el modelo de atribución de responsabilidad propuesto por Lagnado y colaboradores (2014, 2015) puede explicar los juicios de responsabilidad que realizan los/as niños/as. Como comentamos anteriormente, este modelo asume que los juicios de responsabilidad son producto de una atribución causal y, más específicamente, que los juicios de pivotalidad dependen exclusivamente del pensamiento contrafáctico. En esta sección veremos el curso evolutivo de la atribución de causalidad y el pensamiento contrafáctico en niños/as.

Es interesante notar que, aunque se han estudiado la causalidad en niñas/os por un lado, y la comprensión del funcionamiento de dispositivos (p.ej. una balanza) que requieren pensamiento hipotético por otro, no se ha relacionado el modelo de atribución causal de responsabilidad con el desarrollo y las posibles limitaciones de ambos fenómenos. Este modelo indica que se requiere pensamiento contrafáctico para atribuir responsabilidad causal y esta forma de pensamiento se desarrolla tarde.

4.4.1. Atribución de causalidad en niños/as

La mayoría de las investigaciones que han explorado la atribución de causalidad en niñas/os han hecho uso de eventos de movimiento. Estos estudios han demostrado que a los 4 años de edad los/as niños/as ya entienden que las causas han de preceder a las consecuencias (Bullock y German, 1979), siendo capaces incluso de detectar qué tipo de factores (p. ej. Espaciales o temporales) son los relevantes en un evento causal determinado (Schultz, 1982). A esta misma edad, los/as niños/as también que la causa real de un resultado es aquella que esta presente cuando se da el resultado (y ausente cuando este no se produce), es decir, ya han adquirido el principio de covariación (Shultz y Mendelson, 1975). Sin embargo, la capacidad para comprender cómo diferentes causas se interrelacionan para dar lugar a un resultado no aparece hasta finales de la infancia o

principios de la adolescencia. De acuerdo con Inhelder y Piaget (1958), los/as niños/as no son capaces de razonar sobre varias hipótesis hasta el estadio de operaciones formales. Así, por ejemplo, se ha demostrado que hasta los 12 años de edad los/as niños/as no son capaces de integrar la información sobre peso y distancia al eje para predecir si una balanza se mantendrá o no en equilibrio (Siegler, 1976, 1978). Así, tanto la comprensión del funcionamiento de determinados aparatos electrónicos como la capacidad para aceptar o refutar hipótesis en base a la información disponible parece no estar completamente establecida hasta finales de la infancia o principios de la adolescencia (veasé, Klahr, 2000).

Pese a la abundante literatura, existe un gran debate a cerca del tipo de representaciones que se encuentran a la base del razonamiento causal. Algunos/as autores/as defienden que éste surge de las representaciones de movimiento mientras que otros sostienen que el razonamiento causal surge a partir de las representaciones que construyen los/as niños/as a partir de las acciones que perciben en los agentes (Muentener y Bonawitz, 2017). Incluso, hay estudios que sugieren que el origen del razonamiento causal podría ser probabilístico (Saffran y col., 1996; Wu y col., 2011). No obstante, investigaciones recientes sugieren que el razonamiento causal basado en las representaciones de las acciones de agentes podría aparecer más tarde en el desarrollo dada su estrecho vínculo con aspectos como la Teoría de la Mente o habilidades metacognitivas (Muentener y Bonawitz, 2017).

4.4.2. El desarrollo del pensamiento contrafáctico.

El pensamiento contrafáctico o habilidad para imaginar cómo las cosas podrían haber ocurrido de un modo diferente a cómo ocurrieron en realidad (Byrne, 2016) ha sido ampliamente investigado en niñas/os. Sin embargo, no existe consenso a cerca de la edad a la que los/as niños/as son capaces de razonar correctamente con condicionales contrafácticos. Hasta finales de 2008 se pensó que, ya a los 3 o 4 años de edad, los/as niños/as eran capaces de comprender un condicional contrafáctico del mismo modo en el que lo hacen los adultos (Beck y col., 2006; Guajardo y col., 2009; Roldán-Tapia y col., 2017). Sin embargo, investigaciones recientes sugieren que los/as niños/as podrían no adquirir esta habilidad hasta finales de la infancia o inicios de la adolescencia.

El principal argumento esgrimido para explicar esta disparidad de resultados ha sido el tipo de tareas empleadas para la evaluación de dicha habilidad. En concreto, Raftseder y colaboradores (2010) argumentaron que, a edades tempranas, los/as niños/as proporcionan repuestas correctas a condicionales contrafácticos no porque

verdaderamente comprendan un condicional contrafáctico sino porque interpretan éste como un condicional básico de razonamiento, el cual conduciría a la misma respuesta correcta que la interpretación contrafáctica del mismo.

Diversos estudios han puesto a prueba la propuesta Raftseder y colaboradores (2010) introduciendo condiciones en las que una interpretación bicondicional de un condicional contrafáctico conduce a una respuesta incorrecta. En concreto, estos estudios han incluido información contextual en los condicionales contrafácticos la cual anulaba la relación entre antecedente y consecuente (p.ej. Raftseder y col., 2013) y/o han hecho uso de condicionales semifacticos (Gómez-Sánchez y col., 2020; Moreno-Ríos y García-Madruga, 2002), un subtipo de condicionales contrafácticos que anulan y/o debilitan la relación entre antecedente y consecuente (p. ej., Roese y Olson, 1996) para explorar el desarrollo del razonamiento contrafáctico. De manera sistemática, estos estudios han demostrado que cuando las inferencias a realizar no pueden ser resueltas aplicando el condicional básico de razonamiento, sólo los/as preadolescentes (12 años) y los/as adolescentes (13-15 años) son capaces de razonar contrafáctica (o semifácticmente) de modo similar a como lo hacen los adultos (Gómez-Sánchez y col., 2020; Moreno-Ríos y García-Madruga, 2002; Raftseder y col., 2013, 2020). Así, si tal y como defiende Lagnado y colaboradores (2014, 2015) el componente de pivotalidad depende del razonamiento contrafáctico y éste a su vez no esta plenamente adquirido hasta finales de la infancia o inicios de la adolescencia, cabría esperar que las atribuciones de responsabilidad que realizan los/as niños/as reflejasen estos cambios en su razonamiento contrafáctico.

Por otra parte, es posible que las limitaciones que identifican Piaget (1958) y Siegler (1976, 1978) en la comprensión del funcionamiento de dispositivos como la balanza, dónde existen causas potenciales múltiples, estuviese relacionado con la dificultad para comprender el papel de los diferentes elementos y su responsabilidad en producir consecuencias por la limitación del pensamiento contrafáctico reproducido en el componente de pivotalidad.

5. OBJETIVOS GENERALES Y ESPECÍFICOS

El objetivo de la presente tesis doctoral es identificar cambios cognitivos en el desarrollo de dos aspectos centrales de la cognición: la atribución de conocimiento y la atribución de responsabilidad. Para ello, se compara a escolares de entre 8 y 13 años y adultos. En la primera parte de este trabajo se investiga la atribución de conocimiento a través de dos de los denominados sesgos egocéntricos: el sesgo de transparencia ilusoria en la intención (Keysar, 1994), relacionado con la atribución del conocimiento de la intención comunicativa; y el sesgo retrospectivo (Fischhoff, 1975), relacionado con el efecto que el conocimiento de un resultado posee en la atribución de conocimiento a otras personas (Fischhoff, 1975). Como vimos en el Capítulo 1, algunos autores (Epley y col., 2004a; Nickerson, 1999; Royzman y col., 2003) han propuesto que todos los sesgos egocéntricos son producto de un mismo mecanismo: el Heurístico de Anclaje y Ajuste. El estudio del curso evolutivo de ambos sesgos nos permitirá obtener evidencias a favor o en contra de la existencia de un mecanismo común a ambos sesgos. En la segunda parte, se examina cómo son las atribuciones de responsabilidad que realizan niñas/os (de entre 8 y 13 años) y adultos a componentes de un aparato electrónico.

El sesgo de transparencia ilusoria en la intención (Keysar, 1994), no ha sido evaluado en niñas/os; muy probablemente, debido a que la tarea tradicional requiere de la comprensión del sarcasmo. Sin embargo, en la actualidad sabemos que alrededor de los 6 años los/as niños/as adquieren la habilidad para comprender el sarcasmo (p.ej., Harris y Pexman, 2003; Pexman y col., 2011). No obstante, su conocimiento sobre las leyes de comprensión del sarcasmo, así como su conocimiento pragmático en general, continúa desarrollándose a lo largo de la infancia media y tardía (p. ej., Burnett, 2015; Creusere, 2000; Nilsen y col., 2011). Uno de los objetivos de la presente tesis doctoral es determinar si el sesgo de transparencia ilusoria esta presente en niñas/os que ya han adquirido la habilidad para comprender el sarcasmo y, de ser así, explorar el curso evolutivo del mismo. Por otra parte, investigaciones con adultos han mostrado que el sesgo desaparece cuando se les pide a los/as participantes que pronostiquen cómo un lector accidental (p. ej., La secretaria de Julia) interpretaría el mensaje ambiguo (Moreno-Ríos y col., 2011). En este sentido, otro de los objetivos de la presente tesis es conocer si, en escolares, el sesgo desaparece (o disminuye) cuando no existe una relación entre emisor del mensaje y su receptor (receptor accidental). En lo que respecta a la naturaleza del sesgo, pese a que se han ofrecido varias explicaciones al fenómeno (Epley y col., 2004a; Keysar, 1994, 2000; Gerrig y col., 2000; Royzman y col., 2003) aún se desconoce los procesos

subyacentes a éste. Estudios con adultos sugieren que en el sesgo podrían estar actuando principios pragmáticos (Gerrig y col., 2000; Moreno-Ríos y col., 2011). Sin embargo, éstos resultarían insuficientes para explicar el fenómeno de transparencia ilusoria. Así, otro de nuestros objetivos en la presente tesis es explorar los mecanismos que subyacen al sesgo de transparencia ilusoria, examinando las diferentes explicaciones ofrecidas al sesgo de transparencia ilusoria (Gerrig y col., 2010; Keysar, 1994, 2000) vistas en el Capítulo 1, e investigando una explicación alternativa según la cual en el sesgo de transparencia ilusoria, al igual que en el componente de inevitabilidad del sesgo retrospectivo (*creeping determinism*), podrían estar actuando procesos de razonamiento causal.

Por su parte, el sesgo retrospectivo ha sido evaluado en niñas/os empleando diseños de memoria y la versión social de los diseños hipotéticos. En el caso de los diseños de memoria, los resultados son contradictorios. Mientras que un estudio sugiere que la magnitud del sesgo permanece estable entre los 9 y 12 años de edad (Pohl y col., 2010), otro sugiere que alrededor de los 9 años el sesgo comenzaría a disminuir hasta estabilizarse en la adolescencia (Pohl y col., 2018). En cambio, en los diseños hipotéticos, el único estudio que ha explorado el sesgo en niñas/os de estas edades muestra que, a partir de los 6, la magnitud del sesgo permanece estable hasta llegar a la senectud (Bernstein y col., 2011). En el caso de los diseños hipotéticos, las investigaciones que han investigado el sesgo haciendo uso de estos diseños (Bernstein y col., 2004, 2007, 2011), muestran dos limitaciones. Por una parte, las tareas empleadas en estos estudios, a diferencia de las utilizadas con adultos (p. ej., Higham y cols., 2017), no requiere que los escolares emitan un juicio sobre el número de personas que serían conocedoras de una determinada pieza de información, si no que emplean una medida más "implícita" o "indirecta" del sesgo (cuándo otra persona podría identificar un estímulo visual degradado que progresivamente se va clarificando). Como señalan Ghrear y colaboradores (2016), esto podría hacer a estas tareas menos sensibles a los cambios evolutivos. Así, otro de los objetivos en el presente trabajo es desarrollar tareas aplicables a niñas/os (y adultos) que ofrezcan una medida cuantitativa de la ejecución. Por otra parte, como se señaló en el capítulo de introducción, los estímulos utilizados en las tareas hipotéticas para la evaluación del sesgo han sido exclusivamente visuales. Por tanto, se desconoce si los niños exhibirán el sesgo cuando los estímulos empleados para su evaluación pertenecen a otra modalidad perceptiva como, por ejemplo, la auditiva. Otro de los objetivos del presente trabajo es determinar si los/as niños/as exhiben sesgo

retrospectivo con estímulos auditivos y, de ser así, determinar el curso evolutivo del mismo. Se sabe que con el desarrollo cambian las preferencias de los/as niños/as por atender a estímulos de una u otra modalidad perceptiva. Hasta los 6 años, los/as menores muestran una preferencia por atender a estímulos auditivos (p. ej., Sloutsky y Napolitano, 2003; Sloutsky y Robinson, 2008). Entre los 9 y 12 esa preferencia cambia, siendo los estímulos de carácter visual los predominantes a nivel atencional (Nava y Pavani, 2013). En el período intermedio (6 a 8 años), no habría una preferencia vinculada a la modalidad perceptiva del estímulo (Nava y Pavani, 2013). Estas diferencias evolutivas vinculadas a la modalidad perceptiva de los estímulos podrían tener un impacto en la magnitud del sesgo de retrospectivo en función de los estímulos empleados para su evaluación. Así, podría ocurrir que la preferencia por atender a estímulos visuales entre los 9 y 12 años, se tradujese en un procesamiento más profundo de estos estímulos (en comparación con los auditivos) y que, a su vez, esto se tradujese en un mayor sesgo en la tarea visual (frente a la auditiva) a estas edades. Otro de los objetivos del presente trabajo es conocer si la magnitud y el desarrollo del sesgo retrospectivo difiere en función de la modalidad perceptiva de los estudios empleados para su evaluación (visuales y auditivos).

Recientemente se ha comprobado que en adultos el sesgo retrospectivo auditivo es susceptible de ser eliminado haciendo conscientes a los participantes de la dificultad para identificar un estímulo auditivo distorsionado en ausencia de información sobre su identidad (Bernstein y col., 2018). En este sentido, otro de los objetivos de este trabajo es determinar si el sesgo retrospectivo auditivo puede ser eliminado, o reducido, en la infancia. Por otra parte, investigaciones recientes (Bernstein y col., 2018; Higham y col., 2017) han demostrado que los procesos que subyacen al sesgo retrospectivo auditivo difieren en función del tipo de diseño empleado para su evaluación (diseño de memoria o diseño hipotético). Estos estudios han basado en el efecto diferencial que el priming posee en la magnitud del sesgo retrospectivo cuando es evaluado a través de uno u otro diseño. En concreto, estos estudios han comprobado que incrementar la fluencia con la que un estímulo distorsionado es procesado a través de priming antes de recibir información sobre el resultado (identidad del estímulo) reduce el efecto del conocimiento de un resultado, únicamente en los diseños hipotéticos. En este sentido, otro de nuestros objetivos es investigar si diferentes procesos subyacen al sesgo retrospectivo auditivo en cada diseño experimental, analizando sí, al igual que ocurre en adultos (Bernstein y col., 2018; Higham y col., 2017), el primming modula la magnitud del sesgo retrospectivo sólo en los diseños hipotéticos.

La segunda parte de la presente tesis doctoral la dedicaremos a investigar los cambios cognitivos asociados a la atribución de responsabilidad, utilizando como marco el Modelo Estructural de Atribución de Responsabilidad (Lagnado y col., 2015). La idea central del modelo es que las personas asignan responsabilidad a un agente de un resultado común basándose en dos factores: criticidad y pivotalidad. La criticidad del agente es evaluada antes de que el resultado sea conocido y refleja cómo de importante o necesaria es la contribución del agente para lograr un resultado positivo. La pivotalidad, en cambio, refleja el grado en el que el resultado depende contrafácticamente de la acción del agente, es decir, de la medida en que cambiando la acción del agente el resultado habría sido diferente (pivotalidad). En adultos, el modelo ha sido evaluado en situaciones que requieren asignar responsabilidad de un resultado a una persona o a un componente de un aparato electrónico (Lagnado y col., 2014, 2015). Sin embargo, en este último caso, sólo el componente de pivotalidad ha sido evaluado (Lagnado y col., 2015). Además, se ha propuesto que la atribución de responsabilidad opera igual si se atribuye a personas o a dispositivos mecánicos. Existen dos razones para elegir dispositivos mecánicos en nuestra evaluación: 1) podemos aislar la influencia de la atribución de creencias y la TOM que cambian con el desarrollo y 2) podemos contribuir a analizar un dominio de estudio clásico en el desarrollo cognitivo no resuelto desde hace años: explicar las dificultades de los/as escolares al comprender el funcionamiento de dispositivos mecánicos.

Por otra parte, Lagnado y colaboradores (2014) defienden que el factor de pivotalidad es puramente contrafáctico ya que requiere que los individuos consideren un mundo hipotético (o varios) en el que, deshaciendo la acción del agente, el resultado no habría ocurrido. Investigaciones recientes sugieren que la habilidad para razonar contrafácticamente se desarrolla paulatinamente durante la infancia media y tardía (p.ej., Gómez-Sánchez y cols., 2020; Raftseder y Perner, 2013). Otro de nuestros objetivos es determinar si existe una relación entre las habilidades de razonamiento contrafáctico de los/as niños/as y su sensibilidad a los diferentes niveles de pivotalidad de un componente del dispositivo electrónico.

En síntesis, los objetivos generales y específicos de la presente tesis doctoral son:

1. Investigar el sesgo de transparencia ilusoria en niños de entre 8 y 13 años y adultos.
   1.1. Evaluar el sesgo de transparencia ilusoria en niños que ya han adquirido la habilidad para interpretar el sarcasmo.
   1.2. Determinar, de exhibir los niños el sesgo, el curso evolutivo del mismo.

1.3. Evaluar si, al igual que ocurre en adultos, el sesgo desaparece (o disminuye) cuando no hay relación entre emisor y receptor del mensaje (receptor accidental).

1.4. Investigar la naturaleza de los procesos implicados en el sesgo de transparencia ilusoria

1.4.1. Analizar la relación entre el sarcasmo atribuido al receptor y la magnitud del sesgo.

1.4.2. Explorar si, en niños, están actuando principios pragmáticos y/o procesos de *construal*.

1.4.3. Determinar si en el sesgo están actuando procesos de razonamiento causal.

2. Investigar el sesgo retrospectivo visual y auditivo en niños de entre 8 y 13 años.

2.1. Desarrollar medidas cuantitativas del sesgo que ofrezcan una medida cuantitativa de la ejecución.

2.2. Evaluar si niños/as de entre 8 y 13 años exhiben el sesgo retrospectivo auditivo.

2.3. De ser así, determinar el curso evolutivo del sesgo retrospectivo auditivo.

2.4. Explorar si el sesgo retrospectivo auditivo en niños/as puede ser eliminado o reducido.

2.5. Investigar los procesos psicológicos que subyacen al sesgo retrospectivo auditivo en la infancia en los diseños hipotéticos y de memoria.

2.6. Determinar si la magnitud y curso evolutivo del sesgo retrospectivo difieren en función de la modalidad perceptiva de los estímulos empleados para su evaluación (visuales o auditivos).

3. Investigar el modelo de atribución de responsabilidad desarrollado por Lagnado y colaboradores (2015) en niños de entre 8 y 13 años y adultos en un contexto de atribución de responsabilidad a componentes de un dispositivo electrónico.

3.1. Evaluar si niños y adultos son sensibles al nivel de criticidad y pivotalidad del componente.

3.2. Analizar si el grado de responsabilidad asignada al componente es el producto del nivel de criticidad y pivotalidad de éste.

3.3. En niños, determinar el curso evolutivo de los juicios de criticidad y pivotalidad.

3.4. Evaluar si existe una relación entre las habilidades de razonamiento contrafáctico de los niños y su sensibilidad al componente de pivotalidad.

El objetivo 1 se aborda en los estudios 1 y 2 donde se evalúa el sesgo de transparencia ilusoria en niños/as y adultos respectivamente. Los objetivos 1.1 y 1.2 se ponen a prueba en el estudio 1 donde se evalúa el sesgo de transparencia ilusoria en niños/as de entre 8 y 13 años. En este mismo estudio, también se evalúan los objetivos 1.3 y 1.4.2. En el estudio 2, se aborda el objetivo 1.4 (en concreto, 1.4.1 y 1.4.3) investigando los procesos que subyacen al sesgo de transparencia ilusoria. En concreto, se investiga si en el sesgo de la transparencia ilusoria, al igual que en el componente de inevitabilidad del sesgo retrospectivo (*creeping determinism*), están actuando procesos de razonamiento causal. Para ello se incluye en la tarea tradicional una serie de manipulaciones que estudios previos han mostrado que alteran la magnitud del *creeping determinism* y se evalúa su efecto en la magnitud del sesgo de transparencia ilusoria. Además, a través de una de estas manipulaciones, se evalúa una de las explicaciones propuestas por Keysar ("transparencia de la intención").

El objetivo 2 se aborda, en dos experimentos, en el estudio 3 donde se evalúa el sesgo retrospectivo en escolares de entre 8 y 13 años de edad. Todos los objetivos específicos contemplados en este, salvo el objetivo 2.6, se abordan en el Experimento 1 donde se investiga el sesgo retrospectivo auditivo y visual, y se incluye una condición en la tarea auditiva donde los/as niños/as realizan dos estimaciones sobre una misma canción (sin información y con información). El objetivo 2.6 se aborda en el Experimento 2 donde se incluye en la tarea auditiva una manipulación de priming por repetición con el fin de alterar la fluencia con la que el estímulo auditivo es procesado y se pide a los/as participantes que completen dos tareas de sesgo auditivo, una de diseño hipotético y otra de diseño de memoria.

El objetivo 3 se pone a prueba en el estudio 4 a través de dos experimentos. En ambos experimentos se les pide a los/as participantes que completen una tarea que requiere que juzguen la criticidad y pivotalidad de varios cables para el funcionamiento de una batería. Los objetivos 3.1 y 3.2. se ponen a prueba en ambos experimentos (Experimento 1 con adultos y Experimento 2 con niños/as). En cambio, los objetivos 3.3. y 3.4. específicamente se abordan en el Experimento 2 con escolares donde estos, además de completar la tarea de la batería, responden a un cuestionario orientado a evaluar sus habilidades de razonamiento contrafáctico.

SECCIÓN EMPÍRICA

En los siguientes capítulos veremos la sección empírica de la presente tesis doctoral que consta de 4 estudios, divididos en cuatro capítulos:

- Capítulo 2: *Children's "illusory transparency" of intention and interpretation: the construal vs pragmatic view.*
- Capítulo 3: "*I know that you know because it was known": a link between illusory transparency and hindsight bias.*
- Capítulo 4: *Auditory hindsight bias in school-age children*
- Capítulo 5: *Responsibility attribution by children and adults with mechanical devices*

Como comentamos en la sección anterior, en el capítulo 2 y 3 investigamos el sesgo de transparencia ilusoria en niños/as y adultos. En concreto, en el capitulo 2 se estudia el sesgo de transparencia ilusoria en niños/as y se evalúa si el sesgo desaparece cuando el receptor del mensaje es un lector accidental. Además, se ponen a prueba dos de las explicaciones ofrecidas al sesgo de transparencia ilusoria: la explicación pragmática y la explicación de la *construal*. En el capítulo 3 se investiga la naturaleza del sesgo de transparencia ilusoria haciendo uso de manipulaciones que estudios previos han mostrado que alteran la magnitud del sesgo retrospectivo (*creeping determinism*). Asimismo, se pondrá a prueba la explicación de "transparencia de la intención" propuesta por Keysar (1994).

En el capitulo 4 se evalúa el sesgo retrospectivo auditivo y visual en escolares, y se examina si el sesgo retrospectivo auditivo es susceptible de ser eliminado o reducido. Además, se investiga la naturaleza de los procesos que subyacen al sesgo retrospectivo auditivo en cada tipo de diseño experimental (de memoria e hipotético) introduciendo una manipulación de priming por repetición y evaluando el efecto de ésta en la magnitud del sesgo retrospectivo auditivo en cada diseño.

Por último, en el capítulo 5 se pone a prueba el Modelo Estructural de Atribución de responsabilidad propuesto por Lagnado.y colaboradores (2015) en niños/as y adultos a través de una tarea que requiere que los/as participantes juzguen la criticidad y pivotalidad de un componente de un dispositivo electrónico para el funcionamiento (o no funcionamiento, en el caso de la pivotalidad) de dicho dispositivo. Además, investiga si, tal y como contempla el modelo, el factor de pivotalidad depende el pensamiento contrafáctico. Para ello, se pide a los/as niños/as que respondan a un cuestionario

orientado a evaluar sus habilidades de razonamiento contrafáctico y se relacionan estas con su sensibilidad al factor de pivotalidad.

Capítulo 2

Children's "illusory transparency" of intention and interpretation: the construal vs the pragmatic view

Abstract

This paper explores the illusory transparency phenomenon in childhood and tests two accounts of the phenomenon: the construal and the naïve (pragmatic) accounts. One hundred and nine children between 8 and 13 years old participated in an adapted task of illusory transparency. Children read a story in which a character wrote an ambiguous message to another (addressee) that could be interpreted as sarcastic, this being dependent on information known only to the writer (privileged knowledge). The reader of the message could be the addressee or an accidental reader. We asked children about the intention of the writer to be sarcastic and whether the reader interpreted it as sarcastic. Also, we asked explicitly whether the reader could have had access to the privileged knowledge. The results showed illusory transparency of intention in children, as had been shown in adults. Moreover, while attribution of sarcastic intention increased with age, the magnitude of the illusory transparency did not. The results from an explicit measure support the pragmatic account: children attributed more capacity of knowledge to the addressee than to the accidental reader. However, the implicit (traditional) measure suggests the implication of construal operations. We propose that these results are due to the fact that information processing operates in two different ways, consistent with each approach. The developmental results are discussed.

Keywords*:* illusory transparency, sarcasm, children, cognitive development, knowledge attribution, perspective-taking.

Manuscript published as: Gordo, C., & Moreno-Ríos, S. (2019). Children's Illusory Transparency of Intention: Construal versus Pragmatic View. *Discourse Processes, 56*(4), 331-344. https://doi.org/10.1080/0163853X.2018.1508264

Introduction

Healthy social interactions between schoolchildren require some expertise in distinguishing what they know and what others know in order to interpret correctly other people's feelings, behaviours and attitudes. This ability develops through life. However, some research with adults has documented the existence of several cognitive biases which make perspective-taking difficult. One of them, called illusory transparency (Keysar, 1994; Mante-Estacio & Bernardo, 2015; Moreno-Ríos, Rodríguez-Menchén, & Rodríguez-Gualda, 2011; Ståhl & Ellemers, 2016; Todd, Hanko, Galinsky & Mussweiler, 2011; Weingartner & Klin, 2005, 2009), consists of attributing our privileged knowledge of an event to another individual. That is, assuming that another person will know something that only we know. That information could be necessary to interpret an event correctly. This happens, for example, when in a horror film we expect the main character not to open the door when only we know that behind it is the murderer.

Illusory transparency has been evaluated through a task in which participants had to read a story about two colleagues. In this story, David asked Julia, his partner, about a restaurant to go to that night with his parents. The next day, David could not find Julia and left her the following message: "You wanted to know about the restaurant? Well, it was marvellous, just marvellous" (Keysar, 1994, p. 173). Depending on the kind of privileged information that participants received, positive (the food was delicious, and the service was impeccable) or negative (the food was revolting, and the service was mediocre), the message left by David was interpreted as sincere when the information was positive or as sarcastic when the information was negative. If participants considered just the stated facts and were able to discard their own knowledge (privileged knowledge), they would have attributed to Julia a literal interpretation of the message, without sarcasm. However, participants attributed a sarcastic interpretation, showing what was called the "illusory transparency of intention" bias (Epley, Keysar, Van Boven & Gilovich, 2004).

Sarcasm is a form of non-literal language. Its comprehension requires a recipient to be able to detect the incongruence between what is said and what is meant (e.g., Capelli, Nakagawa, & Madden, 1990; Ivanko & Pexman, 2003; Nilsen, Glenwright, & Huyder, 2011). In illusory transparency tasks, the recipient knows what the writer of the message says but not what he or she means in the negative condition. In this condition, to understand what the writer really means (that is, to note that he/she is being sarcastic), the addressee should have information about the experience. This information would enable the addressee to interpret the message as sarcastic. However, the incongruity

between the message and the experience is not available to the recipient. Thus, this bias reflects the extent to which "the privileged information about an event causes subjects to attribute the speaker's intention to the addressee" (Keysar, 1994; p. 172). The illusory transparency was computed by the difference in scores in the attribution of sarcastic interpretation of the message by the addressee depending on whether participants received privileged positive or negative information (Keysar, 1994).

Illusory transparency with the sarcasm tasks has been demonstrated with adults, but as far as we know, not yet in children. The present study explores children's ability to take a character's perspective in stories and, particularly, to discover whether illusory transparency of intention is present in children and whether the magnitude of this bias changes with age.

Other biases, such as the "curse of knowledge" or "hindsight bias", have also been studied in children. In the case of the "curse of knowledge" (Camerer, Lowenstein, & Weber, 1989) tasks, children have to ignore a piece of information (for example, the contents of a bag) in order to predict the response of an uninformed other (for example, whether she/he would or wouldn't know the contents of the bag). Results showed children's tendency to be biased by their own knowledge when judging the perspective of a less informed other. The greatest bias was exhibited in early years, and decreased with age (e.g. Birch & Bloom, 2003. Similarly, in the "hindsight bias" (Fischhoff, 1975), very young children (and older adults) are those exhibiting the greatest bias (Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011). This bias shows people's tendency to assert, once they know the outcome of an event, that they or another person knew it all along (e.g. Bernstein, Atance, Meltzoff, & Loftus, 2007). Some authors have argued that the two biases are, in fact, the same phenomenon with different labels (Birch, Brosseau-Liard, Haddock & Ghear, 2017). These distinctive biases share a core feature with the illusory transparency bias: the tendency to contaminate a naïve perspective with children's own privileged knowledge (Birch & Bernstein, 2007; Royzman, Cassidy & Baron, 2003).

This bias has been studied mainly with Keysar's sarcasm task, and in a few cases with other tasks (see for example, Rai, Mitchell, Kadar & Mackenzie, 2016). The advantage of Keysar's task over others is that it tests illusory transparency of intention by including the assessment of a character's perspective in talking about the intention to communicate some information. In this way, illusory transparency can be more clearly distinguished from other perspective-taking biases, such as the "curse of knowledge" or "hindsight bias" (see Royzman et al., 2003).

In spite of it being a robust phenomenon in adults, it is not clear which mechanisms are involved in illusory transparency. Two different accounts have tried to explain the bias. On the one hand, the defendants of the construal account (Keysar, 2000) argue that the attribution of a reader's own knowledge, and therefore, the sarcastic interpretation attribution to the addressee (and speaker's intention) in the negative condition is a product of the message's "change" of meaning. Namely, they "change" the original message in their mind by a "construal" operation of disambiguating that "ambiguous" message. Participants mentally transform the original message, adding "the experience was bad" when actually that was not present in the message, and they attribute that whole meaning to the addressee.

On the other hand, from the "naïve-pragmatic" account, it is argued that Keysar's task involved pragmatic principles that could explain the experimental results. According to Gerrig, Ohaeri & Brennan (2000), the greater degree of sarcasm attribution obtained in the negative condition could be explained not by illusory transparency, but by the reader's resistance to interpreting that the writer intended a violation of Grice's (1975) cooperative principle and, more specifically, of the maxim of quality (however, see Moreno-Ríos et al., 2011). That is, the reader could anticipate that if the information provided by David was not complete, it was because he knew that Julia (his friend) should have additional information to understand the message completely.

Although both accounts assume that the bias originates in a wrong attribution to the addressee (e.g. "bad experience" after the message "good experience"), they differ in *what* is wrongly attributed by the reader: a reconstructed message (e.g. "bad experience" instead of "good experience"; construal account, Keysar, 2000), or the original message (e.g. "good experience") with additional pragmatic information (e.g. "the addressee should have clues to let them know that the experience was bad"), which allows the addressee to interpret the message properly (naïve account, Gerrig et al., 2000).

Testing what is wrongly attributed to the addressee, and what is the genesis of such bias in childhood is another aim of this research. For this, we included a condition in which the message to Julia left by David is read by an accidental reader (e.g. Julia's secretary; see Experiment 2 Moreno-Ríos et al., 2011) and we ask participants whether the reader of the message (the addressee or an accidental reader) can know the actual experience of the sender of the message. This measure will tell us whether the shared knowledge identified by the naïve account as the element causing the bias is present only

when there is a link between sender and receiver. On the other hand, from the "construal" account the bias will be present in both readers of the message.

Some studies have included a third character to manipulate the spread of the attribution of the privileged information between different characters. For example, when the third character did not have the privileged knowledge and was the writer of the message, the attribution of intention of sarcasm to the writer of the message was reduced (Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009). However, although the pragmatic effect cannot explain the results of illusory transparency, it does seem to influence it. In an extreme condition with adults, the illusory transparency disappeared when, in the story, the writer and the addressee of the ambiguous message did not know each other (see Experiment 2; Moreno-Ríos et al., 2011).

Results, such as those previously mentioned, also showed that the process of adjustment in the illusory transparency bias is mediated by the knowledge of how communication works and that only people who know each other can share knowledge. Accordingly, Savitsky, Keysar, Epley, Carter, and Swanson (2010) enunciated what they called "the closeness-communication bias". They found that egocentrism was higher among friends than with strangers. It is uncertain whether this effect, tested with adults, is also shown by children.

Study Aims

The aim of this study is to test whether illusory transparency of intention is present in children aged between 8 and 13 years and to determine whether the magnitude of this bias changes with age. We chose this age group for several reasons. First, younger children could have difficulties in perspective-taking. In addition, previous studies have found that the ability to understand sarcasm, that is, the ability to detect that a sarcastic speaker does not really believe what he or she has said, emerges around the age of 5 or 6, and continues to develop through middle childhood (Burnett, 2015; Creusere, 2000; Dews et al., 1996; Glenwright & Pexman, 2010; Harris & Pexman, 2003; Pexman et al., 2011). Finally, children from 8 years are able not only to infer from a positive statement after a negative outcome what a speaker really believes but, are also capable of adjusting their answers about how a listener would interpret this sarcastic statement according to the listener's knowledge state - knowledgeable or ignorant (Nilsen et al., 2011).

Although previous findings suggest that children of this age group are able to complete the traditional illusory transparency task, we adapted the task to guarantee its correct comprehension by schoolchildren. Specifically, as in other studies which have

assessed sarcasm comprehension in children (see for example, Glenwright & Pexman, 2010; Nilsen et al., 2011), we avoided using the term "sarcastic" in the questions, asking instead whether the ambiguous message would be interpreted by the addressee as sincere (instead of sarcastic) and whether the speaker's intention was to be sincere.

With these changes in the task, eleven additional stories were created with the same structure as the traditional task. In addition, we included a condition in which the message was read by an accidental reader instead of the addressee (Moreno-Ríos et al., 2011).

Based on previous theoretical accounts of the development of biases (see Royzman et al., 2003) at least two contrary predictions can be expected for illusory transparency in childhood. On the one hand, if illusory transparency and hindsight bias share a common mechanism, as some authors have proposed (e.g., Epley et al., 2004), it will be present in childhood and the magnitude of the bias will decrease with age until it stabilises (Pohl Bayen & Martin, 2010). On the other hand, as we have seen in studies with adults, the effect is, at least, sensitive to the knowledge about how communication works between known and unknown people. Children are gaining experience about this knowledge so older children will have a better comprehension of social functioning rules than young children. For this reason, if the illusory transparency is based on that increasing knowledge, we would expect an increase in the bias during the school years. This better comprehension of social functioning rules that older children have leads us to predict that, as occurred in studies with adults (Moreno-Ríos et al., 2011), illusory transparency will disappear when the receiver of the message is an accidental reader, but only for older children, given their better comprehension of those rules.

With the previously mentioned question, we test the presence and magnitude of the transparency illusion effect, as tested with adults. Also, we include an additional question that helps us to contrast the two theoretical explanations of this phenomenon: Can the reader know the experience of the sender of the message? Table 1 shows predictions from the naïve and the construal approaches in the "sarcastic" task. Participants interpret the ambiguous message ("the restaurant was marvellous, just marvellous") in the positive (the experience in the restaurant was good) and negative (the experience was bad) conditions, when the message was read by the addressee (June, friend of David) and by an accidental reader (June`s secretary) who did not know the writer, and therefore, could not share knowledge about the experience (David in the

restaurant). The comparison of conditions with accidental readers and with addressees can help to contrast the different predictions for the two accounts.

Participants will tend to conclude that in the positive condition the reader can know what actually happened (David says the experience was marvellous, and in fact, it was), but this is not so in the negative condition. In these conditions, different results are expected depending on the approach (see Table 1). The naïve approach clearly predicts an effect of the addressee: participants will claim that the reader of the message can know the experience, but only when the reader is the addressee (Julia, David's friend) and not when the reader is an accidental reader (Julia's secretary). That is, they use the assumption of shared knowledge only in the addressee condition because the two people know each other, but not in the accidental reader condition.

Table 1. Naïve and Construal predictions in negative conditions depending whether the reader of the message is the addressee or an accidental reader, and what causes the wrong attribution to the addressee.

	Addressee	**Accidental Reader**
Naïve	Shared Knowledge	
Construal	Construal Operation	Construal Operation

In the construal approach, no effect of reader of the message is expected. In the negative conditions, the ambiguous message will induce a change in the representation of the message, assuming erroneously that the reader will read what actually happened (will perceive the message as sarcastic), and therefore, will know what happened (David's awful experience in the restaurant). Moreover, if no construal operation is working, we expect that in the negative condition, all the participants will say that the accidental reader cannot know what actually happened because they do not have any source of information about it.

Method

Participants: A hundred and nine schoolchildren (50 girls and 59 boys) from 3rd to 6th grades, aged between 8 and 13, performed the task. They were all native speakers of Spanish and were enrolled in state schools in Granada, Spain. The only selection criterion was consent from both parents and children to participate in the study.

Materials: Following Keysar (1994) and Moreno-Ríos et al. (2011), twelve stories were constructed and adapted for children, each with three characters (see Appendix for an example). All the stories ended with an ambiguous message that could be interpreted as sincere or sarcastic depending on the content of the story. Unlike in studies with adults, we asked children whether the message could be interpreted as sincere and whether the writer's intention was to be sincere. Four versions were made for each story (see Table 2), depending on whether the privileged information was positive or negative (e.g., the food was delicious and the service was impeccable *versus* the food was revolting and the service was mediocre); and the reader of the message, an addressee or an accidental reader (e.g., Julia, addressee of the message versus Julia's secretary, accidental reader of the message).

Table 2. Text structure in the four conditions

Conditions	**Text structure**
Positive version addressee reader	Julia recommends a restaurant to her partner David. David has a *good experience* in the restaurant. David writes to Julia: "the restaurant was marvellous, just marvellous". Julia reads the message.
Negative version addressee reader	Julia recommends a restaurant to her partner David. David has a *bad experience* in the restaurant. David writes to Julia: "the restaurant was marvellous, just marvellous". Julia reads the message.
Positive version accidental reader	Julia recommends a restaurant to her partner David. David has a *good experience* in the restaurant. David writes to Julia: "the restaurant was marvellous, just marvellous". *A different person reads the message.*
Negative version accidental reader	Julia recommends a restaurant to her partner David. David has a *bad experience* in the restaurant. David writes to Julia: "the restaurant was marvellous, just marvellous". *A different person reads the message.*

Eight booklets were constructed, each with twelve stories. In each booklet, the same story was presented in a different version (addressee-positive; addressee-negative;

accidental reader-positive; accidental reader-negative) and in a different order. The version of each story was assigned randomly for each booklet. In the first four booklets, the stories were presented from Story 1 to Story 12, while the remaining four booklets were arranged from Story 12 to Story 1. Therefore, for example, in Booklet 1, we presented the addressee-positive version of the story, Story 1, whereas Booklet 6 told the accidental reader-negative version of that story. In Booklet 1, Story 1 was the first, while in Booklet 6, this story was the last.

Each story included five questions with dichotomous answers ("yes" or "no"). Two of these were comprehension questions while the three remaining questions assessed the writer's intention (Is David sincere when he says to Julia that his experience in the restaurant was marvellous?), addressee's interpretation (Will Julia believe that David is sincere when he says that the restaurant experience was marvellous?) and knowledge attribution to the addressee (for addressee: Can Julia know whether David's experience in the restaurant was good? For Accidental reader: Can Julia's secretary know whether David's experience in the restaurant was good?). This last question provided us with a measure of the comprehension of reality whereby the two main accounts could be contrasted.

Procedure: Assessments were carried out at the school, in a silent room without distractions. Before the experimental session started, we assessed whether children knew that a sarcastic message involves a non-sincere interpretation of it through a number of social situations in which a sarcastic message can appear (e.g. When you ask your mother for a mobile phone and she says to you: I am going to buy you two mobile phones instead of one, is your mother being sincere?). Children did not start the experimental session until they had responded properly to three of these situations. If children made any errors, the experimenter provided further explanation. Each participant received a booklet with instructions. Participants were instructed to read the stories carefully and to stay quiet at the end of the task until everybody had finished. Participants from 5th and 6th grades were evaluated in groups of about 15 to 20 students, whereas those from 3rd and 4th grades were divided into eight small groups of two or three students (equal to the number of conditions) and the reading was made jointly with the experimenter in order to guarantee its comprehension. The experimenter was instructed to read the stories in a neutral tone of voice. In both cases, the order of presentation was counterbalanced.

Results

Four per cent of the participants were excluded for answering the comprehension questions incorrectly. The responses of sarcasm were coded, following Epley et al. (2004), with scores of -1 for "Yes" answers (indicating sincerity, the absence of sarcasm), and 1 for "No" (meaning interpretation or intention of sarcasm). Thus, for each participant in every condition, an index of responses of sarcasm (from 1 to -1) was calculated. The illusory transparency effect involves more sarcasm in the negative than in the positive condition. Finally, the averages of intention attribution and sarcastic interpretation were submitted to an ANOVA 2 source of the sarcasm (writer's intention or addressee's interpretation), X 2 reader of the message (addressee or accidental reader) x 2 polarity of the event (positive or negative) x 2 levels (2nd or 3rd levels) with the three first factors manipulated within-participants and the fourth between groups. Analyses are reported across both subjects (F1) and items (F2). Results are shown in Tables 3 and 4.

The analysis revealed a higher effect for intention of the writer of the message than for the interpretation, 0.85 vs. -0.44, $F1(1,107) = 530.78$, $p<.001$, $\eta p^2=.83$; $F2(1,22) = 154.20$, $p<.001$, $\eta p^2=.88$ and for negative than for positive events, 0.28 vs 0.13, $F1(1,107) = 21.41$, $p<.001$, $\eta p^2=.17$; $F2(1,22) = 352.78$, $p<.001$, $\eta p^2=.94$. There was not a significant effect of the reader of the message, $F1(1, 107) = 0.18$, $p>.67$, $\eta p^2= .00$; $F2(1,22) = 1.29$, $p >.05$, $\eta p^2=.05$, or for children's level, $F1(1, 107) = 0.02$, $p>.90$, $\eta p^2<.01$; $F2(1,22) = 0.67$, $p >.05$, $\eta p^2=.03$. The analysis also showed only two interactions: one between the intention of the writer of the message and the polarity of the event, $F1(1, 107) = 24.71$, $p<.001$, $\eta p^2=.19$; $F2(1,22) = 416.30$, $p<.001$, $\eta p^2=.95$, and another between the same two factors and children's level, $F1(1, 107) = 5.84$, $p<.05$, $\eta p^2=.05$; $F2(1,22) = 27.30$, $p<.001$, $\eta p^2=.55$. To analyse these interactions, we decided to carry out two separate analyses for intention and for interpretation (including the reader factor).

Sarcasm from the intention of the writer. Analysis.

The averages of intention attribution were submitted to an ANOVA 2 reader of the message (addressee or accidental reader) x 2 polarity of the event (positive or negative) x 2 levels (2nd or 3rd levels) with the first two factors manipulated within-participants and the third between groups. Results are shown in Table 3.

The analysis showed a main effect for polarity of the event: more sarcasm was attributed to the intention of the character when the polarity of the event was negative than when it was positive, $F1(1,107) = 422.06$, $p<.001$, $\eta p^2=.80$; $F2 (1,22) = 840.81$,

p<.001, ηp^2=.98. The effect was shown for younger, 0.60 vs. -0.76, F1(1, 51) = 110.81, p< .001, ηp^2=.69; F2(1,11) = 263.88, p<.001, ηp^2=.96, as well as older, 0.87 vs. -0.80, F1(1, 56) = 437.64, p<.001, ηp^2=.89; F2(1,11) = 674.32, p<.001, ηp^2=.98, children. However, this effect was greater for third level students than for younger ones, F1(1, 107) = 8.14, p<.05, ηp^2=.06; F2(1,22) = 14.53, p<.001, ηp^2=.40. No other significant effect was found.

Table 3. The degree of sarcasm (from 1 to -1) attributed to the intention of the character according to the privileged information and the reader of the message for 2nd level (3rd 4th) and 3rd level (5th 6th graders). Standard deviations between parentheses.

		Positive Version	**Negative Version**
Addressee	2nd level	-0.76 (0.51)	0.61 (0.68)
	3rd level	-0.80 (0.52)	0.87 (0.37)
	Total	-0.78 (0.53)	0.74 (0.55)
Accidental reader	2nd level	-0.69 (0.61)	0.49 (0.77)
	3rd level	-0.78 (0.52)	0.94 (0.30)
	Total	-0.73 (0.56)	0.72 (0.61)

We made an additional analysis to check whether the increase of the bias in the attribution of intention was also present by grades in the levels. Results showed that the overall effect was greater for 4th grade than 3rd grade, (1.57 vs 1.04, F1[1,51] = 4.22, p<.05, ηp^2=.08; F2(1,22) = 5.81, p<.05, ηp^2=.21), in level 2 and also for 6th grade than for 5th grade, 1.84 vs 1.57, F1[1,56] = 4.52, p<.05, ηp^2=.08; F2(1,22) = 12.68, p<.01, ηp^2=.37.

Sarcasm attributed by the addressee's interpretation. Analysis.

The same analysis was made for the sarcasm attributed to the addressee's interpretation. The results are shown in Table 4. The analysis of results showed the illusory transparency effect: more sarcasm was attributed when the polarity of the event was negative than when it was positive, F1 (1,107) = 28.86, p<.001, ηp^2 = .21; F2(1,22) = 32.21, p<.001, ηp^2=.60. There was no interaction of polarity and student's level, F1 (1,107) = 3.01, p>.09; ηp^2=.03; F2(1,22) = 2.54, p >.05, ηp^2=.10. No other significant effect was found.

Table 4. Degree of sarcasm (from 1 to -1) attributed to the addressee's interpretation according to the reader of the message, addressee or accidental reader, and the polarity of the event, positive or negative. Standard deviations between parentheses. 2nd level (3rd 4th graders) 3rd level (5th 6th graders).

		Positive Version	Negative Version
Addressee	2nd level	- 0.61 (0.56)	- 0.22 (0.68)
	3rd level	- 0.59 (0.61)	- 0.40 (0.67)
	Total	- 0.60 (0.59)	- 0.31 (0.68)
Accidental reader	2nd level	- 0.59 (0.59)	- 0.14 (0.70)
	3rd level	- 0.60 (0.61)	-0.35 (0.69)
	Total	- 0.60 (0.60)	-0.25 (0.70)

Knowledge of the reader about the experience of the writer. Analysis.

The participants' answers to the knowledge attribution to the reader were coded with scores of "0" (meaning that the reader cannot know the writer's experience) and "1" (indicating reader capacity to know the writer's experience). The results were submitted to an ANOVA 2 Reader (addressee or accidental reader) x 2 polarity of the event (positive or negative) x 2 School levels (2nd or 3rd levels) with the first two factors manipulated within-participants and the third between groups. The results are shown in Table 5. The analysis revealed a main effect of polarity of the event: greater capacity of interpretation was attributed to the reader when the experience was positive than negative, 0.60 vs. 0.36, $F1(1,107) = 63.62$, $p<.001$, $\eta p^2=.37$; $F2(1,22) = 77.64$, $p<.001$, $\eta p^2=.78$. Also, greater capacity of knowledge was attributed to the addressee than to the accidental reader; 0.51 vs 0.45, $F1(1,107) = 8.12$, $p<.05$, $\eta p^2=.07$; $F2(1,22) = 4.63$, $p<.05$, $\eta p^2=.17$. There was no effect of children's level, $F1(1,107) = 0.14$, $p>.71$, $\eta p^2< .01$; $F2(1,22) = 0.87$, $p >.05$, $\eta p^2=.04$, nor any significant interaction, Polarity x Level $F1(1,107) = 1.52$, $p>.22$, $\eta p^2=.01$; $F2(1,22) = 4.11$, $p >.05$, $\eta p^2=.60$; and Reader x Level $F1(1,107) = 1.81$, $p>.18$, $\eta p^2= .02$; $F2 (1,22) = 1.77$, $p >.05$, $\eta p^2=.07$.

Table 5. Capacity of interpretation attributed to the reader of the message (from 0 to 1) according to the reader: addressee or accidental reader, and the polarity of the event: positive or negative. Standard deviations between parentheses. 2nd level (3rd 4th graders) 3rd level (5th 6th graders).

		Positive Version	Negative Version
Addressee	2nd level	0.65 (0.33)	0.37 (0.32)
	3rd level	0.60 (0.35)	0.40 (0.34)
	Total	0.63 (0.34)	0.39 (0.33)
Accidental reader	2nd level	0.56 (0.34)	0.29 (0.27)
	3rd level	0.57 (0.41)	0.37 (0.40)
	Total	0.57 (0.38)	0.33(0.34)

In order to test whether these conditions were significantly different from zero, we submitted the data to a simple t-test. The analyses showed that participants considered that the addressee knew what actually happened in all conditions: in positive, $t(108) = 19.16$, $p < .001$, $d = 1.84$, and negative, $t(108) = 12.22$, $p < .001$, $d = 1.17$, conditions when the reader is the addressee, and in positive, $t(108) = 15.73$, $p < .001$, $d = 1.51$, and negative, $t(108) = 10.03$, $p < .001$, $d = .96$, conditions when the receiver is an accidental reader.

Discussion

Consider the following story, similar to those used in Keysar's task: a girl finds a dog in the street and tries to pet it. Suddenly the dog growls. The girl, who is talking on the phone with her friend, says, "it's a very loving dog!" The girl is trying to be sarcastic. Obviously, the girl knows that her friend has heard the dog growling, otherwise her friend could not interpret her message as sarcastic. However, in a situation in which the friend does not know about the growling, that is, in a situation in which the girl's friend cannot detect the incongruence between what the girl says and what she means (for example, if the two friends were texting on the phone), Keysar showed that adult readers of that kind of story still (1) attribute intention to be sarcastic to the girl and still (2) affirm that the friend will interpret the message as sarcastic. The second aspect was called illusory transparency of intention. In this paper, we have explored the illusory transparency phenomenon in childhood with Keysar's task.

The most remarkable result in this study is that children aged between 8 and 13 years showed illusory transparency of intention with the "sarcasm" task, replicating other results found with adults (e.g., Keysar, 1994; Moreno-Ríos et al., 2011). That is, children failed to take the character's perspective. In our example, children thought that the friend would interpret the message (lovely dog!) as sarcastic in a situation where the girls were texting on the phone, although she did not know what had happened when the girl met the dog. Specifically, children attributed more sarcasm to the reader's interpretation of the message when the polarity of the event was negative (the dog growled) than when it was positive (in another version of the story, in which the dog was wagging its tail). The same occurred with the attribution of intention to the writer (the intention of the girl in the previous example): children attributed more sarcasm to the intention of the writer when the event was negative than when it was positive. In addition, as occurred in Keysar's study (Experiment 2) with adults, children attributed more sarcasm to the intention of the writer (the girl's intention) than to the addressee's interpretation (the friend's interpretation).

The second finding is that older children attributed a stronger intention to be sarcastic to the writer of the message (the girl was trying to be sarcastic). Previous studies with younger children have shown an increase in the ability to detect sarcasm from the age of six years (Burnett, 2015; Creusere, 2000; Dews et al., 1996; Glenwright & Pexman, 2010; Harris & Pexman, 2003; Nilsen et al., 2011; Pexman et al., 2011). When sarcasm is used in a written utterance, the addressee of a message does not have clues, such as sarcastic prosody and/or a specific facial expression to detect the sarcasm. Therefore, using sarcasm in this situation has two requirements: 1) a negative experience is followed by a positive message in a context in which 2) *the addressee of the message knows of the negative experience*. Younger children seem to fulfil the first requirement, but not the second. Interestingly, Nilsen and col. (2011), using a different task, found that only older children (from 8 to 10) and not younger ones (from 6 to 7) were sensitive to the knowledge of the addressee about the experience: children attributed more ironic intention when the addressee knew the negative situation (the friend knew that the dog growled) than when she did not know. Adults also showed this effect with a greater appreciation of this feature. In our task, as in Keysar's task, there is an important difference from traditional sarcasm tasks: stories satisfy the first requirement of sarcasm (negative story + affirmative message), but not the second (the friend did not know that the dog growled), and therefore, participants have to mentally "fill" that requirement to interpret sarcasm. Older

children did this more frequently. What is the reason? The same two explanations tested in this study for "illusory transparency of intention" attributed to the addressee can operate in the attribution of sarcastic intent to the writer: Older children have a better understanding of sarcasm. Thus, they could be more prone to activating the sarcastic schemata (that is, they are more prone to attributing sarcasm when they detect inconsistency between what the speaker says and what she/he means) based on their pragmatic assumption (if the girl wrote the positive message from the negative situation, it was because she knew that her friend would know what really happened) or just that older children reconstructed the positive message as negative more easily (the girl actually said that the dog was not nice). The manipulation to disentangle the two possible explanations was made only in the interpretation question, as discussed below. However, we would expect the same mechanism to operate as a part of processing the attribution of intention and the attribution of interpretation.

In contrast, the tendency to interpret a more sarcastic intention in the writer (in the girl) did not translate into a significantly greater sarcasm attributed to the reader (the friend interpreted the message as sarcastic). According to Keysar, the bias reflects the extent to which "the privileged information about an event causes subjects to attribute the speaker's intention to the addressee" (Keysar, 1994; p. 172). Consequently, it would be expected that those participants who attributed more sarcasm to the sender's intention, would attribute more sarcasm to the addressee's interpretation. However, illusory transparency was shown by younger and older children without a significant change in the magnitude of the bias. In any case, what is clear is that there is no reduction in the bias with age. Other egocentric biases, such as "the course of knowledge", are reduced during childhood, at least in the early years. Older children tend to consider that the other person could not know what they know (e.g. Birch & Bloom, 2003). Something similar happens with the hindsight bias (Bernstein et al., 2011). Epley et al. (2004), proposed the existence of a common mechanism between the illusory transparency phenomenon and hindsight bias: the anchoring and adjustment heuristic. However, they did not test it empirically. From a different approach (see Blank, Nestler, von Collani & Fischer, 2008; Nestler, Blank & Egloff, 2010), hindsight bias has been analyzed in components: one of them, memory distortion, is based on the difficulty people have leaving aside new knowledge when they need to access a previous state of knowledge. The distortion in accessing previous information is shown at around 9 years old, but not later, demonstrating a decrease in older age groups. Other strategies, such as the reconstruction

bias, seem to remain constant, without developmental changes after the age of 8 (Pohl et al., 2010). The reconstruction bias in hindsight bias maintains some resemblance to the "construal" explanation in the illusory bias: the initial information is changed, and a different state of facts is recovered. If the illusory transparency effect is related to the memory distortion component of the hindsight bias, looking at the developmental trend without a reduction in the magnitude of the biases, it seems more likely to be related to the reconstruction bias rather than to distortion of access.

However, illusory transparency of intention differs from the other biases, not only in the sarcastic interpretation during perspective-taking, but also in that the tasks are more cognitively demanding and imply cognitive competences that develop during childhood, such as working memory skills (Filippova & Astington, 2008) and theory-of-mind skills (Nilsen et al., 2011; Pexman & Glenwright, 2007). Children can only make errors in the attribution of knowledge if they have access to them. That could be the reason for the absence of a reduction of this kind of bias.

Another objective in this study was to contrast the naïve-Pragmatic account with the construal shown in Table 1. We did this using two different measures: first, an indirect measure (traditional measure), computing the "bias" effect after asking whether the reader interpreted the message as sincere/sarcastic. From these responses, we could indirectly infer whether participants attributed the knowledge about what really happened to the characters. And, second, a new and additional measure was included in the present study: we asked participants directly about the privileged knowledge, whether the reader (friend of the girl) knew about the experience with the dog (direct measure).

The indirect measure (traditional effect) provide counter-evidence to the pragmatic principle implication, and therefore, evidence of the presence of construal operations: unlike adults (Moreno-Ríos et al., 2011), children did not attribute less sarcasm to the reader of the message's interpretation when he was an accidental reader, that is, the illusory transparency effect did not disappear when the reader was an accidental reader. From the construal account, the same process of modification of knowledge is expected from the addressee as from the accidental reader in the negative condition. It is assumed that participants will try to create an unambiguous consistent representation from the ambiguous message. Only the naïve account predicts differences between the real addressee and the accidental reader. Because the accidental reader does not know the writer, the participants cannot assume that they are trying to be collaborative. Therefore, no sarcasm is predicted for the accidental reader.

However, the opposite happened when participants were asked about the knowledge of the characters (direct measure): they asserted that the real addressee would have more access to the negative experience than would the accidental reader according to the naïve account, but against the construal view. In the negative condition with an accidental reader, the writer and the addressee do not know each other, and therefore, they do not have "shared" knowledge. So these results would support the idea that the attribution of sarcasm was produced because readers knew that the author and addressee of the message were known to each other, and would maintain the pragmatic principles of communication. Readers interpreted that the message could not be "ambiguous" and, therefore, the addressee would have the additional information needed to understand the message completely.

Therefore, the implicit (indirect-traditional) measure of the knowledge supports the construal account with no differences between the addressee and the accidental reader while the explicit response to the knowledge reflects the differences between the two conditions, supporting the naïve account. Two ways of processing that could relate to the difference between automatic (system I) and controlled (system II) processes, are widely discussed in reasoning literature (e.g. see Evans & Stanovich, 2013; Kahneman, 2011; Khemlani & Johnson-Laird, 2012).

In order to maintain this classical task, implicit-questions about interpretation (was the message interpreted as sincere?) were always presented at the beginning, followed by the explicit-knowledge-question (did the addressee know what happened?). One potential problem in this study was to know whether always giving the explicit question at the end meant it could be influenced by the previous questions. We must be cautious in interpreting the results, but there is one compelling result that allows us to make conclusions: previous questions did not "normalise" later questions. Actually, the results are systematically different in the conditions from the previous question: participants gave the same response to the sarcastic interpretation for addressee and accidental addressee, but they gave a different response in the two conditions in the later questions; only addressees and not accidental addressees could know what happened. We think that one key factor in the differences in results with the accidental reader between adults in previous experiments and children is their respective experience about what people can know about others. Their increasing knowledge of what is known can make adults more sensitive to applying pragmatic principles even in automatic processing.

To summarise, the present study shows the presence of the illusory transparency effect in childhood, along with other egocentric cognitive biases. In addition, we tested the two main explanations offered for the phenomenon. Our results suggest that in childhood the naïve theory principles were operating when children were explicitly asked about reader's knowledge (reader's capacity of knowledge question) since they attributed more knowledge to an addressee than to an accidental addressee. However, when they were implicitly asked about reader's knowledge (addressee's interpretation question) there were not significant differences in sarcasm attributed to the addressee or an accidental reader. This result linked to the fact that they significantly attributed to addressee capacity to know speaker's experience in all conditions (including negative conditions), suggests that in childhood both naïve principles as well as "construal" operations could be operating in illusory transparency of intention but each of those in a different way. Finally, the tendency to infer a speaker's sarcastic intention has also been shown when the receiver did not know the facts. It increases during these developmental years but this better ability was not linked to a significant change in the bias magnitude of illusory transparency of intention.

Appendix

Negative version-real addressee

As they do every day, Pedro and Natalia speak animatedly in the school playground during the break. On this occasion, they're talking about the new maths teacher. The children say that he's a rather ugly man, they understand nothing he has explained today in class and, as if this weren't bad enough, he gives them a lot of homework. At this moment Andrés, a schoolmate who is going to have lessons with the new teacher, arrives. Andrés asks Natalia and Pedro:

- What is the new maths teacher like?
- Natalia answers: Excellent, just excellent

Positive version-real addressee

As they do every day, Pedro and Natalia speak animatedly in the school playground during the break. On this occasion, they're talking about the new maths teacher. The children say that he's a very likeable man, and moreover, they have understood all his explanations and he gives them almost no homework. At this moment Andrés, a schoolmate who is going to have lessons with the new teacher, arrives. Andrés asks Natalia and Pedro:

- What is the new maths teacher like?
- Natalia answers: Excellent, just excellent

Negative version-accidental addressee

As they do every day, Pedro and Natalia speak animatedly in the school playground during the break. On this occasion, they're talking about the new maths teacher. The children say that he's a rather ugly man, they understand nothing he has explained today in class and, as if this weren't bad enough, he gives them a lot of homework.

Andrés, a schoolmate of Pedro and Natalia, wants to know what the new teacher is like. He intended to ask their schoolmate during the break but he was being kept in as a punishment. For this reason, he asks Pablo, a friend, to ask Natalia and Pedro what the new teacher is like.

- Hi guys! I'm Pablo, one of Andres' friends, he's sent me to ask you what the new maths teacher is like because he was being punished and couldn't go out during the break.
- Natalia answers: Tell Andrés that he's excellent, just excellent!

Positive version-accidental addressee

As usually, Pedro and Natalia speak animatedly in the school playground during the break. On this occasion, they're talking about the new maths teacher. The children say that he's a very likeable man, they have understood all his explanations and he gives them almost no homework.

Andrés, a schoolmate of Pedro and Natalia wants to know what the new teacher is like. He intended to ask their schoolmate during the break but he was being kept in as a punishment. For this reason, he asks Pablo, a friend, to ask Natalia and Pedro what the new teacher is like.

- Hi guys! I'm Pablo, one of Andres' friends, he's sent me to ask you what the new maths teacher is like because he was being punished and couldn't go out during the break.
- Natalia answers: Tell Andrés that he's excellent, just excellent!

- *Questions.*

1. Is Natalia sincere when she says the new maths teacher is excellent?
2. Will Andres think Natalia is sincere when she says the new maths teacher is excellent?
3. Will Pablo, Andres' friend, think Natalia is sincere when she says the new maths teacher is excellent? (Accidental)
4. Have Pedro and Natalia understood the teacher's explanations?
5. Does the new teacher teach physical education?
6. Can Andres know what the new maths teacher's lessons are like?
7. Can Pablo, Andres' friend, know what the new maths teacher's lessons are like? (Accidental)

Capítulo 3

"I know that you know because it was known": a link between illusory transparency and hindsight bias

Abstract

We report on three experiments investigating the common role of causal attribution processes in two epistemic biases: hindsight bias and illusory transparency. Based on hindsight bias literature, we propose an alternative explanation for illusory transparency bias that involves the actuation of causal reasoning processes. To test this hypothesis, an outcome was included in the traditional illusory transparency task. Results show that participants exhibit illusory transparency and hindsight bias in the same task and that a positive relation exists between the magnitudes of the two biases (Experiment 1). In addition, as occurred in hindsight bias studies, we found that the degree of illusory transparency is sensitive to the participant's capacity to build a causal model event (Experiments 2 and 3) as well as to the strength of the perceived causal link (Experiment 3). Taken together, these results support the hypothesis that causal attribution processes are acting in the illusory transparency bias.

Key words: epistemic biases; illusory transparency; hindsight bias; common mechanism; causal attribution.

Pre-print version of the manuscript: Gordo, C. & Moreno-Ríos, S. (2021). "I know that you know because it was known": a link between illusory transparency and hindsight bias. Manuscript submitted for publication.

Introduction

Imagine that you are watching a typical TV quiz show. It is the final round. If a contestant answers the question correctly, then she wins a million euros. If the contestant fails, she loses everything. The host abruptly asks the question:

- What is the capital of Finland?
- She knows! you immediately think, clearly envisaging the contestant winning a million euros.
- Oslo, the capital of Finland is Oslo, the contestant replies.
- What? How could she fail? It was so easy… Helsinki! you say.

Why did you think the contestant would know? Maybe, just because *you* know. A large body of research has shown how individuals are unable to discard their own knowledge when they reason about others' knowledge (e.g., Bernstein et al., 2018; Birch & Bloom, 2003; Fischhoff, 1975; Gordo & Moreno-Ríos, 2019; Higham et al., 2017; Keysar, 1994; Moreno-Ríos et al., 2011). The previous example illustrates the effect of an egocentric bias known as illusory transparency of intention (Keysar, 1994). Illusory transparency bias leads people to attribute their own privileged knowledge about an event (for example, the correct answer to the question) to another individual (e.g., the contestant) who does not have this information.

The illusory transparency effect has been related to other egocentric biases, such as the hindsight bias (Fischhoff, 1975), that reflects the effect of outcome knowledge in participants' judgements. In the previous TV programme example, knowledge about the contestant's failure would lead us to perceive the unfortunate result as more predictable and/or more inevitable than it would be in foresight. From the social cognition perspective, illusory transparency and hindsight bias have been considered as epistemic biases (for details, see Royzman et al., 2003). Some authors (e.g., Epley et al., 2004; Nickerson, 1999; Royzman et al., 2003) have proposed the existence of a common mechanism for these epistemic biases: the anchoring and adjustment heuristic (Tversky & Kanheman, 1974). Epley and cols. (2004), assuming the existence of such a common mechanism, assessed the effect of illusory transparency and hindsight bias in perspective-taking in two independent studies. Results showed that both biases hinder people's ability to adopt others' perspectives. In this work, we aim to explore whether, as these authors argue, there is a common mechanism in illusory transparency and hindsight bias. More specifically, we will test whether causal attribution is a common mechanism in these two biases.

Although traditionally, hindsight bias has been treated as a unitary phenomenon, it is currently assumed that at least three different hindsight components or manifestations can be distinguished ("separate components view"): impressions of inevitability, impressions of foreseeability, and memory distortions (for details, see Blank et al., 2008; Nestler et al., 2010). It has been shown that one of the hindsight bias components, impressions of necessity or creeping determinism (see Blank et al., 2008; Nestler et al., 2010 for details), is a product of causal reasoning processes (e.g., Nestler et al., 2008a; Nestler et al., 2008b; Nestler & von Collani, 2008; Oeberst et al., 2014; Roese & Olson, 1996; Yopnick & Klim, 2012). In the present work, we will focus on the manifestation of creeping determinism.

Creeping determinism

Creeping determinism leads individuals to perceive an outcome as more inevitable, more probable and/or more causally determined after being informed about it. In the previous TV show example, when we watch the TV programme again, our knowledge about the participant's error would lead us to claim: "it could not have been otherwise", contrary to what we thought the first time we saw the show. Typically, creeping determinism has been investigated using historical events. In Fischhoff's study (1975), for example, participants received information about the war between the British and the Gurkhas in the 19th century and were asked to make probability judgements about several alternative outcomes (e.g., British victory). The scenario included antecedents that spoke in favour of a British victory and of a Gurkha victory (e.g., there were only 12,000 Gurkhas). Before making their judgements, only half the participants were informed about the outcome (e.g., British victory). The results showed that participants who received the outcome information judged that outcome as more probable than participants who did not receive any outcome information.

Previous creeping determinism studies have found that the presence of the bias and its magnitude depend on the causal information available in the event description (see for example, Nestler et al., 2008a; Oeberst et al., 2014; Wasserman et al., 1991). In this regard, Yopchick and Kim (2012) found that the bias only appeared when participants were provided with relevant causal information as to the outcome. These authors asked participants to read scenarios that described, for example, a battle between the Hutus and the Tutsi. In one of the hindsight conditions, these scenarios included relevant causal information (e.g., the Hutus had superior troop discipline) for the outcome provided (e.g., the Hutu won the battle), whereas in the other condition, the causal information was

irrelevant to the outcome (e.g., the Hutus beginning the day by travelling west). In addition to these conditions, participants completed two foresight conditions that included information that was either relevant or irrelevant. The participants' task was to rate how likely each outcome (e.g., the Hutus victory and the Tutsi victory) was. Results showed that participants judged both outcomes equally likely in the hindsight condition that included irrelevant information. The effect of causal attribution that influences creeping determinism could also be playing a role in the illusory transparency bias.

Illusory transparency

Illusory transparency has been investigated through stories that involve the interaction between two characters (see for example: Gordo & Moreno-Ríos, 2019; Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009). In one of them, Julia recommends to her colleague, David, a restaurant to go to that night. David goes to the restaurant and has an experience. Next day, David cannot find Julia in her office and he leaves her the following message: "*You wanted to know about the restaurant? Well, it was marvellous, just marvellous*" (Keysar, 1994, p. 173). In one condition, David's experience is described as positive ("*the food was delicious and the service was impeccable*"), whereas in the other condition it is described as negative ("*the food was revolting, and the service was mediocre*"). The participants' task is to predict whether David has the intention of being sarcastic. Typically, they attribute more intention of sarcasm to David (the writer) in the negative condition than in the positive one (e.g., Epley et al., 2004; Gordo & Moreno-Ríos, 2019; Keysar, 1994; Moreno-Ríos et al., 2011). Additionally, participants are asked to predict whether Julia will interpret the message as sarcastic or not. Note that in both conditions the information about the experience was inaccessible to Julia, the addressee. Thus, if participants were able to ignore their own knowledge about the experience, they should attribute to the addressee (Julia) a literal interpretation (a sincere interpretation) in both conditions (see Figure 1A). However, participants attribute more sarcasm to the addressee's interpretation when the experience is described as negative than when it is positive (Epley et al., 2004; Gerrig et al., 2000; Gordo & Moreno-Ríos, 2019, Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009), showing the so-called "illusory transparency of intention" bias.

Despite being a robust phenomenon, it remains unclear what mechanisms underlie the illusory transparency effect. Keysar (1994) defended the theory that the bias arises from the reader's illusion that David's intentions (writer's intentions) are transparent to

Julia (addressee). Namely, readers would attribute to Julia the knowledge about David's communicative intentions (sincere or sarcastic) in spite of Julia's lack of access to the key information (David's experience at the restaurant) –see Figure 1B-. In a later publication this account was shown less clearly (see Keysar, 2000 for details), but the term "illusion of intention" became general (see for example, Gerrig, et al., 2000; Gordo & Moreno-Ríos, 2019; Weingartner & Klin, 2005, 2009). In the present work, we propose an alternative explanation, where the bias arises from readers' causal reasoning processing.

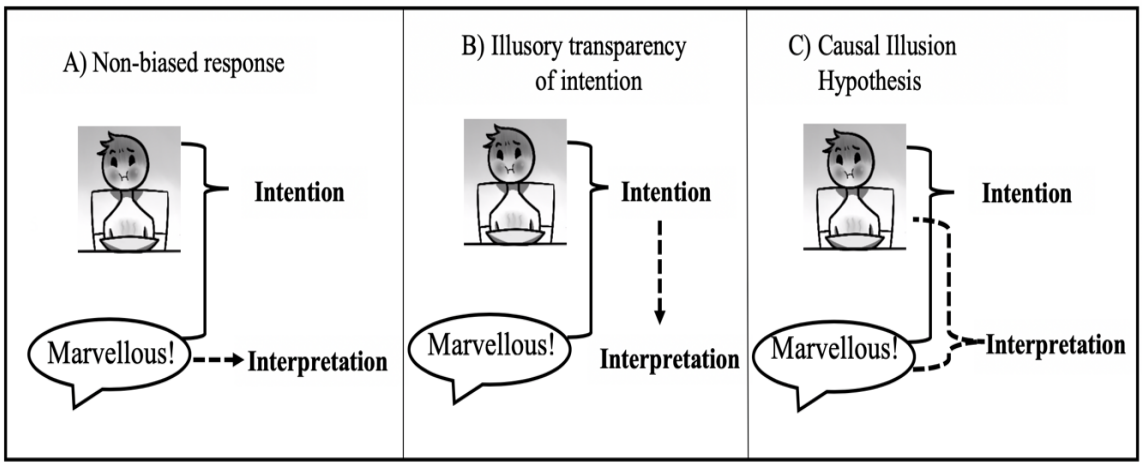
The traditional illusory transparency task involves the comprehension of sarcasm. Sarcasm comprehension requires understanding that the meaning of an utterance is the opposite of what is stated, by detecting incongruity between the event (i.e. David's bad experience in the restaurant) and the statement –i.e. David's positive message: *"Well, it was marvellous, just marvellous*" - (e.g. Matsui et al. 2016). In this way, individuals, as they develop, learn an automatic rule: a positive message preceded by a negative experience is interpreted as sarcastic (Burnett, 2015; Creusere, 2000; Dews et al., 1996; Glenwright & Pexman, 2010; Harris & Pexman, 2003; Pexman et al., 2011). In the traditional illusory transparency task, participants activate that pragmatic rule to attribute a sincere intention to the writer in the positive condition and a sarcastic intention in the negative one (e.g., Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009). Similarly, participants who exhibit the bias seem to follow this rule to predict the addressee's interpretation: they state that the addressee will interpret the message as sarcastic in the negative condition, without considering what the addressee knows. In illusory transparency stories, the addressee is always ignorant of the event (the experience), and therefore, unable to detect the incongruity between the event (restaurant experience) and the statement (positive message). Why then do biased participants predict a sarcastic interpretation? Our hypothesis is that these participants create an illusory causal link between the writer's experience and the addressee's interpretation to conclude that the interpretation is sarcastic. It is illusory because the addressee does not know the writer's experience and therefore, a sarcastic interpretation cannot follow. Thus, participants would detect an incongruence between the experience and the message that leads them to attribute sarcasm to the addressee's interpretation (see Figure 1C). This causal link could only be shown in the negative condition, where the experience-message relation changes the interpretation of the message from positive to negative. Hereinafter, we refer to this idea as *the causal illusion hypothesis.* Note that this hypothesis assumes the illusory transparency effect is independent of the perception of sarcasm in the writer's

intention, unlike in the original proposal (Keysar, 1994). Participants would build a causal event model where the addressee's interpretation is directly determined by the experience of the writer. That is, where a negative writer's experience is followed by a sarcastic addressee's interpretation (and a positive experience by a sincere one) - "*the hypothesised causal model built by biased participants*-.

Experiment 1

In order to test *the causal illusion hypothesis*, we adapted the traditional illusory task, adding to the traditional scenarios a final sentence. In this sentence, participants were informed that the addressee interpreted the message as sincere when the event was positive but as sarcastic when it was negative. Note that the informed addressee's interpretation (the outcome) is always coherent with the hypothesised causal model built by biased participants. Participants completed traditional scenarios (without outcome information) and outcome scenarios. Half the participants received traditional scenarios first and then outcome scenarios while the other half received them in the opposite order. After each scenario, participants were asked to determine whether the writer's intention was to be sarcastic and whether the addressee would interpret the message as sarcastic.

Figure 1. Sources for attribution of intention (solid lines) and attribution of interpretation (dotted lines) in (A) a non-biased response, (B) Keysar's original proposal and (B) the causal illusion hypothesis.



Based on previous illusory transparency studies (e.g., Gerrig et al., 2000; Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009), we will expect that participants attribute more sarcasm to the addressee's interpretation in the negative traditional condition than in the positive one, showing the traditional illusory transparency effect. With respect to creeping determinism, as we mentioned above, it only appears when participants are informed about an outcome and the event description includes causal information to explain such an outcome (Yopchick & Kim, 2012; Wasserman et al., 1991). Following these findings, only if biased participants activate causal processing as we hypothesised, will the creeping determinism appear in the negative conditions. That is, only if participants perceive the writer's experience as relevant to predicting the addressee's interpretation, will they attribute more sarcasm to the addressee's interpretation in the negative outcome condition than in the traditional one. The presence of the outcome would make participants perceive the sarcastic addressee's interpretation as more inevitable ("it could not be otherwise") than when no outcome is given. Moreover, if causal attribution processes are acting in the illusory transparency effect in the same way as in creeping determinism, more sarcasm should be attributed to the addressee's interpretation in the outcome conditions than in the traditional ones.

With regard to the positive conditions, previous illusory transparency studies have reported that almost all participants attribute a sincere interpretation to the addressee (e.g., Gerrig et al., 2000; Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005, 2009). Therefore, we do not expect any effect of outcome knowledge. Similarly, if the causal illusion hypothesis is right, the inclusion of an outcome should not affect participants' judgements about the writer's intention because it is independent of the hypothesised causal link. Thus, participants will always attribute more sarcasm to the writer's intention in negative conditions than in positive ones, and their sarcasm attribution will not differ between outcome and traditional scenarios.

Additionally, the inclusion of an outcome in the traditional illusory task allows us to obtain two indices for each participant: an illusory transparency bias index and a creeping determinism index. It could happen that the two indices are not related, but if the causal illusory hypothesis is true and the same processes are acting in both biases, then the magnitudes of the biases should be positively related: those participants who show a higher illusory transparency bias will also exhibit a greater creeping determinism, because the presence of an outcome will reinforce the hypothesised causal link.

Method

Participants

Fifty-one undergraduate students (14 males and 37 females) from the University of Granada (age: *M* = 19.82; *SD* = 1.53) participated in the study in exchange for course credits. One participant from the group that completed the outcome scenarios first was excluded for answering more than three comprehension questions incorrectly. Therefore, our final sample consisted of fifty participants (13 males and 37 females). All participants were native speakers of Spanish. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the set of experiments.

As recommended by the journal, we computed sensitivity power analyses using G*Power3 (Faul et al., 2007) for the main predicted interaction effects. Because G*Power does not calculate interactions between repeated measures factors, we computed the sensitivity power analyses, treating one of the repeated measures factors as a between groups factor (note that this is a more conservative power analysis than our repeated measures design). The sample size of this experiment allowed us to detect an effect (i.e., the main hypothesis regarding the outcome knowledge × polarity of the experience interaction) as small as f = 0.279 ($\eta 2$= .072; $\alpha = 0.05$, $1 - \beta = 0.80$, the average correlation among all repeated measures; r =.05).

Materials

Twelve scenarios were constructed with the same structure as the traditional illusory transparency task (Keysar, 1994): a character has an experience that may be positive or negative. S/he then leaves a message that could be interpreted as either sincere or sarcastic depending on the character's experience (see Appendix 1). For example, in one scenario, Julia recommends a restaurant to her co-worker David. Next day David leaves a message to Julia about his experience at the restaurant: "*You wanted to know about the restaurant? Well, it was marvellous, just marvellous*". In one condition, David's message is preceded by a positive experience (positive version), whereas in the other condition it is preceded by a negative experience (negative version).

In addition to these two traditional illusory versions (positive and negative), we created two additional versions of each scenario. These versions included an outcome. That is, they ended with an utterance that informed participants about the addressee's interpretation of the message (see Table 1). In one version, after David's positive experience, participants read that Julia had interpreted the message as sincere (positive-sincere version). In the other version, David's negative experience is followed by Julia's

sarcastic interpretation (negative-sarcastic version). Thus, we created four versions of each scenario: a positive (traditional) version, a negative (traditional) version, a positive-sincere (outcome) version and a negative-sarcastic (outcome) version.

Table 1. Structure of text in the four conditions. Differences between them are in italics.

Conditions	**Text structure**
Positive version	Julia recommends a restaurant to her partner David. David has a *good experience* in the restaurant. David writes to Julia: ‘‘the restaurant was marvellous, just marvellous’’. Julia reads the message.
Negative version	Julia recommends a restaurant to her partner David. David has a *bad experience* in the restaurant. David writes to Julia: ‘‘the restaurant was marvellous, just marvellous’’. Julia reads the message.
Positive-Sincere version	Julia recommends a restaurant to her partner David. David has a *good experience* in the restaurant. David writes to Julia: ‘‘the restaurant was marvellous, just marvellous’’. Julia reads the message. *Participants read that Julia interpreted the message as sincere.*
Negative-Sarcastic version	Julia recommends a restaurant to her partner David. David has a *bad experience* in the restaurant. David writes to Julia: ‘‘the restaurant was marvellous, just marvellous’’. Julia reads the message. *Participants read that Julia interpreted the message as sarcastic.*

Eight different booklets were constructed. Each booklet had two blocks of scenarios: a traditional scenarios block and an outcome scenarios block. Each block was composed of six scenarios: six traditional scenarios (three positive and three negative) and six outcome scenarios (three positive-sincere and three negative-sarcastic). There was nothing in the booklets that could enable participants to distinguish between the two blocks. In half of the booklets, the traditional scenarios appeared first, whereas in the other half, the outcome scenarios were presented at the beginning. The order and the content of the scenarios were counterbalanced.

After each scenario and on the following page, four questions relating to the story appeared. These questions asked participants to indicate whether the writer of the message intended to be sarcastic (e.g., Does David intend to be sarcastic?) and whether the addressee would interpret the message as sarcastic. The wording of this last question differed slightly between the traditional and the outcome scenarios. In the traditional scenarios, participants were asked “Will Julia interpret David’s message as sarcastic?”,

whereas in the outcome scenarios, the question wording was as follows: "If you didn't know how Julia interpreted the message, would you expect a sarcastic interpretation?" In both kinds of scenario, participants answered the writer's intention and the addressee's interpretation questions by marking "Yes" or "No". The remaining two questions were comprehension questions (e.g., Are David and Julia co-workers?).

Procedure

Fifty-one participants were evaluated in groups in a silent class at the University of Granada. They were instructed to stay quiet at the end of the task until everybody had finished. Every participant received a booklet. The booklet began with one-page instructions on how to carry out the task. The instruction page also explained the objective of the study. However, the real objective was hidden to avoid any possible effect on their responses. Instead of revealing this, participants were told that the purpose of the study was to evaluate their reading comprehension. On the following pages, the twelve scenarios were presented one by one on separate pages. After each scenario, and on the following page, the four questions appeared: the two illusory transparency questions (writer intention and addressee's interpretation) and two comprehension questions.

Results

Following Epley et al. (2004) and Moreno-Ríos et al., (2011), participants' responses were coded with scores of -1 for "no" answers (indicating sincerity, the absence of sarcasm) and 1 for "yes" answers (meaning intention or interpretation of sarcasm). In every condition, an index of sarcasm (from 1 to -1) was calculated for each participant. The more positive the index, the more sarcasm it reflected. The averages of participants' responses to writer's intention and addressee's interpretation questions are shown in Table 2.

Table 2. Degree of Sarcasm (from 1 to -1) attributed to the writer's intention and the addressee's interpretation, depending on the polarity of the experience and the outcome knowledge (traditional-no outcome- and outcome).

		Traditional	Outcome
Intention	Positive	-0.93 (0.24)	-0.97 (0.13)
	Negative	0.77 (0.37)	0.85 (0.36)
Interpretation	Positive	-0.75 (0.42)	-0.64 (0.50)
	Negative	-0.47 (0.59)	0.00 (0.70)

Sarcasm from the Intention of the Writer

The averages of sarcasm attribution to the writer's intention question were submitted to an ANOVA 2 (outcome knowledge: traditional, outcome) x 2 (polarity of the experience: positive, negative) x 2 (order: 1st traditional, 2nd traditional) with the two first factors manipulated within participants and the third between groups.

The analysis yielded a main effect of the polarity of the experience factor, $F(1, 48) = 1876.03$, $p < .001$; $\eta 2 = .98$: more sarcasm was attributed to the writer's intention when the experience was negative than when it was positive. However, neither the outcome knowledge factor, $F(1, 48) = 0.33$, $p = .570$; $\eta 2 = .01$, nor the order of presentation factor, $F(1, 48) = 2.05$, $p = .158$; $\eta 2 = .04$, were significant. No significant interaction was found (outcome knowledge x order: $F(1, 48) = 1.51$, $p = .226$; $\eta 2 = .03$; polarity of the experience x order: $F(1, 48) = .33$, $p = .569$; $\eta 2 = .01$; outcome knowledge x polarity of the experience: $F(1, 48) = 2.58$, $p = .115$; $\eta 2 = .05$; outcome knowledge x polarity of the experience x order: $F(1, 48) = 2.58$, $p = .115$; $\eta 2 = .05$.)

Sarcasm attributed to the addressee's interpretation

The averages of sarcasm attribution to the addressee's interpretation were submitted to the same ANOVA as the writer's intention responses. The results of this analysis revealed a main effect of the outcome knowledge factor $F(1, 48) = 15.51$, $p < .001$; $\eta 2 = .24$, showing that participants' sarcasm attributions depended on whether the result was provided. *This is the general effect of outcome knowledge.* In addition, the polarity of the event factor was also significant $F(1, 48) = 37.80$, $p < .001$; $\eta 2 = .44$: participants attributed more sarcasm to the addressee's interpretation when the polarity of the experience was negative than when it was positive. This is the *illusory transparency of intention effect*. The interaction between the outcome knowledge and the polarity of the event was also significant, $F(1, 48) = 5.12$, $p = .028$; $\eta 2 = .10$. Further analysis showed that although the illusory transparency effect was found in traditional, $F(1, 49) = 7.51$, $p = .009$; $\eta 2 = .13$, as well as outcome scenarios, $F(1, 49) = 29.74$, $p < .001$; $\eta 2 = .38$, the effect was higher in outcome scenarios. The illusory transparency effect was greater for outcome scenarios because participants attributed significantly more sarcasm to the addressee's interpretation in the negative outcome condition than in the traditional negative condition, $F(1, 49) = 13.57$, $p = .001$; $\eta 2 = .22$. This is the *creeping determinism effect* that we predicted. Participants' ratings in the positive conditions did not significantly differ, $F(1, 49) = 1.62$, $p = .209$; $\eta 2 = .03$. Finally, the analysis yielded an effect of the order of scenarios that did not reach significance, $F(1, 48) = 3.21$, $p = .079$;

η2= .06: participants who started the booklet by answering the outcome scenarios showed a higher tendency to attribute sarcasm to the addressee's interpretation than those who completed the traditional scenarios first. No other significant interactions were found (outcome knowledge x order: F (1, 48) = 0.10, p = .753; η2 < .01; polarity of the experience x order: F (1, 48) = 0.09, p = .755; η2 < .01; outcome knowledge x polarity of the experience x order: F (1, 48) = 0.83, p = .775; η2 < .01).

Correlations between bias indices

We computed indices for both biases. The illusory transparency index reflected the difference in participants' sarcasm attributions to the addressee's interpretation between positive and negative conditions in traditional scenarios. By contrast, creeping determinism was obtained in the negative conditions by subtracting participants' sarcasm attributions in the traditional condition from those in the negative outcome condition. The analysis revealed a positive correlation between the illusory transparency index and the creeping determinism index, r(49) = 0.60; p < .001, showing that participants who displayed a higher illusory transparency (*M* = -0.27, *SD* = 0.71) were those who exhibited a greater creeping determinism (*M* = 0.47, *SD* = 0.91).

Discussion of Experiment 1

Results with the adapted illusory transparency task replicate the main results in previous studies (Gerrig et al., 2000; Gordo & Moreno-Ríos, 2019; Keysar, 1994; Moreno-Ríos et al., 2011; Weingartner & Klin, 2005,2009) showing the effect of illusory transparency, and what is more interesting, showing that it is possible to use the same task to test creeping determinism. Although with both traditional and outcome scenarios, participants attributed more sarcasm to the addressee's interpretation in the negative condition, their sarcasm attributions were higher in the outcome condition. By contrast, as predicted, the inclusion of an outcome did not affect the judgements of the addressee's interpretation in the positive outcome condition.

In addition, as causal illusion predicts, inclusion of the information about how the addressee had interpreted the message did not have an effect on the participants' ratings for the writer's intention question: participants' sarcasm attributions (and sincerity attributions) did not differ between traditional and outcome scenarios. Finally, the magnitudes of the two biases were positively related: participants who exhibited a higher illusory transparency showed a greater creeping determinism.

These results are consistent with the causal illusion hypothesis. Firstly, creeping determinism appeared in the traditional illusory transparency task. Note that according to

Yopchick and Kim's (2012) findings, creeping determinism only appears when the information provided to participants has a high causal relevance. Thus, the presence of the bias in the negative condition suggests that participants perceive a causal relationship between the available antecedents (e.g. David's experience at the restaurant) and the provided outcome (e.g., Julia's message interpretation). Secondly, the inclusion of a sentence informing participants about the addressee's interpretation did not have an effect either on the participants' judgements about the addressee's interpretation in the positive outcome condition or on their judgements about the writer's intention. And thirdly, the magnitudes of the two biases were positively related, suggesting that the two biases could share a common mechanism: a causal attribution process. Although, as just mentioned, these results support the causal illusion hypothesis, we conducted a second experiment to investigate whether the illusory transparency bias, like the creeping determinism one (see for example: Nario & Branscombe, 1995; Nestler et al., 2008a; Nestler & von Collani, 2008), is sensitive to manipulations that alter participants' capacity to build a causal model about how things came about.

Experiment 2

Prior works have shown that the magnitude of creeping determinism depends on how well the available antecedents can explain the outcome. Thus, the better the possibility of explaining the outcome from the causal information available, the greater the creeping determinism (e.g., Nario & Branscombe, 1995; Nestler et al., 2008a; Nestler & von Collani, 2008; Roese & Olson, 1996; Oeberst et al., 2014; Yopnick & Kim, 2011). By contrast, when the available causal information makes it difficult to create a causal model of how things came about, the bias diminishes (e.g., Ash, 2009; Calvillo & Gomes, 2011; Nestler & Eggloff, 2009; Pezzo, 2003; Schkade & Kilbourne, 1991) or even disappears (Yopchick & Kim, 2012; Wasserman et al., 1991).

In terms of causal processing, we can see the traditional illusory task as an event with an antecedent: the writer's experience (bad food at the restaurant) and a positive message. From that antecedent, participants predict the consequent outcome: the addressee's interpretation (it was sarcastic), and assess to what extent the available antecedent can explain the provided outcome (the consequent). In Experiment 1, we tested the relation between illusory transparency and creeping determinism in an experimental context of strong causal relation. In Experiment 2 we aimed to investigate whether, as occurs in creeping determinism studies, the magnitude of illusory

transparency diminishes when participants' readiness to set up a coherent model about how things came about is hindered. To this end, we included two additional conditions in which experience and outcome are incoherent with the hypothesised causal model of the event built by biased participants. In one condition, after a negative experience, participants read that the addressee had interpreted the message as sincere (negative-sincere condition). For example, participants were informed that after having a bad experience in the restaurant, David wrote a message saying that the experience was great and Julia interpreted it as sincere. In the other condition, a sarcastic interpretation of the message was preceded by a positive experience (positive-sarcastic condition). Thus, participants completed scenarios from six different conditions: positive traditional, negative traditional, positive-sincere, negative-sarcastic, positive-sarcastic and negative-sincere. Note that in the two coherent conditions, experience and outcome are relevant to explain the outcome (positive-sincere and negative-sarcastic condition, whereas in the two incoherent conditions they are irrelevant (positive-sarcastic and negative-sincere).

As in Experiment 1, we predict that illusory transparency will be present in traditional and outcome scenarios with a coherent outcome (positive-sincere and negative-sarcastic). That is, we expect that in these scenarios participants will attribute more sarcasm to the addressee's interpretation in the negative conditions. However, if the construction of a causal model is a basic component in illusory transparency, as the *causal illusion hypothesis* establishes, the illusory transparency effect should disappear in the scenarios where participants' capacity to set up a model of how things came about is hindered. In particular, we do not expect differences in participants' sarcasm attribution to the addressee's interpretation between positive-sarcastic and negative-sincere conditions (the incoherent scenarios). As in Experiment 1, the inclusion of an outcome should not affect participants' judgements in the positive conditions that act as control conditions. Similarly, in this Experiment, and for the same reasons as in Experiment 1, the presence of an outcome should not affect participants' sarcasm attributions to the writer's intention.

Method

Participants

A total of forty-seven students (10 males and 37 females) from the University of Granada (age: M = 21.20; SD = 3.71) took part in Experiment 2. They participated in exchange for course credits. Three participants were removed from the sample for

answering more than three comprehension questions incorrectly. Therefore, our final sample consisted of forty-four participants (9 males and 35 females). According to the sensitivity power analysis, following the same strategy as in Experiment 1, the sample size of this experiment allowed us to detect an effect (i.e., the main hypothesis regarding the outcome knowledge × polarity of the experience interaction) as small as f= 0.258 ($\eta 2$= .062; $\alpha = 0.05$, $1 - \beta = 0.80$, r=0.06).

Materials

The same twelve scenarios used in Experiment 1 served as stimuli. In addition to the four versions used in Experiment 1, we created two new versions of each scenario that included outcome information. In one of them, after the positive experience, participants read that the addressee had interpreted the message as sarcastic. In the other version, the negative experience was followed by a sincere addressee's interpretation. Thus, there were six different versions of each scenario: two no-outcome versions (positive and negative – traditional-), two versions with a coherent outcome (positive-sincere and negative-sarcastic) and two versions with an incoherent outcome (positive-sarcastic and negative-sincere).

Therefore, six different booklets were constructed, each with twelve scenarios. Each booklet was composed of two scenarios from the 6 different versions (or conditions): 2 positive (traditional), 2 negative (traditional), 2 positive-sincere, 2 negative-sarcastic, 2 positive-sarcastic and 2 negative-sincere. As in Experiment 1, we counterbalanced the scenario versions, ensuring that each scenario was presented in all its versions. However, unlike in Experiment 1, we randomised the order of presentation of conditions in the booklets to prevent any effect of order.

Procedure

The procedure was the same as in Experiment 1. After each scenario, and on the following page, participants answered the same four questions as in Experiment 1.

Results

As in Experiment 1, we calculated an index of sarcasm for each participant for the two illusory transparency questions (intention and interpretation). Participants' responses to the writer's intention question and the addressee's interpretation question were submitted to an ANOVA 3 (outcome knowledge: traditional, coherent outcome, incoherent outcome) x 2 (polarity of the experience: positive, negative) with both factors manipulated within participants. Results are showed in Table 3.

Table 3. Degree of Sarcasm (from 1 to -1) attributed to the writer's intention and the addressee's interpretation depending on the polarity of the experience and the outcome provided (traditional, coherent, incoherent).

		Traditional	Coherent	Incoherent
Intention	Positive	-0.95 (0.21)	-0.91 (0.29)	-1.00 (0.00)
	Negative	0.77 (0.52)	0.80 (0.55)	0.82 (0.50)
Interpretation	Positive	-0.59 (0.62)	-0.52 (0.63)	-0.39 (0.78)
	Negative	-0.57 (0.59)	-0.02 (0.79)	-0.36 (0.72)

Sarcasm attributed to the writer's intention

The analysis yielded the polarity of the event factor as the only significant one, $F(1, 43) = 683.91$, $p < .001$, $\eta^2=.94$: participants attributed more sarcasm to the writer's intention when the experience was described as negative than when it was described as positive. Neither outcome factor, $F (2, 42) = 0.40$, $p = .672$, $\eta p^2= .02$, nor the interaction, $F (2, 42) = 0.68$, $p = .512$, $\eta^2=.03$, were significant.

Sarcasm attributed to addressee's interpretation

The results of this analysis revealed a main effect of outcome knowledge, $F (2, 42) = 6.31$, $p = .004$, $\eta^2=.23$: participants' sarcasm attribution depended on the kind of outcome provided. By contrast, the polarity of the experience factor was not significant, $F (1, 43) = 2.83$, $p =.100$, $\eta^2=.06$. The interaction between outcome knowledge and polarity of the experience was significant, $F (2, 42) = 3.87$, $p = .029$, $\eta^2=.16$. Further analyses showed that the *illusory transparency effect* was only significant in the coherent outcome scenarios, $F (1, 43) = 10.51$, $p = .002$, $\eta^2=.19$, but not in the incoherent, $F (1, 43) = 0.02$, $p =.901$, $\eta^2 = .01$, or traditional ones, $F (1, 43) = 0.03$, $p =.855$, $\eta^2 <.00$.

Regarding the *creeping determinism*, more sarcasm was attributed in the negative coherent outcome scenarios than in the incoherent, $F (1, 43) = 5.01$, $p =.030$, $\eta^2=.10$, or traditional, $F (1, 43) = 17.11$, $p <.001$, $\eta^2=.30$, ones. These differences were only shown in the negative conditions, as predicted (coherent vs incoherent: $F (1, 43) = 1.21$, $p = .278$, $\eta^2=.03$; coherent vs traditional: $F (1, 43) = 0.36$, $p = .555$, $\eta^2=.01$).

Discussion of Experiment 2

In Experiment 2, we tested whether illusory transparency bias, like creeping determinism (e.g., Nestler et al., 2008a; Nestler & von Collani, 2008), is sensitive to the experimental manipulations that alter participants' capacity to set up a coherent model

about the event. As predicted, again in Experiment 2, neither the inclusion of an outcome nor the sort of outcome affected participants' judgements about the writer's intention. In every kind of scenario, participants attributed to the writer a higher intention of sarcasm when the experience was negative. In addition, as we hypothesised, the illusory transparency effect was present in the coherent scenarios, whereas it disappeared in the incoherent ones. Both the presence and the absence of the effect were due to the differences in sarcasm attribution to the addressee's interpretation in the negative conditions. These results showed that the relevance of the causal information available is crucial to elicit the bias in the illusory transparency task, as occurs in creeping determinism (Yopchick & Kim, 2012). Thus, the illusory transparency effect disappears in the conditions where the causal information available does not allow participants to explain the outcome, whereas it is present in the conditions where it allows them to do so. In the positive control conditions, participants' ratings did not differ significantly.

Nevertheless, one initially unpredicted result was shown: unlike in Experiment 1, the illusory transparency effect was not shown in the traditional condition. There is a key difference between Experiment 1 and Experiment 2 that could explain these results: in Experiment 2 only, the outcome cannot be predicted from the value of the experience. Here, participants were informed that an interpretation was sarcastic and sincere the same number of times, regardless of whether the experience was positive or negative. In this context, participants could have learnt that in the absence of a known outcome, it is not possible to establish a causal connection (e.g. a negative experience with a positive message leads to a sarcasm interpretation). This could explain why in the trials without an outcome (traditional condition), the illusory transparency effect is not shown. Furthermore, several creeping determinism studies have found that the magnitude of bias diminishes when participants are asked to consider alternative outcomes (Arkes, Faust, Guilmette, & Hart, 1988, Nestler et al., 2008b). According to Roese & Olson (1996), this reduction occurs because the perceived causal connection between antecedent conditions and outcome is weakened. It happens not only when individuals can imagine an alternative outcome preceded by the same antecedent conditions but also when they realise that different antecedent conditions can lead to the same outcome. In our study, it might be that making sense when different outcomes were preceded by the same antecedent conditions weakened the perceived causal connection between the experience (the antecedent) and the addressee's interpretation (outcome), resulting in illusory transparency bias disappearing in traditional and incoherent scenarios. To further explore

this possibility, we conducted a third study where we manipulated the actual experience-message causal relation using a causal conditional that strengthened or weakened that relation. In addition, to test once again whether illusory transparency is sensitive to the capacity to build an event causal model, we asked participants to complete only outcome scenarios (incoherent and coherent).

Experiment 3

In the previous experiment, it was shown that the differences in coherence between the polarity of the experience (positive or negative) and the outcome provided (sincere or sarcastic addressee's interpretation) has an effect on the magnitude of illusory transparency. This effect has been explained because coherence influences the readiness to establish a causal relationship between the experience and the attributed interpretation. However, it could be argued that differences found in the illusory transparency effect could be due not to coherence producing a causal relation but to coherence influencing in some way the other basic processes involved. For example, in incoherent contexts, but not in coherent ones, people could be activating deliberative processing (Brannon & Gawronski, in press; Ragni et al., 2014) that might reduce biases. In Experiment 3, we induce the causal relationship in a different way, maintaining the coherence factor. To this end, we manipulated the actual experience-message relation by including a sentence that established either a counterfactual or semifactual causal relation between the two. Counterfactual conditional expressions such as "If David's experience at the restaurant had been different, his message would have been different" lead people to establish strong causal relations between the antecedent and the consequent. In contrast, semifactual conditional expressions such as "even if David's experience at the restaurant had been different, his message would have been the same" lead readers to cancel that causal relation (see Byrne, 2016, Moreno-Ríos et al., 2008; Ruiz-Ballesteros & Moreno-Ríos, 2017). Two groups of participants read scenarios that included information like that mentioned above, one with counterfactual expressions and the other with semifactual expressions. With this manipulation, we led readers to interpret that the message depended on the experience in the counterfactual group but it did not do so in the semifactual group. In all the other aspects, the participants' task was the same as in Experiments 1 and 2.

Roese & Olson (1996) showed that creeping determinism decreased when participants were encouraged to establish semifactual relations between antecedents and

outcome: this undermined the link, whereas with the condition in which they were encouraged to create counterfactual relations, the link was strengthened. If in Experiment 2, the bias in the traditional and the incoherent scenarios disappeared because the experience-message relation was weakened, then in this experiment, the bias should also disappear (or at least diminish) in the semifactual group, because the causal relation is broken. In contrast, we expect that participants in the counterfactual group will show a higher illusory transparency than participants in the semifactual group. Nevertheless, participants from both groups will complete incoherent and coherent scenarios. In Experiment 2, the magnitude of the bias was higher in the coherent scenarios than in the incoherent ones, where the bias disappeared. Therefore, we expect the same effect in both counterfactual and semifactual groups.

Finally, unlike in the previous experiments, the manipulation of the actual experience-message relation could have an effect on the participants' judgements about the writer's intention. This is because the conditional statement refers to how much the writer's message depends on the experience. Interestingly, one important difference between the causal illusion hypothesis and other proposals is that the magnitude of the illusory transparency can be modulated regardless of whether participants attribute more or less intention to the writer. As other illusory transparency accounts argue, (see Gerrig et al., 2000; Keysar, 2000 for details), if the bias arises from participants' perception of sarcasm in the writer's intention, then a lower attribution to the writer's intention will be expected in the semifactual group, when participants are informed that the message did not depend on the experience. On the other hand, if the illusory transparency is based on the establishment of a causal relation between the writer's experience and the addressee, as the causal illusion hypothesis predicts, that manipulation should not affect judgements about the writer's intention.

Method

Participants

A total of eighty-four undergraduate (6 males and 78 females) students from the University of Granada (age: $M = 20.98$; $SD = 4.92$) participated in the study in exchange for course credits. Three participants (one from the semifactual group and two from the counterfactual one) were excluded from the final sample because they answered more than three comprehension questions wrongly. Therefore, the final sample was composed of eighty-one participants. Forty participants (2 males and 38 females) completed the

counterfactual booklet and forty-one (3 males and 38 females) the semifactual booklet. According to the sensitivity power analysis, the sample size of this experiment allowed us to detect an effect (i.e., the main hypothesis regarding the outcome knowledge × polarity of the experience interaction × conditional) as small as f= 0.242 ($\eta 2$= .056; $\alpha = 0.05$, $1 - \beta = 0.80$, the average partial correlations among all repeated measure; r=0.02).

Materials

The same twelve outcome scenarios (coherent and incoherent) used in Experiment 2 were used in this study. In the four versions of each scenario (positive-sincere, negative-sarcastic, positive-sarcastic and negative-sincere), we included after the positive message and the informed addressee's interpretation, a conditional statement informing participants about the actual relation between the experience and the message (see Roese & Olson, 1996 for the most effective location of counterfactuals). In the counterfactual booklets, the conditional conveys that the message depended on the experience (e.g., "If David's experience at the restaurant had been different, his message would have been different"). By contrast, in the semifactual booklets the conditional conveys that the message did not depend on the experience (e.g., "even if David's experience at the restaurant had been different, his message would have been the same").

Twelve different booklets were created: six booklets for the counterfactual group and six for the semifactual one. The only difference between the booklets was the wording of the conditional statement (as semifactual or counterfactual conditional). As in Experiments 1 and 2, each booklet was composed of twelve scenarios. However, here all the scenarios included an outcome. Specifically, each booklet was composed of six scenarios with a coherent outcome (3 positive-sincere and 3 negative-sarcastic) and six scenarios with an incoherent outcome (3 positive-sarcastic and 3 negative-sincere). Eight of these twelve scenarios (two per version) included the conditional statement, expressed as a counterfactual or a semifactual conditional. The remaining four scenarios (one per version) were filler scenarios where the outcome appeared but not the conditional statement. We included these filler scenarios in order to highlight the presence of the conditional statement. As in Experiments 1 and 2, we counterbalanced the different versions of the scenarios across the booklets. However, unlike in Experiment 1, the order of presentation of the scenarios was randomised.

Procedure

The procedure was exactly the same as in Experiments 1 and 2.

Results

Only the participants' responses to the experimental scenarios were analysed. As in Experiments 1 and 2, we calculated an index of sarcasm for the participants' sarcasm attributions to the writer's intention and the addressee's interpretation. We submitted their responses to both questions to an ANOVA 2 (outcome knowledge: coherent, incoherent) x 2 (polarity of the experience: positive, negative) x 2 (conditional: counterfactual, semifactual) with the first two factors manipulated within participants and the third between groups. Results are shown in Tables 4 and 5.

Sarcasm attributed to the writer's intention

The analysis revealed a main effect for the polarity of the event factor, $F(1,79) = 793.50$, $p < .001$, $\eta^2=.91$. In both counterfactual and semifactual groups, participants attributed more sarcasm to the writer's intention when the polarity of the experience was negative than when it was positive. In addition, their sarcasm attributions were not influenced either by the sort of outcome provided, $F(1,79) = 1.92$, $p = .170$, $\eta^2=.02$, or by the type of conditional relation $F(1,79) = 0.58$, $p= .449$, $\eta^2=.01$. No significant interactions were found (outcome knowledge x conditional: $F(1,79) = 1.24$, $p = .269$, $\eta^2=.02$; polarity of the experience x conditional: : $F(1,79) = 0.32$, $p = .572$, $\eta^2 <.01$; outcome knowledge x polarity of the experience x conditional: : $F(1,79) = 0.85$, $p = .361$, $\eta^2=.01$).

Table 4. Degree of sarcasm (from 1 to -1) attributed to the writer's intention in the counterfactual and semifactual groups according to the polarity of the experience and the outcome knowledge (coherent and incoherent).

	COUNTERFACTUAL		**SEMIFACTUAL**	
	Coherent	**Incoherent**	**Coherent**	**Incoherent**
Positive	-0.80 (0.46)	-0.95 (0.22)	-0.90 (0.30)	-0.88 (0.40)
Negative	0.78 (0.42)	0.70 (0.52)	0.68 (0.61)	0.63 (0.62)

Sarcasm attributed to the addressee's interpretation

The analysis yielded a main effect for the polarity of the event factor, $F(1,79) = 29.06$, $p < .001$, $\eta^2=.27$: participants attributed more sarcasm to the addressee's interpretation when the polarity of the event was negative than when it was positive in both counterfactual and semifactual groups. The interaction between the polarity of the

event and the conditional was also significant, $F(1,79) = 4.63$, $p = .034$, $\eta^2=.06$. The interaction reflected the fact that in negative conditions, participants attributed significantly more sarcasm to the addressee's interpretation in the counterfactual group than in the semifactual one, $F(1,80) = 11.26$, $p = .001$, $\eta^2=.13$. In positive conditions, participants' ratings did not differ significantly, $F(1,80) = 0.09$, $p = .755$, $\eta^2 < .01$. Finally, the sarcasm attribution to the addressee's interpretation was slightly higher in the counterfactual group than in the semifactual one (-0.19 vs -0.35 respectively) but the difference did not reach significance $F(1,79) = 3.58$, $p = .062$, $\eta^2=.04$. Neither outcome knowledge factor, $F(1,79) = 1.68$, $p = .199$, $\eta^2=.02$, nor the other interactions were significant (outcome knowledge x conditional: : $F(1,79) = 0.32$, $p = .571$, $\eta^2 <.01$; outcome knowledge x polarity of the experience: : $F(1,79) = 1.28$, $p = .262$, $\eta^2=.02$; outcome knowledge x polarity of the experience x conditional: $F(1,79) = 0.25$, $p = .621$, $\eta^2 <.01$).

In order to further explore whether the coherence of the causal event model provided to participants has an effect on the magnitude of the illusory transparency, we analysed independently in each group (semifactual and counterfactual), participants' responses to coherent and incoherent scenarios. Results showed that participants from the counterfactual group exhibited the bias in the coherent $F(1,39) = 24.72$, $p < .001$, $\eta^2=.40$, and incoherent scenarios, $F(1,39) = 11.47$, $p < .005$, $\eta^2=.23$. However, in the semifactual group, the illusory transparency effect disappeared in the incoherent scenarios, $F(1,40) = 1.68$, $p = .202$, $\eta^2=.04$, and differences did not reach significance in the coherent ones, $F(1,40) = 3.52$, $p = .068$, $\eta^2=.10$.

Table 5. Degree of sarcasm (from 1 to -1) attributed to the addressee's interpretation in the counterfactual and semifactual groups according to the polarity of the experience and the outcome knowledge (coherent and incoherent).

	COUNTERFACTUAL		**SEMIFACTUAL**	
	Coherent	**Incoherent**	**Coherent**	**Incoherent**
Positive	-0.53 (0.64)	-0.50 (0.55)	-0.49 (0.75)	-0.49 (0.68)
Negative	0.25 (0.74)	0.00 (0.78)	-0.17 (0.70)	-0.27 (0.81)

Discussion of Experiment 3

Experiment 3 showed that manipulating the actual experience-message relation alters the magnitude of illusory transparency. As predicted, the magnitude of illusory transparency was higher when that relation was strengthened. In the negative conditions, participants from the counterfactual group attributed more sarcasm to the addressee's interpretation than did participants from the semifactual group. In addition, we found that in the semifactual group, the illusory transparency disappeared in incoherent scenarios and did not reach significance in the coherent ones. On the other hand, the bias was present in both sorts of scenario in the counterfactual group. Finally, in this experiment too, the manipulation did not affect participants' judgements about the writer's intention. Regardless of the relation of dependence or independence established between experience and message, participants attributed more sarcasm to the writer's intention when the experience was negative than when it was positive.

General Discussion

In three experiments, we investigated the role of causal attribution as a possible common process implicated in two epistemic biases: illusory transparency and creeping determinism. We obtained measures for each participant in the two biases, using a single task. In this way, we were able to test an alternative explanation for the illusory transparency bias. It was originally formulated based on the attribution of the communicative intention from one character to another in a story (Keysar's, 1994). As in Keysar's original proposal, we suggested that the bias arises from participants' illusory attribution. However, we proposed that what participants attribute to the uninformed addressee is their own knowledge about the event rather than their knowledge about the writer's intention. That knowledge attribution would lead biased participants to create a causal link between the writer's experience and the addressee's interpretation (the *causal illusion hypothesis*), and therefore, to predict a sarcastic addressee's interpretation. Thus, in our proposed account, the bias is related to a causal mechanism, as was shown in creeping determinism. To test this hypothesis, we modified the traditional illusory transparency task by including a final sentence informing participants about how the addressee had interpreted the message (outcome scenario). As predicted, although the illusory transparency effect appeared in both the traditional and the outcome scenarios, the magnitude of the bias was higher in the outcome scenarios. Yopnick and Kim (2012) showed that creeping determinism only appears when participants are provided with

relevant causal information; otherwise, it disappears. In Experiment 1 and in the coherent scenarios from Experiment 2, the creeping determinism emerged, suggesting that participants perceive a causal connection between the experience (antecedents) and the addressee's interpretation (the outcome). The presence of such a causal link is supported by another of our findings: participants who exhibited a greater illusory transparency showed a greater creeping determinism. Thus, the presence of an outcome would have reinforced the causal connection perceived by biased participants between the experience and the addressee's interpretation.

In addition, the illusory transparency effect, like creeping determinism (e.g., Nestler et al., 2008a; Nestler & von Collani, 2008), was sensitive to the manipulations that altered participants' capacity to set up a coherent causal model about how things came about. In Experiment 2 we tested that idea by asking participants to complete scenarios where the available antecedent was relevant to explaining the outcome (coherent scenarios) and scenarios where it was not (incoherent scenarios). As predicted, the illusory transparency effect only appeared in the coherent scenarios. By contrast, in the scenarios where the outcome provided was incoherent with the hypothesised causal model, the differences in sarcasm between the positive and the negative condition disappeared. This result is consistent with previous creeping determinism studies, which found that the bias disappears when participants are unable to make sense of the outcome provided (Nestler et al., 2010; Pezzo, 2003; Yopchick & Kim, 2012; Wasserman, et al., 1991). Based on the hypothesised causal model built by biased participants, there is no way to explain how the addressee could have interpreted the message as sarcastic after a positive experience and as sincere after a negative one. Thus, it might be that the bias in the incoherent scenarios disappeared because biased participants were unable to make sense of the outcome provided (e.g., Pezzo, 2003). However, that unsuccessful sense-making could not explain why in Experiment 2 the bias also disappeared in the traditional scenarios. In our view, a more reasonable explanation is that the bias disappears because the causal link between the writer's experience and the addressee's interpretation is weakened in those two situations. Roese and Olson (1996) reported that the fact of making sense of an outcome (e.g., sincere) preceded by different antecedent conditions (positive and negative experience) weakened the perceived causal link between antecedent (the experience) and outcome (addressee's interpretation). To test this explanation, we manipulated the actual relation between experience and message in the third experiment. In one condition, that relation was strengthened (counterfactual group), whereas in the

other condition, the relation was weakened (semifactual group) by giving conditional information (see e.g. Byrne, 2016 for the effect of counterfactual and semifactual conditionals). Results showed that although participants from both counterfactual and semifactual groups showed illusory transparency, participants from the counterfactual group attributed more sarcasm to the addressee's interpretation in the negative conditions. In addition, in the semifactual group, the illusory transparency effect did not appear in the incoherent scenarios, whereas it was present in the counterfactual scenarios. Similarly, the illusory transparency effect was present in the coherent scenarios in the counterfactual group but was reduced in the semifactual group (actually, it did not reach significance). Thus, as occurred in creeping determinism studies (Roese & Olson, 1996), strengthening the perceived causal link between experience and message increases the magnitude of the illusory transparency. And, more importantly regarding the focus of the present work, the magnitude of the bias depended on the participant's own interpretation of how much the message described the experience.

Finally, as the causal illusion hypothesis predicted, informing participants about how the addressee had interpreted the message did not have any effect on participants' judgements about the writer's intention. In Experiments 1 and 2, participants always attributed more sarcasm to the writer's intention when the experience was negative. In Experiment 3, we tested directly the dependency between participants' attribution of the intention and their interpretation by including the writer's intention in the hypothesised causal link. We argued that if the illusory transparency phenomenon, as defended in traditional illusory transparency accounts (e.g., Gerrig et al., 2000; Keysar, 1994), arises from participants' perception of sarcasm in the writer's intention, then the same effect found in judgements of the addressee's interpretation should appear in those of the writer's intention. However, there were no differences in sarcasm attribution to the writer's intention between the counterfactual and the semifactual groups. In both groups, and as in Experiments 1 and 2, participants' sarcasm attributions were higher in the negative condition than in the positive one. In our view, this finding provides strong evidence for the causal illusion hypothesis by showing that the magnitude of the bias is independent of the participants' perception of sarcasm in the writer's intention. Specifically, we believe that this happens because the mechanisms involved in judgments of the writer's intention and the addressee's interpretation are different. In the latter judgements, causal attribution processes would lead participants to predict a sarcastic addressee's interpretation in the negative condition. On the other hand, judgments of the

writer's intention would be driven by pragmatic principles. In particular, these judgements would be based on their general knowledge about sarcasm comprehension rules (Matsui et al., 2016). It explains why the sarcasm attributed to the addressee's interpretation is independent of the sarcasm attributed to the writer's intention. This explanation is consistent with other illusory transparency findings in childhood. In a recent study, Gordo and Moreno-Ríos (2019) found that each judgement seems to follow a different developmental trend. Nevertheless, we are not ruling out that in some participants and/or under certain conditions, there could be transference from the writer's intention to the addressee's interpretation. Previous studies (e.g., Moreno-Ríos et al., 2011; Weingartner & Klin, 2009) have found that the illusory transparency effect is sensitive to pragmatic manipulations. For example, Moreno-Ríos and cols. (2011) found that the illusory transparency disappeared when the addressee of the message was an accidental reader. However, it might also be that in that study the bias disappeared because the presence of an accidental addressee weakened the illusory causal link created by biased participants, as also occurs in the present work.

Conclusion

The new hypothesis is also consistent with previous results and helps us deepen our knowledge of the mechanism that causes the bias and its relation to other egocentric biases. Results in the present study show that causal attribution processes are involved in the illusory transparency phenomenon (the causal illusion hypothesis). This challenges the traditional explanation of the phenomenon (e.g., Keysar, 1994), which posits that the bias arises from transference of the writer's intention. In addition, our results suggest that, as some authors have posited (e.g., Epley et al., 2004; Royzman et al., 2003), there is a common mechanism in illusory transparency and the creeping determinism manifestation of hindsight bias. Moreover, both epistemic biases have been related to other cognitive biases, such as *"realistic bias", "I-knew-it-all-along effect",* or *"the curse of knowledge,* which are also caused by people's inability to discard their own knowledge when making judgments about others' knowledge and/or perspective (e.g., Bernstein et al., 2018; Birch et al., 2017). Future research should investigate whether similar causal reasoning processes could be acting in these cognitive biases.

Open Practices

The three experiments in this article earned an Open Data badge for transparent practices. Data for the experiments are available at https://osf.io/7tecj/?view_only=16ce4a27b18b4afda6c04ca61cd3ba44

Appendix 1: Scenarios structure example

David asked his office mate, Julia, to recommend a restaurant. His parents were in town and he wanted to take them to a good place. ‘‘I strongly recommend this new Italian place, called Tony’s. I just had dinner there last night and it was marvellous. Let me know how you all enjoy it,’’ said Julia who really liked Italian food.

That evening, David and his parents ate there. The food was unimpressive and the service was mediocre (or instead, in positive version: ‘‘The food was indeed delicious and the service was impeccable’’)

When David arrived at work the next morning, he did not find Julia at her desk. He remembered she was taking the morning off, so he left a note on her desk: ‘‘You wanted to know about the restaurant you recommended to me? Well, it was marvellous, just marvellous.’’

(Experiment 3: “If David’s experience at the restaurant had been different, his message would have been different” –counterfactual booklet-; “even if David’s experience at the restaurant had been different, his message would have been the same” –semifactual booklet-).

Julia interpreted the message as sarcastic (sincere) –outcome provided-

Capítulo 4

Auditory hindsight bias in school-age children

Abstract

We report two experiments investigating hindsight bias in children, focusing on a rarely studied age range of 8 to 13 years. In Experiment 1, we asked children to complete both an auditory and a visual hindsight task. Children exhibited hindsight bias in both tasks, and the bias decreased with age. In Experiment 2, we further explored children's auditory hindsight bias by contrasting performance in hypothetical and memory designs (which previous research with adults had found to involve different mechanisms: fluency vs. memory reconstruction). Children exhibited auditory hindsight bias in both tasks, but only in the hypothetical design the bias magnitude was modulated by a priming manipulation designed to increase fluency, replicating and extending the pattern found in adults to children.

Keywords: auditory hindsight bias, children's cognitive biases, hypothetical design, memory design, priming.

Pre-print version of the manuscript: Gordo, C., Moreno-Ríos, S., & Blank, H. (2021). Auditory hindsight bias in school-age children. Under review with minor changes in Journal of Experimental Child Psychology (14/06/2021).

Introduction

"How old are you?"

"I am thirty-three years old, but if I had not told you, what would you have said?"

This is an example of an everyday situation in which we must ignore provided information in order to issue a judgment. However, once people know an outcome (e.g. a person's age), it is difficult for them to ignore this knowledge in order to reason about their own prior state of knowledge (e.g., what I would have thought if she/he had not told me his/her age), or someone else's (e.g., what another person would think about his/her age). While in this situation people might say "I thought you were younger" out of politeness, a large body of research (Blank et al., 2007; Guilbault et al., 2004; Hawkins & Hastie, 1990; Musch & Wagner, 2007) also suggests that their judgements would be biased in favour of the provided outcome (e.g., his/her age). This phenomenon has been called hindsight bias (Fischhoff, 1975): people's tendency to believe, once they know the outcome, that they were able to predict it, claiming that they "knew it all along" – "I always thought you were in your early thirties" (e.g., Bernstein et al., 2004).

Hindsight bias has been documented across cultures (Pohl et al., 2002) and in a large variety of judgement contexts, including medical diagnoses (Arkes, 2013; Arkes et al., 1981; Dawson et al., 1988), sport events (Blank et al., 2015; Gray et al., 2007; Leary, 1981), legal decisions (Harley, 2007; Kamin & Rachlinski, 1995; Hawkins & Hastie, 1990), election outcomes (Blank et al., 2003; Blank & Nestler, 2006; Leary, 1982), and consumer satisfaction surveys (Zwick et al., 1995). Also, hindsight bias has been obtained with different stimuli: verbal (e.g. Bernstein et al., 2011), visual (e.g. Bernstein et al., 2004; Bernstein et al., 2007; Bernstein et al., 2011), gustatory (Pohl et al., 2015) and auditory (Bernstein et al., 2018; Bernstein et al., 2012; Higham et al., 2017). Hindsight bias has been studied in both adults and children (see below), using different types of materials but never auditory stimuli. Our aim in this work was to test whether children exhibit auditory hindsight bias and, further, whether its magnitude changes with age.

Assessing hindsight bias

Two main experimental designs have been used to investigate hindsight bias: memory designs and hypothetical designs. In memory designs, participants first answer questions such as "How tall is Mount Kilimanjaro?". Later, they receive feedback about the solution (e.g., the height of Mount Kilimanjaro is 5895 meters). At a third point in time (often directly after receiving feedback), they are asked to remember their foresight judgements. Typically, participants' recollections of their foresight judgements are biased

towards the solution feedback they received earlier. By contrast, participants in hypothetical designs are asked to make a hypothetical judgement after having been provided with the solution ("the height of Mount Kilimanjaro is 5.895 metres – what would have been your estimate had I not told you?"). In a control condition, participants answer the same questions without solution feedback. Here again, the (hypothetical) hindsight estimates are biased by the solution feedback.

Of particular interest in the context of the present research, there is a social version of the hypothetical design in which participants are not being asked to make hindsight judgements about themselves (i.e. their naïve "prior" self) but about the extent to which naïve others (typically peers) would have known the correct answer to a knowledge question (e.g. "out of 100 of your peers, how many would have known the correct answer?"). This design variant was first introduced by Fischhoff (1975), who found the magnitude of hindsight bias to be very similar to the one obtained in the traditional hypothetical design. It has been used particularly in research with children, because it is arguably more intuitive for them than hypothetical judgements.

Auditory hindsight bias

In adults, auditory hindsight bias has been tested mainly using the social version of the hypothetical design. Different studies have found that individuals overestimate their peers' ability to identify an auditory target when they receive feedback about the auditory target's identity. For example, Epley, Keysar, Van Boven and Gilovich (2004) told half of their participants that, in a song by the rock band Queen played backwards, it is possible to identify the message "it is fun to smoke marihuana". Informed participants predicted that a higher percentage of their naïve peers would identify the message than uninformed participants did. More recently, several studies assessed auditory hindsight bias using muffled words (e.g., Higham et al., 2017) or sentences (Bernstein et al., 2012). Again, results showed that participants are unable to discard the provided feedback when asked to predict how many of their naïve peers would identify the distorted auditory stimuli (Bernstein et al., 2012, 2018; Higham et al., 2017).

Hindsight bias in children

Hindsight bias research in children has so far used knowledge questions or visual materials as stimuli. For example, Bernstein et al. (2011) asked children to identify, as soon as possible, blurred objects that progressively became clearer. Thereafter, children were asked to estimate at which point a puppet named Ernie (the naïve other) would be able to identify the same degraded objects. Children claimed that Ernie was able to

identify the same objects at a more degraded level than they themselves had been able to previously. Children also showed the bias in memory designs. In these studies, the children's task was to answer several numerical knowledge questions (such as 'How many countries are there in Africa?'). After receiving solution feedback, they were asked to recall their previous answers. Results showed that children's recollections of their previous foresight ratings were biased towards the provided feedback (e.g., Bernstein et al., 2011; Pohl, Martin & Bayen, 2010). In terms of developmental differences, these and other studies suggest that the magnitude of hindsight bias decreases from 3 to 5 years (Bernstein et al., 2004, 2007, 2011), reaching a stable level during middle and late childhood (Bernstein et al., 2011; Pohl et al., 2010). Little is known, however, about the precise development of the bias in middle childhood, the period covered in the present research in one-year increments.

Experiment 1

The goal of the present work was to investigate auditory hindsight bias in school children and to find out if the bias magnitude changes (or not) with age. To this end, we created a task involving songs that seemed to contain a "hidden" message. This "hidden" message was a sentence that could be "heard" when the song was played backwards (see Epley et al., 2004). In fact, these messages are very difficult to detect unless their content is provided in advance. In some trials, children were informed about the message content (e.g., "There are no tomatoes in your garden") before playing the song (hindsight judgements condition; HJ). In other trials, the songs were played without information about the content (foresight judgements condition; FJ).

In all conditions, the children's task was to estimate how many of a group of six naïve peers would be able to identify the hidden message in each song. That is, we asked children to provide an explicit numerical judgement, following the principal methodology used in studies of auditory hindsight bias with adults (for example, see Epley et al., 2004; Higham et al., 2017). To allow comparisons to previous studies of hindsight bias in children that used visual materials (e.g., Bernstein et al., 2004, 2007; Harley, Carlsen & Loftus, 2004), we also included a visual hindsight bias task. Children were shown pictures of celebrities and were asked to estimate how many of a group of six naïve peers would recognize these celebrities. For half of the pictures, the children received feedback about the celebrities' identities before making their judgements (hindsight judgements

condition; HJ); for the other half, they made them without identity information (foresight judgements condition; FJ).

We tested children from third grade (~8 years) onwards, for several reasons. (1) While it seems clear that hindsight bias decreases from 3 to 5 years (Bernstein et al., 2004, 2007, 2011), there is only one study (to our knowledge) that explored the bias trajectory during middle and late childhood (Bernstein et al., 2011). (2) Although metacognitive skills develop during early childhood, one aspect takes longer: source monitoring, or the ability to accurately attribute the origins of one's memories, knowledge and beliefs (Johnson et al., 1993; Lindsay, 2008). This skill is crucial in the present study because the question is whether children will attribute their recognition of the 'hidden messages' to the information provided by the experimenter or to their own unaided understanding. Several studies indicate that there is an increase in this skill from about four to eight years of age (Drummey & Newcombe, 2002, Ruffman et al., 2001; Sluzenski et al., 2004), but little is known about its later development. (3) Before 7 years, children's judgements of the frequency of events are poor (Sharman et al., 2011). Given that we asked children about frequencies (how many peers?), we had to ensure that children had a basic grasp of frequency judgements.

Based on the available previous research (e.g., Bernstein et al., 2004, 2007, 2011; Pohl et al., 2010), we expected that children aged between 8 to 13 years would exhibit both auditory and visual knowledge hindsight bias, and at a relatively stable level (Bernstein et al., 2011), but in light of the sparsity of research on the development of hindsight bias and source monitoring over this age range we kept an open mind.

Debiasing in children?

We included a further experimental condition in the auditory task as a test of the robustness of auditory hindsight bias in children. In research with adults, Bernstein et al. (2018, Exp. 2) found that providing participants with an initial unaided identification trial before making their hindsight judgements successfully reduced auditory hindsight bias. Similarly, we included a condition where children listened to the same song twice (two-presentations condition), first without knowing the message content (two-presentations foresight judgements condition; 2FJ) and then for a second time after being informed about it (two-presentations hindsight judgements condition; 2HJ). Like Bernstein et al. (2018), we reasoned that this allows them to personally experience the difference in detection difficulty and, perhaps, subsequently avoid the hindsight bias. On the other hand, other research in adults has shown that auditory hindsight bias and hindsight bias

in general persist despite instructions or opportunities to avoid it (Bernstein et al., 2012; Bernstein et al., 2018, Exp. 3; Harley et al., 2004; Pohl & Hell, 1996); thus, we treated this issue as exploratory.

Method.

Both of our studies (Exps. 1 and 2) were approved by the university's ethical review board (Comité de Ética en Investigación Humana de la Universidad de Granada: 1068/CEIH/2020).

Participants

Ninety-nine schoolchildren (46 girls and 53 boys) aged between 8 and 13 years completed the auditory and visual identification tasks: n = 27 from 3rd grade (M = 8.60; SD = 0.51), n = 23 from 4th grade (M = 9.70; SD = 0.70), n = 23 from 5th grade (M = 10.43; SD = 0.51) and n = 26 from 6th grade (M = 11.70; SD = 0.70). All of the participants were native speakers of Spanish. One participant was excluded from the visual identification task sample for not answering all items. Therefore, the visual identification task sample consisted of 98 participants (46 girls and 52 boys).

Materials

We briefly pilot-tested auditory and visual task materials (using the exact same procedure) with adults (N = 22, 8 males and 14 females) before applying it to children. The main purpose of this was to pre-test the relative difficulty of the materials (see below).

Auditory task

Twenty-one segments of songs with a "hidden" message (when played backwards) served as auditory stimuli. Thirteen songs were in Spanish and eight in English, but the 'hidden message' for the children was always in Spanish (see https://www.youtube.com/watch?v=TvhCThWWy3Q for an illustrative English example). Songs were selected following Epley et al. (2004; p. 333): *"the selection played for participants is sufficiently unclear to ensure that very few (if any) of those who were not told about the critical phase would hear it on their own, but virtually everyone told about the phrase in advance would have no difficulty detecting it"*.

On the basis of the pilot study, we arranged the auditory stimuli into three sets of seven songs according to the difficulty of identifying the "hidden" message (based on adults' responses in the FS judgement condition of the pilot study), an easy set, a medium set and a difficult set. We counterbalanced the assignment of these song sets to the three

experimental conditions, resulting in three different slideshow presentations (A, B and C). In the Foresight Judgements Condition (FJ), the children listened to the song without being informed of the message. In the Hindsight Judgements Condition (HJ), they were told about the message before the song was played. Finally, in the Two-Presentations condition, children listened to the song twice: firstly, without being informed of the message (2FJ), and secondly, after being informed (2HJ; see Figure 1). The songs appeared in a random order in each presentation. Children were assigned to one of the slideshow presentations (A, B or C) sequentially following the order of arrival to the experiment. They provided their answers in a booklet containing twenty-one scales from 0 to 6, each preceded by a song number (e.g., Song 3). This number also appeared on the screen.

Figure 1. Auditory hindsight bias task. Experimental procedure per condition: Foresight Judgments, Hindsight Judgements and Two-Presentations condition. The English translation of the message in the example is "*There are no tomatoes in your garden*"

Visual identification task

Twenty-four pictures of celebrities from politics, music, sports, and literature, with whom the children were unlikely to be familiar (e.g., Margaret Thatcher, Keith Richards, etc.), were combined into a slide show presentation. In this presentation, we provided children with feedback about the identity of twelve celebrities (HJ stimuli). No information was provided about the identity of the twelve remaining pictures (FJ stimuli). The slide show included six blocks composed of four pictures each (two from a first set, e.g., George Bush and Angela Merkel, and two from a second set, e.g., Cristina Kirchner and Tony Blair). In each block (see Figure 2), children received information about the identity of one celebrity from set 1 (e.g., Angela Merkel) and one celebrity from set 2 (e.g., Tony Blair), whereas no information was provided about the remaining celebrities (e.g., George Bush and Cristina Kirchner); this assignment was counterbalanced across participants (i.e., there were two respective versions of the slide show). A detailed instruction sheet ensured that all groups received the same instructions and the same amount of information for each celebrity (e.g., "This is George Bush. He was president of the USA").

As in the auditory task, children gave their answers in a booklet. The booklet was divided into six numbered blocks, each of which contained four clearly numbered scales from 0 to 6 (for example: Picture 3). The block and picture numbers also appeared on the screen.

Procedure

The two tasks were applied in the same experimental session in a counterbalanced order. Children were tested in a quiet room, in groups of three to five. We started each experimental session explaining the task and the meaning of the scale: "*Your task is to indicate, on a scale like this one, how many of a group of six naïve peers would identify the "hidden" message in the songs (or recognise the celebrity's identity).*" We defined naïve peers as: "*… classmates who are not present in the room, and who therefore do not have the information about the message that I will give you".* In both tasks, before completing the experimental trials, children completed two practise trials in order to check whether they had understood the meaning of the scale. The practise trials involved questions with clear and unambiguous answers (e.g., How many of your peers would recognise Dani, the principal of their school?). Once all children had answered the practise trials satisfactorily (partly after having received further explanation and two more practise trials as necessary), the main task started.

Auditory task

The task began with the following instructions: "*Now, we will listen to some songs. Some people say there are hidden messages in these songs. In some cases, I will tell you the message people say can be heard, whereas in other cases I will not tell you anything about the content of the message. Moreover, there will be some cases in which I will play the same song twice. In these cases, in the first playback I won't tell you the content of the message, but I will do before playing the song for the second time."* Thereafter, children completed the experimental trials. In FJ and 2FJ trials, the song was played and children provided their ratings. In the information trials (HJ and 2HJ), we told children about the content of the hidden message before playing the song ("*In this song, some people say you can hear: "There are no tomatoes in your garden"*). Before each hindsight judgement, we reminded children that they should think of peers who do not have any information about the content of the message.

Visual identification task

Like the auditory task, this task started with a general task explanation before completing the experimental trials: *"Now I will show you two pictures of celebrities. I will introduce them to you first, and then these celebrities and two additional celebrities will appear on the screen one by one and you should give your answer. Remember that you should think of peers who were not present in the room and therefore did not receive the information about the celebrities that I will give you"*. After that, children began the experimental trials. Each trial had two phases (see Figure 2): a first phase where we told children about the identity of two celebrities, and a second phase where children made their judgements. In the first phase, two pictures of celebrities appeared on the screen and were introduced to the children[1] (e.g., *This is George Bush. He was president of the USA*). In the second phase, four pictures (these and two others) were presented to children one by one on screen and they made their judgements. Before each judgement, the experimenter reminded children that they should ignore the identity feedback when making their judgements.

Figure 2. Visual task. Experimental procedure per trial in the information and judgments phases

First Phase: Information Phase


Celebrity's identity information

Tony Blair was Former prime minister of UK

Angela Merkel: Prime minister of Germany

Second Phase: Judgements Phase

Cristina Kirchner's picture + FJ (0-6)

Angela Merkel's Picture + HJ (0-6)

George Bush's picture + FJ (0-6)

Tony Blair's Picture + HJ (0-6)

Results

Auditory task

As explained above, children provided four different judgements, the two traditional hindsight judgements (FJ and HJ) and the two judgements in the two-presentations condition (2FJ and 2HJ). Thus, we obtained two different hindsight bias measures: a traditional bias measure and a two-presentations bias measure. The judgements were submitted to a 2 (measure: traditional, two-presentations) x 2 (knowledge: foresight, hindsight) x 4 (grade: 3rd, 4th, 5th, 6th) mixed ANOVA, with the first two variables being manipulated within participants. This overall ANOVA revealed main effects for measure, $F(1, 95) = 4.28$, $p < .05$, $\eta 2 = .04$), knowledge, $F(1, 95) = 599.22$, $p < .01$, $\eta 2 = .86$), and grade $F(3, 95) = 9.72$, $p < .001$, $\eta 2 = .24$). We also obtained a significant three-way interaction between these factors, $F(1, 95) = 4.29$, $p < .05$, $\eta 2 = .12$); therefore, we subsequently analysed participants' responses in the traditional hindsight and the two-presentations conditions separately.

Children's estimates of the numbers of peers that would identify the message in the traditional conditions (FJ and HJ) were submitted to a 2 (knowledge: foresight, hindsight) x 4 (grade: 3rd, 4th, 5th, 6th) mixed ANOVA. Results are shown in Table 1.

The analysis revealed a strong main effect for knowledge, $F(1, 95) = 475.84$, $p < .001$, $\eta 2 = .83$): participants stated that a higher number of peers would identify the hidden message when they were informed about its content in advance. This is the *traditional hindsight bias effect*. The interaction with grade was also significant, $F(3, 95) = 8.43$, $p < .001$, $\eta 2 = .21$. Further analysis of the interaction showed a significant cubic trend, $M = 0.67$, 95% CI [0.22, 1.11], reflecting different amounts of hindsight bias across the age groups. Post-hoc comparisons (using Bonferroni correction for multiple comparisons, resulting in an adjusted alpha of 0.008, revealed that 3rd graders and 4th graders exhibited more bias than 6th graders, $F(1, 52) = 10.21$, $p = .002$, $\eta 2 = .17$ and, $F(1, 48) = 20.71$, $p < .001$, $\eta 2 = .31$, respectively. Also, 4th graders showed more bias than 5th graders, $F(1, 44) = 12.67$, $p = .001$, $\eta 2 = .22$. No other significant differences were found.

Table 1. Auditory task: Mean ratings (SDs) in the traditional foresight and hindsight judgements conditions, and bias magnitude (SD) per grade.

	FJ Condition	**HJ Condition**	**Bias magnitude**
3 grade	2.29 (1.16)	4.93 (1.02)	2.64 (0.98)***
4 grade	2.07 (0.99)	5.19 (0.59)	3.12 (1.12)***
5 grade	2.24 (1.13)	4.09 (1.28)	1.85 (1.30)***
6 grade	3.57 (0.22)	5.39 (0.69)	1.82 (0.88)***
Total	2.56 (1.26)	4.92 (1.04)	2.35 (1.19)***

Note: ***p < .001

The same analysis was performed on the ratings in the two-presentations condition (see Table 2) and revealed main effects of knowledge, $F(1, 95) = 427.88$, $p < .001$, $\eta 2 = .82$, and grade, $F(3, 95) = 8.99$, $p < .001$, $\eta 2 = .22$. Again, children exhibited hindsight bias – after being told about the content of the 'hidden messages', they thought a higher number of their peers would be able to identify these. Unlike in the previous ANOVA, however, there was no interaction with age here (although, descriptively, the pattern of bias magnitudes was the same – in fact, there is a perfect rank order correlation of the bias means in Tables 1 and 2).

Table 2. Auditory Task: Mean ratings (SDs) in the two-presentations foresight (2FJ) and hindsight (2HJ) judgements conditions, and bias magnitude (SD) per grade.

	2FJ Condition	2HJ Condition	Bias magnitude
3 grade	2.88 (1.39)	5.23 (0.95)	2.34 (1.11)***
4 grade	2.42 (1.02)	5.01 (1.00)	2.60 (1.35)***
5 grade	2.05 (1.02)	4.26 (1.16)	2.21 (1.07)***
6 grade	3.31 (1.10)	5.51 (0.40)	2.20 (0.94)***
Total	2.69 (1.23)	5.02 (1.01)	2.33 (1.12)***

Note: ***p < .001

Visual identification task

The findings obtained in the visual identification task were submitted to a 2 (knowledge: foresight, hindsight) x 4 (grade: 3rd, 4th, 5th, 6th) mixed ANOVA and are shown in Table 3. There was a main effect for the knowledge factor, $F(1, 98) = 327.48$, $p < .001$, $\eta 2 = .78$: children's ratings were higher for introduced celebrities than for celebrities whose identity had not been explained. This is the hindsight bias effect. The interaction with grade was also significant, $F(3, 94) = 13.82$, $p < .001$, $\eta 2 = .31$, and included a significant quadratic component, $M = -0.69$, 95% CI [-1.13, -0.24]. Post-hoc comparisons (using the same Bonferroni correction and adjusted alpha of 0.008 as in the auditory task) revealed that all other age groups showed more bias than 6th graders; 3rd graders: $F(1, 52) = 35.61$, $p < .001$, $\eta 2 = .41$; 4th graders: $F(1, 47) = 34.44$, $p < .001$, $\eta 2 = .43$; and 5th graders $F(1, 47) = 28.32$, $p < .001$, $\eta 2 = .38$. Finally, note again that the rank order of bias means was the same as in Tables 1 and 2.

Table 3. Visual identification task: Mean ratings (SDs) in the foresight and hindsight judgements conditions, and bias magnitude (SD) per grade.

	FJ Condition	HJ Condition	Bias magnitude
3 grade	2.07 (1.23)	4.56 (1.28)	2.49 (1.26)***
4 grade	2.38 (1.30)	5.02 (1.15)	2.65 (1.41)***
5 grade	2.37 (0.85)	4.45 (1.04)	2.08 (0.97)***
6 grade	3.36 (1.45)	4.21 (1.22)	0.85 (0.62)***
Total	2.55 (1.32)	4.55 (1.20)	1.99 (1.30)***

Note: ***p < .001

Relations between measurements

Three indexes of hindsight bias were computed for each child: *a visual identification* index, a *traditional auditory* index and a *two-presentations* index. These indexes (each reflecting the difference between children's ratings in the respective foresight and hindsight judgements conditions) were used to determine whether there were similarities between the three different bias measures. Only children who completed both the auditory and the visual tasks (N = 98) were included in this analysis. Results showed a strong positive relation between the *traditional auditory* index and the *two-presentations* index (r = 0.51, p < .001), a moderate relation between the *traditional auditory* index and the *visual identification* index (r = 0.28, p < .001), but no significant relation between the *visual identification* index and the *two-presentations* index (r = 0.17, p > .05).

Discussion

In Experiment 1, we found that auditory hindsight bias (in both conditions, traditional and two-presentations) is present in childhood, like other types of hindsight bias (e.g., Bernstein et al., 2004, 2007). Children aged between 8 to 13 years claimed that a higher number of their peers would identify the "hidden" message when they had been informed about its content previously. In fact, hindsight bias was present even when they experienced how difficult it is to identify the message without outcome feedback (two-presentations condition). Therefore, it seems that, in childhood as much as in adulthood, hindsight bias is a phenomenon that is hard to eliminate, or even reduce. Similarly, in the visual identification task, children claimed that more peers would recognise the celebrities when they had received feedback about their identity in advance.

Unlike previous studies (e.g., Bernstein et al., 2011), we found that the bias magnitude decreased substantially at some point in later childhood. Specifically, in all three tasks, hindsight bias peaked in 4th graders (~ age 9) and then declined. This result is consistent with Lagattuta et al.'s (2010) finding that between ages six and nine children tend to overly rely on their own previous personal experiences when asked to predict or explain behaviours or mental states. Thus, younger children's larger bias may reflect this tendency to base judgements of others' mental states on their own personal experiences.

Returning to the two auditory biases (traditional and two-presentations): While the pattern of findings was largely similar between these conditions, it remains an open question if the underlying processes are comparable. Relevant clues to answering this

question come from auditory hindsight bias research with adults, linking different hindsight designs (hypothetical and memory) to different underlying cognitive processes. Specifically, hindsight bias in hypothetical designs (corresponding to our traditional auditory hindsight condition) has been linked to misattributed processing fluency (e.g., Bernstein et al., 2018; Higham et al., 2017) – target knowledge facilitates processing of the auditory target (creating subjective fluency) and in turn leads to overestimating target identifiability (fluency misattribution). By contrast, hindsight bias in memory designs more strongly depends on how the pre-feedback judgements are recollected (if possible) or reconstructed (e.g., Higham et al., 2017).

Our two-presentations condition may be seen as a hybrid between a hypothetical and a memory design: On the one hand, as in the traditional (social) hypothetical task, participants made hypothetical judgements about a group of peers, unlike in a memory design where participants are asked to *remember* their first judgements. On the other hand, the intended effect of the first, unaided, listening experience should depend on participants' ability to remember/reconstruct it, and in this respect, the two-presentations condition resembles a memory design. Possibly, this hybrid design then produces a hybrid auditory hindsight bias that depends on both fluency and memory processes. We designed Experiment 2 to follow up this question and learn more about the involved cognitive mechanisms, by contrasting a (slightly modified, but conceptually equivalent) two-presentations condition with a proper memory design and using a priming procedure to assess the role of processing fluency. It was also of interest to see if the hindsight bias age trajectory observed in Exp. 1 would replicate in a different sample of children.

Experiment 2

In some recent studies linking auditory hindsight bias in hypothetical designs to processing fluency (Bernstein et al., 2018; Higham et al., 2017), *conceptual priming* has been used to demonstrate the role of fluency, based on the following logic: both priming and outcome knowledge are known to facilitate the processing of the auditory stimulus and create subjective fluency. Increases in fluency have an upper limit, though. This leads to the expectation that increases due to outcome knowledge (presumed to produce hindsight bias) should be less pronounced if fluency has already been heightened by conceptual priming (Higham et al., 2017). Therefore, the (auditory) hindsight bias should be smaller in a priming (compared to no-priming) condition, which would indirectly confirm the dependence of hindsight bias on fluency.

In their research, Higham et al. (2017) presented participants with semantically related (e.g., nurse – high fluency) and unrelated (e.g., grass – low fluency) prime words before playing muffled versions of targets words (e.g., 'dtr' for doctor). In the hindsight phase, participants heard a clear version of a muffled target word before hearing the prime word (related or unrelated) and the muffled version of the target word. Their task was to estimate how many of their naïve peers would be able to identify the muffled words without target knowledge. Results showed that prime relatedness interacted with auditory hindsight bias – whereas the bias was found for related and unrelated prime trials, the effect was smaller in the related prime trials than in the unrelated ones. Crucially, Higham et al. also used the same procedure to test priming effects in the memory design (with otherwise identical materials), but only found two independent main effects for priming and knowledge. By implication, this confirms the unique role of fluency in auditory hindsight bias, according to the logic set out above.

More recently, Bernstein et al. (2018) further studied the role of fluency in hypothetical designs and found that (1) repetition priming (i.e. priming the target word by itself) produces effects similar to conceptual priming (i.e. through related words) and (2) the modulating effect of repetition priming was independent of the number of repetitions (one, three, or six). As conceptual priming would be difficult to realise for the 'hidden messages' we used in Exp. 1 (and which we wanted to keep for reasons of comparability), we therefore used repetition priming for our priming manipulation in Exp. 2. Before providing foresight and hindsight judgements, children heard a recording that contained clearly spoken versions of some of the target messages that appeared in the songs, increasing processing fluency for those targets.

Overall, the procedure in Exp. 2 was conceptually similar to the two-presentations condition in Exp. 1, with some necessary adaptations due to the new priming manipulation and memory design: (1) In the hindsight phase, children provided hypothetical judgements (as in Exp. 1) for only half of the trials. In the remaining trials, they were asked instead to *recall their foresight ratings*. That is, the two-presentations condition from Exp. 1 was split into two versions differing only in the nature of the hindsight task – hypothetical (i.e., estimating how many peers would recognise the 'hidden message' without outcome knowledge) vs. memory (i.e., remembering one's own foresight judgement). (2) To accommodate priming phases and to provide meaningful retention intervals for the memory design, the foresight and hindsight phases of the auditory task were implemented as separate blocks (not immediately following each other

for a given song/message as in Exp. 1). (3) The priming procedure also required some necessary minor adaptations of the stimulus materials (see methods section for details). (4) Lastly, we asked questions probing the understanding of task instructions and the identification of the messages (as a check of the priming manipulation; Bernstein et al., 2018; Higham et al., 2017).

If different processes underlie hypothetical and memory auditory hindsight biases not only in adulthood but also in childhood (and if our two-presentations condition in Exp. 1 otherwise 'behaves' like a traditional hypothetical design), then we should expect to find a similar pattern of effects as in the research above, that is, two independent main effects of priming and knowledge in the memory design, and an interaction between these two factors in the (two-presentations) hypothetical design, specifically in the form of a smaller hindsight bias effect for priming as compared to no-priming messages (Higham et al., 2017).

Regarding developmental trajectories, we expected, as in Experiment 1, a decrease starting around 5th grade for the hypothetical design. Regarding the memory design, previous memory studies in children found the bias to be stable during middle and late childhood (Bernstein et al., 2011; Pohl et al., 2010), leading to the expectation that the magnitude of hindsight bias in the memory task should remain stable from 8 to 13 years. Finally, we expected that in both hypothetical and memory tasks the proportion of correct identifications would be higher for priming than for no-priming songs (Higham et al., 2017).

Method

Participants and Design

A total of 197 children (103 boys and 94 girls) aged between 8 and 13 years participated in the study. All of them were native speakers of Spanish. Nineteen children were excluded from analysis for answering comprehension questions incorrectly. Thus, our final sample consisted of 178 participants (94 boys and 84 girls): n = 35 from 3rd grade (M = 8.49; SD = 0.51), n = 45 from 4th grade (M = 9.47; SD = 0.50), n = 52 from 5th grade (M = 10.54; SD = 0.58) and n = 46 from 6th grade (M = 11.43; SD = 0.54). Within each grade, children's participation followed a 2 (knowledge: foresight, hindsight) * 2 (task: hypothetical, memory) * 2 (priming: no-priming, priming); all of these factors were manipulated within subjects.

Materials

The auditory stimuli consisted of twelve songs that seemed to contain a message; these were the eight English-language songs from Exp. 1 and four new English-language songs of the same kind. For the purposes of Exp. 2, we considered the Spanish-language songs from Exp. 1 unsuitable, as some residual unaided recognition of words in the Spanish songs might interfere with the priming procedure. The twelve songs were randomly split into four sets of three and assigned to the four experimental conditions and phases of the experiment as illustrated in Table 4. As shown in the table, six songs were used in the hypothetical task (3 with and 3 without priming) and the other six appeared in the memory task (3 with and 3 without priming). These stimulus sets were combined into four different counterbalanced presentations to ensure that each set of materials was used equally often in each task and condition.

Table 4. Assignment of stimulus materials to hindsight tasks and conditions across phases in Experiment 2.

Task/condition	**Phase 1** (priming)	**Phase 2** (foresight judgements)	**Phase 3** (priming)	**Phase 4** (hindsight judgements)
Hypothetical/ priming	3 sentences (e.g. Set A)	3 songs (e.g. Set A)	3 sentences (e.g. Set A)	3 songs (e.g. Set A)
Hypothetical/ no priming	---	3 songs (e.g. Set B)	---	3 songs (e.g. Set B)
Memory/ priming	3 sentences (e.g. Set C)	3 songs (e.g. Set C)	3 sentences (e.g. Set C)	3 songs (e.g. Set C)
Memory/ no priming	---	3 songs (e.g. Set D)	---	3 songs (e.g. Set D)

Note. The hypothetical and memory hindsight bias tasks were run consecutively (in counterbalanced order). Orthogonal to this, the assignment of stimulus sets to tasks and conditions was also counterbalanced (by rotating through the scheme). The materials in each set consisted of songs played backwards that supposedly each contained a 'message', which corresponded to the sentences used in the priming phases.

Each task involved four phases: a foresight phase, a hindsight phase and two priming phases. Like Higham et al. (2017), we aimed to increase the fluent processing of the auditory targets in both foresight and hindsight judgements. However, our repetition priming did not allow us to present the primes *immediately* before each foresight and hindsight judgement as in Higham et al. (2017); this would have practically eliminated the difference between prime and target knowledge. Therefore, we included two separate

priming phases to ensure that the expected priming effect was present while children made both foresight and hindsight judgements.

For the priming phases, we created four different recordings (corresponding to the four stimulus sets above) that included spoken versions of the "hidden" messages for half of the target songs, in a clear female voice. In each recording, the "hidden" message (a sentence) was repeated three times, for three of the targets. Thus, each recording included a total of nine sentences. The sentences were separated by 1s breaks and appeared in random order but with the restriction that the same message must not appear consecutively. In the foresight phase, the songs appeared one by one together with a song number (e.g., Song 1). In the hindsight phase, the slides also included the sentence (the "hidden message") that could be 'heard' in each song.

For both the hypothetical and memory tasks, children provided their answers in a booklet similar to the one used in Experiment 1. The booklet also included a message identification question after each foresight judgement: "*Did you hear a message in the song? If so, write it down";* and two instruction comprehension questions at the end of each hindsight phase: (1) *"In this activity, your task was: a) to remember your ratings from Activity 1* [i.e., the foresight ratings], *b) to guess how many of your peers would identify the hidden message, c) I can't remember*; and (2) "*when answering the questions where you were asked to think of six of your peers – did these peers receive any information about the content of the "hidden message"? Yes/no").*

Procedure

Children were tested in a quiet room, in groups of three to five. Both hindsight task conditions were applied in the same experimental session in a counterbalanced order. The session started with explaining the tasks and the meaning of the response scale to the children.

Unlike traditional priming manipulations, in which the prime (the message) is presented immediately before the target (the song), we presented all primes in separate independent phases. With this "unspecific" priming method, participants listened to all messages together before starting the foresight and hindsight phases of the hypothetical and memory tasks. Crucially, this method allowed us to differentiate the identity priming from target knowledge, which appeared linked to a specific target.

The procedure was the same in the hypothetical and memory task conditions. Both conditions started with a priming phase, in which children listened to the clear spoken version of three of the "hidden" messages as described above. Children were told: *"Now*

I will play a recording. You should listen to it carefully because it is going to give you some clues about the "hidden messages" that people say can be heard in these songs". Immediately after the recording, the foresight phase started. Six songs were played one by one and children provided their estimates. After each estimate, children answered the message identification question (see above). Once they had completed the foresight phase, the second priming phase started with the same instructions as in the first priming phase, and the same priming recording was played again. Immediately thereafter, they completed the hindsight phase.

In the hindsight phase, the same six songs were played in a different order. Before playing the song, we informed children about the content of the song's "hidden message" (target knowledge). The instructions differed between tasks. In the memory task, they were asked to recall their previous foresight estimates. By contrast, in the hypothetical task, children were asked to estimate – but ignoring the provided target feedback (and the "clues" from the priming phase as well) – how many of a group of six naïve peers would identify the message. At the end of each hindsight phase, children answered the two instruction comprehension questions (see Materials section).

Results

Identification performance

To explore whether priming increased children's identification performance (as a check on the effectiveness of the priming procedure), we conducted a 2 (priming: no-priming, priming) x 4 (grade: 3rd, 4th, 5th, 6th) mixed ANOVA. We only included those participants in this analysis who had given a positive answer to the message identification question and had attempted to write the message down (n = 175; 98% of our sample). Three independent raters then classified these written messages, awarding 1 point for full identifications (e.g., "There are no tomatoes in your garden"), 0.5 points for partial identifications (e.g., "There are tomatoes in your wardrobe") and 0 points for incorrect identifications (e.g., "He wears jeans"). The internal consistency (Cronbach's alpha) of the researchers' classifications was .98. A fourth rater resolved disagreements.

From these ratings we created an index of identification performance (from 0 to 6) for each participant; a rating of 6 means that participants correctly identified all six messages in the respective condition. The analysis yielded a strong main effect of priming, $F(1, 171) = 364.23$, $p < .001$; $\eta 2 = .68$. The identification index was much higher for priming (M = 1.96; SD = 1.29) than for no-priming songs (M = 0.13; SD = 0.31).

Further, the priming x grade interaction was significant, $F(3, 171) = 5.01$, $p < .001$; $\eta 2 = .08$. The trend analysis showed a significant linear component, $M = 0.66$, 95% CI [0.28, 1.05], showing that correct identifications in the priming condition increased with age. Nevertheless, in all grades priming increased children's message identification success.

Hindsight bias

In the memory task, following common procedure, we removed perfect matches between foresight and hindsight ratings given that these cases of veridical recollection distort the memory hindsight bias index (e.g., Erdfelder & Buchner, 1998; Pohl, 2007). Thus, 20% of responses were removed from this analysis. Note that in our study the retention interval was very short; therefore, the large proportion of veridical recollections is not surprising.

Children's (remembered, in the memory task) estimates of the number of peers that would identify the "hidden message" were submitted to a 2 (task: hypothetical, memory) x 2 (knowledge: foresight, hindsight) x 2 (priming: no-priming, priming) x 4 grade (3^{rd}, 4^{th}, 5^{th}, 6^{th}) mixed ANOVA. Results are displayed in Table 5 and Figure 3. The three-way interaction between task, priming and grade was significant, F (3, 139) = 4.72, p = .004 .01, $\eta 2 = .10$. Thus, we decided to analysed participants' responses to each task independently.

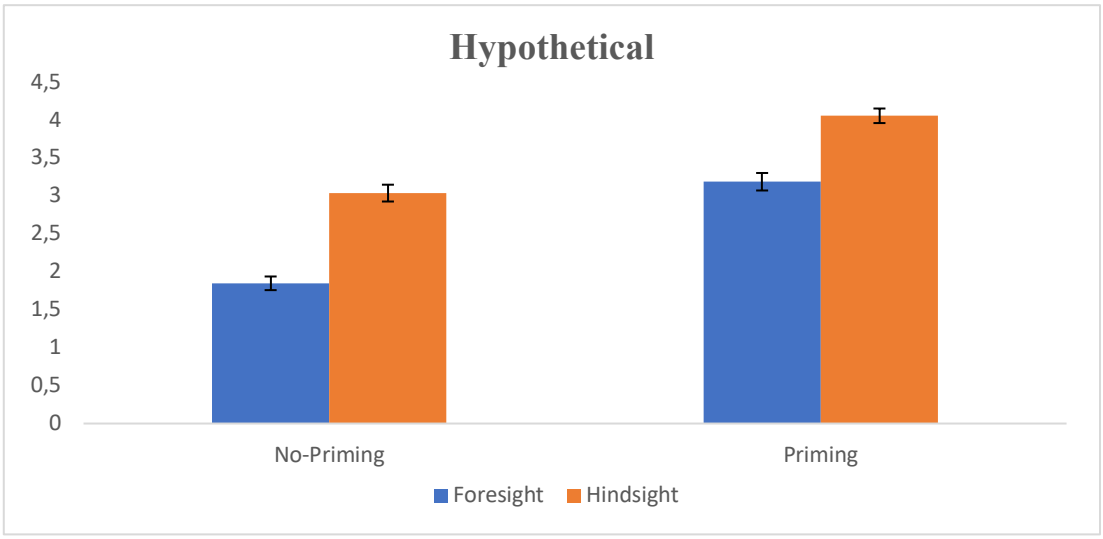
Hypothetical task

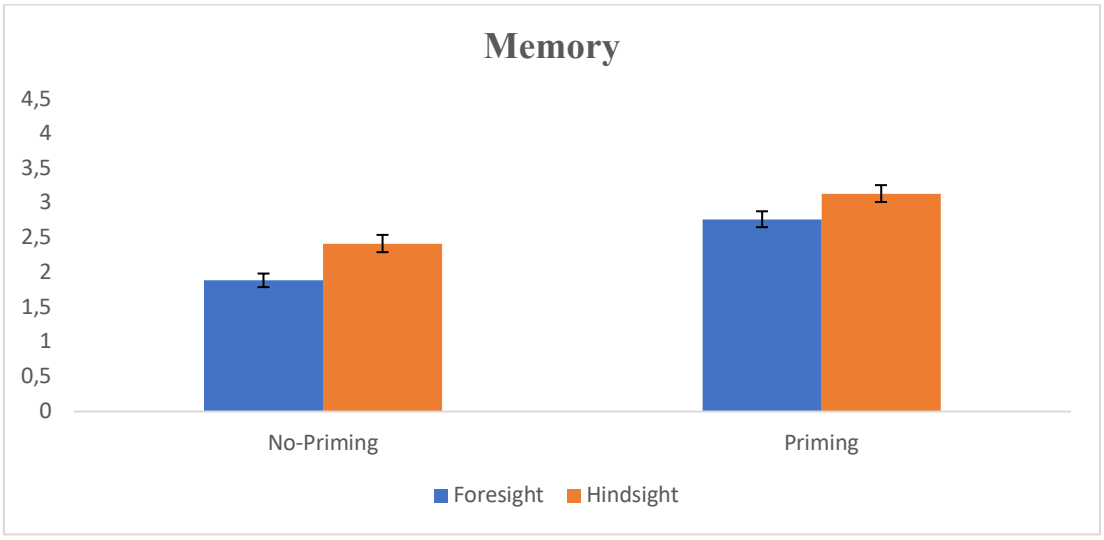
Participants' ratings were subjected to a 2 (knowledge: foresight, hindsight) x 2 (priming: no-priming, priming) x 4 (grade: 3^{rd}, 4^{th}, 5^{th}, 6^{th}) mixed ANOVA. The analysis showed main effects for knowledge, $F(1, 174) = 103.12$, $p < .001$, $\eta 2 = .37$, and for priming, $F(1, 174) = 162.58$, $p < .001$, $\eta 2 = .48$. Participants claimed that a higher number of peers would identify the message when they were informed about the content of the message just before hearing the song ($M = 3.54$; $SD = 1.21$) than when no information was provided ($M = 2.52$; $SD = 1.06$), showing auditory hindsight bias. Similarly, children provided higher ratings for priming ($M = 3.62$; $SD = 1.09$) than for no-priming songs ($M = 2.44$; $SD = 1.07$). Importantly, the interaction between knowledge and priming was significant, $F(1, 174) = 6.91$, $p = .009$, $\eta 2 = .03$. Further analysis showed that the hindsight effect was greater for no-priming than for priming songs, although both effects were significant on their own, $F(1, 177) = 92.93$, $p < .001$, $\eta 2 = .34$ and $F(1, 177) = 42.42$, $p < .001$, $\eta 2 = .19$, respectively.

The knowledge x grade interaction was also significant, $F(3, 174) = 8.53$, $p < .001$, $\eta 2 = .12$. The trend analysis showed a significant linear component, $M = -1.04$, 95% CI

[-1.47, -0.61]. With the exception of 6th graders, $F(1, 45) = 2.45$, $p = .124$, $\eta 2 = .05$, all children showed significant hindsight bias. Finally, there was a weak (and rather uninteresting and difficult to interpret) priming x grade interaction, $F(3, 174) = 5.85$, $p < .01$, $\eta 2 = .09$, due to larger priming effects in 4th and 6th graders (see Table 5).

Figure 3. Mean peer identification ratings in foresight vs. hindsight and in the hypothetical (top panel) vs. memory tasks (bottom panel) for no-priming and priming songs. Error bars represent SEM.





Memory task

The same analysis was performed for participants' ratings in the memory task[1] (see Table 5) and revealed main effects of knowledge, $F(1, 139) = 29.22$, $p < .001$, $\eta 2 = .17$, and priming, $F(1, 139) = 58.65$, $p < .001$, $\eta 2 = .29$. Participants' (remembered) ratings were higher in hindsight ($M = 2.78$; $SD = 1.24$) than in foresight ($M = 2.33$; $SD = 1.05$), showing auditory hindsight bias in the memory design. Ratings were also higher for priming songs ($M = 2.96$; $SD = 1.27$) than for no-priming songs ($M = 2.15$; $SD = 1.11$). Further, there was a marginally significant interaction between knowledge and grade, $F (3, 139) = 2.52$, $p = .06$, $\eta 2 = .05$. The trend analysis showed a significant linear component, M = - 0.47, 95% CI [-0.82, -0.11]. Again here, with the exception of 6th graders, $F(1, 37) = 1.64$, $p = .208$, $\eta 2 = .04$, all children showed significant hindsight bias. Post-hoc comparisons revealed that 3rd and 4th grade children exhibited more bias than 6th graders: $F(1, 66) = 7.30$, $p = .009$, $\eta 2 = .10$ and $F(1, 72) = 6.49$, $p = .013$, $\eta 2 = .08$, respectively. Crucially, the theoretically important priming x grade interaction was *not* significant, $F(3, 139) = 1.91$, $p = .316$, $\eta 2 = .02$).

[1] Remember we excluded 20% of cases from this analysis due to veridical recollection. Including all participants, however, produces the main principal pattern of findings: Main effects for knowledge, $F(1, 174) = 27.82$, $p < .001$, $\eta 2 = .13$, and for priming, $F(1, 174) = 86.63$, $p < .001$, $\eta 2 = .33$. Importantly, the interaction between these two factors was not significant either, $F(1, 174) = 1.72$, $p = .191$, $\eta 2 = .01$.

Table 5. Means peer identification ratings (SDs) for non-priming and priming songs in foresight and hindsigh
magnitudes in the hypothetical and memory tasks.

	Hypothetical					M		
	Foresight		Hindsight			Foresight		
	No-Priming	Priming	No-Priming	Priming	Bias	No-Priming	Priming	P
Grade 3	1.82 (1.23)	3.20 (1.83)	3.90 (1.60)	4.75 (1.21)	1.81*** (1.69)	2.15 (1.41)	3.16 (1.16)	
Grade 4	1.48 (1.13)	3.50 (1.56)	2.93 (1.54)	4.27 (1.23)	1.10*** (1.58)	1.60 (1.15)	2.13 (1.57)	
Grade 5	1.97 (1.22)	2.58 (1.28)	3.00 (1.20)	3.76 (1.18)	1.11*** (1.21)	1.97 (0.92)	3.01 (1.24)	
Grade 6	2.11 (1.18)	3.56 (1.35)	2.52 (1.36)	3.63 (1.26)	0.26 (1.10)	1.86 (1.19)	2.80 (1.29)	
Total	1.85 (1.20)	3.19 (1.54)	3.04 (1.48)	4.06 (1.28)	1.03*** (1.48)	1.89 (1.17)	2.77 (1.37)	

Note: * p < .05. **p < .01. ***p < .001

General discussion

In two experiments, we investigated auditory hindsight bias in children aged between 8 to 13 years, using a methodology similar to the one used in research with adults (e.g., Bernstein et al., 2018; Birch et al., 2017; Higham et al., 2017). Results showed that children exhibit auditory hindsight bias in the context of hypothetical (Exps. 1 and 2) and memory (Exp. 2) designs. Children not only overestimated the number of peers that would identify a "hidden message" but also distorted their recollections about their previous foresight ratings when they had been informed about the message content.

In Exp. 1, we also investigated if auditory hindsight bias can be reduced by experimental means. For this purpose, like Bernstein et al. (2018), we included a condition where children made *two* hypothetical judgements regarding the same song (two-presentations condition), one in foresight and another in hindsight. Results were consistent with the majority of findings in adults (Bernstein et al., 2012, Exp. 2; Bernstein et al., 2018, Exp. 3): hindsight bias persisted. It seems therefore that in childhood, too, hindsight bias is a robust phenomenon and hard to overcome. Despite largely similar findings in this condition and the 'traditional' hypothetical condition (i.e., without foresight judgements), however, there remained a principal ambiguity regarding the cognitive processes involved: The two-presentations condition combines elements from both hypothetical designs and memory designs, and previous research on auditory hindsight bias with adults (Bernstein et al., 2018; Higham et al., 2017) has traced the bias in these designs to different underlying processes – fluency processes in hypothetical designs and memory processes in memory designs.

The goal of Exp. 2 was then to use the priming methodology of the above research to resolve this ambiguity and contrast the two-presentations condition with a proper memory design. Using repetition priming (Bernstein et al., 2018), we manipulated the fluency with which the "hidden messages" were processed and asked children to complete auditory hypothetical and memory tasks. The findings pointed to a role of fluency processes in the two-presentations (hypothetical) task but not in the memory task. Consistent with the idea that priming increases the fluency processing of auditory targets and, in turn, reduces the subjective impact of target knowledge in hypothetical designs (Bernstein et al., 2018; Higham et al., 2017), auditory hindsight bias in the two-presentations condition was modulated by priming (paralleling Higham et al., 2017) – the bias was greater for no-priming than for priming songs. By contrast, there was no such modulation in the memory task. In turn, this suggests that the two-presentations condition

is, on balance, more akin to a hypothetical design than to a memory task. And finally, the findings of Exp. 2 demonstrate that the differential pattern of cognitive processes linked to the two hindsight bias designs in adults (i.e. processing fluency in hypothetical hindsight bias and remembering/reconstruction in memory hindsight bias) is already present in childhood.

Developmental trajectories

Disregarding minor differences between Exps. 1 and 2 in the shapes of the trajectories, we found an overall decline of auditory hindsight bias, in both hypothetical and memory designs, across our age range (8-13 years), with 3rd and 4th graders (8-10 years) showing the strongest bias. Thus, different from previous research (e.g., Bernstein et al., 2011; Pohl et al., 2010), we could demonstrate that hindsight bias continues to decrease after the age of 6 years. This decrease was also present in our visual identification task from Experiment 1 and suggests that the trend does not depend on the type of material (although other types of stimuli might need to be studied in order to be more confident about this generalisation).

There was an important difference, however, between previous research (Bernstein et al., 2004; 2007; 2011) and ours – the bias measure. Those previous studies used a more "implicit" measure of the bias (i.e., the point at which a naïve other would be able to identify a degraded object that progressively becomes clearer). By contrast, we asked children to provide an explicit numerical judgment (i.e., how many peers would identify the "hidden message"): children were asked to ignore their actual experience and select from available options (zero peers, one peer, etc.) the one that fits best with their subjective impression. This type of measure may be more sensitive to developmental changes in metacognitive abilities that occur around this age (Flavell et al., 1995; Flavell et al., 1997; Siegler, 1996).

Our tasks also differed from previous research in terms of the salience or identifiability of the source of target knowledge. In previous research, target knowledge resulted from foresight experience with the target (e.g., an object that progressively becomes clearer). By contrast, in our study this information was explicitly provided by the experimenter, rendering the source more salient than in previous research. Also, older children are generally better at accurately attributing the origins of their knowledge (e.g., Drummey & Newcombe, 2002, Ruffman et al., 2001; Sluzenski et al., 2004). In combination, the developmental trajectory of the bias in our research may thus reflect

improving source attribution ability in older children, facilitated by heightened salience of the source, due to the explicit experimenter provision of target information.

Differential source salience could also explain why, in the hypothetical task, the bias disappeared in 6th graders in Exp. 2, while it was present in this grade in Exp. 1. In Exp. 2, the foresight and the hindsight judgements were provided in two different phases, making the role of target knowledge in message identification more salient than in Exp. 1. This should have facilitated source attribution for older children, making them less prone to succumb to auditory hindsight bias.

In the memory task, 3rd, 4th and 5th graders showed auditory hindsight bias. Previous research in adults has shown that the effect size in memory tasks depends on the retention interval: the longer the interval, the larger the bias (e.g., Blank et al., 2003). In our study, the retention interval was very short: Only a brief priming phase separated the hindsight phase from the foresight phase. Moreover, the number of judgements that children were asked to recall was small (only six), as well as the choice of possible answers (a number between 0 to 6). As older children have better memory skills (e.g., Cowan, 1997; Goswami, 2020; Keil, 1989; Schneider & Pressley, 1997), this might explain the disappearance of the effect in 6th graders.

Conclusions

Our two experiments extended previous research on a phenomenon recently discovered in adults, auditory hindsight bias, to school-age children. We could establish that (1) children of this age show strong auditory hindsight bias and (2) this hindsight bias is as robust against debiasing attempts as in adults. Moreover, (3) the bias shows the same pattern of underlying cognitive processes (fluency and memory processes) as in adults. Finally, (4) partly different from previous conclusions in research with other modalities, auditory hindsight bias declined across our age range. Future research should continue exploring the fascinating phenomenon of hindsight bias in children.

Open Practices

The two experiments in this article earned an Open Data badge for transparent practices. Data for the experiments are available at https://osf.io/a76ct/?view_only=e18b8871c6d44e2c9473505ef065c96a

Capítulo 5

Responsibility attribution by children and adults with mechanical devices

Abstract

In two experiments, we tested the reasoning of children and adults about the functioning of mechanical devices, using the structural model of outcome responsibility. We created a task in which participants were informed about the functioning of four wires in a battery. They had to decide whether a wire was critical in order for it to function and pivotal for the outcome. The structural model establishes that responsibility is computed by two independent factors related to forward and counterfactual inferences, respectively. In Experiment 1, we replicated previous results for adults with the new task and in Experiment 2, the results showed that children used the same mechanism of attribution of responsibility as adults, with older children making better judgments than younger ones. We also obtained a correlation between the pivotality parameter and an independent measure for counterfactual thinking. The results established that older children improved their judgment in the parameters predicted by the structural model and that judgment is related to an improvement in counterfactual thinking.

Keywords: Responsibility attribution, pivotality, criticality, development of causation, counterfactual thinking.

Pre-print version of the manuscript: Gordo, C., Gómez-Sánchez, S., & Moreno-Ríos, S. (2021). Children and adults' responsibility attribution in mechanical devices. Manuscript submitted for publication.

Introduction

Causal reasoning is developed in very young children (Muentener & Bonawitz, 2017). However, understanding the function of devices in which different causes can operate is more complex, not being achieved until late childhood or even adolescence. Inhelder and Piaget (1958) believed that the ability to reason systematically about hypotheses is achieved only when individuals reach the formal operation stage. Siegler (1978) showed that children had great difficulty making predictions and understanding the function of a balance scale when they had to interrelate information about the weights and the distance of the weights from the fulcrum. Some other studies have found ways to enable younger children to give correct responses to these problems (e.g., see Amsel et al., 1996; Wilkening & Anderson, 1991). However, the comprehension of how mechanical devices function and the explicit understanding of how data can support or falsify possible hypotheses in general seem to present a challenge for the development of causal reasoning and scientific reasoning in late childhood (see Klahr, 2000). In order to detect potential causes of a malfunction in a mechanical device, we must mentally undo actions in order to obtain different outcomes. This requires thinking about how things could have happened, which is known as counterfactual thinking. Pre-schoolers show some limitations in this ability, which develops gradually during the school years until around the age of 12 to 14 (Rafetseder et al., 2013, 2021; Gómez-Sánchez et al., 2020).

School experiences of learning, particularly in the later grades, help children to monitor in a more explicit way their process of problem solving. Comparing children of intermediate primary ages with those of an older age could help us discover developmental differences in reasoning about the functioning of devices. In this study, we compare children between 8 and12 years from the second level of Spanish primary school (2nd Level PS; 3rd and 4th grades) and those of the third level (3rd Level PS; 5th and 6th grades) with adults in how they attribute responsibility to components for the functioning of a device. We also test the possible relation of the attribution of responsibility with counterfactual thinking in children.

Causal attribution

On 16th November 2020, the Ingenio satellite was launched into space from the Kourou aeronautical base in French Guyana by the Vega VV17 rocket. Eight minutes after the Vega rocket took off, a problem occurred with the rocket's controls and, after breaking up in the atmosphere, it fell into the sea. An independent committee of inquiry determined that the cause of the rocket's accident was several wrongly connected wires.

But what was the real cause of the accident? According to the standard counterfactual model of causation (Lewis, 1973, 1986), there are two requirements needed to establish that a cause "X" causes an outcome "Y". Firstly, X and Y should both be true. And secondly, there must be a counterfactual dependence between "X" and "Y". That is, if "X" had not occurred, then "Y" would not have happened. We will call this a counterfactual test. Let us imagine that there were four badly connected wires in the Vega rocket (A, B, C, and D) and that each one is crucial to the rocket's safe flight. In this case, the four wires fulfilled the first requirement: they were badly connected and the rocket had an accident. However, none of them fulfilled the second. Thus, if we submit, for example, wire A's action to the counterfactual test, we realise that the rocket's accident would have occurred even if wire A had been properly connected. The Vega rocket case illustrates the over-determination problem, one of the main limitations of the standard counterfactual model of causation (Lewis, 1973, 1986): although adults established a causal relation between wire A and the accident, this was not predicted by the model.

More recently, Lagnado and colleagues (2014, 2015) developed a psychological model of outcome responsibility that deals with over-determination cases: the structural model of outcome responsibility, which is based on the standard account of causation (Halpern & Pearl, 2005) and causal responsibility research (Chockler & Pearl, 2004). The core idea in the model is that outcome responsibility judgments are products of prospective and retrospective factors. Prospective factors are termed 'criticality' in the model while retrospective ones are called 'pivotality'. Criticality is evaluated before knowing the outcome and reflects the importance of an agent's contribution to reaching a joint outcome. Criticality is defined as "the probability of an agent A_i's contribution x_i in a situation S that is necessary for the positive group outcome y" (Lagnado et al., 2015, p. 222):

$$\text{Criticality } (A_i, S) = \frac{p\left(\frac{y}{x_i}\right) - p\left(\frac{y}{\neg x_i}\right)}{p\left(\frac{y}{x_i}\right)}$$

In the Vega rocket case, if one wire is badly connected, the rocket does not work. Therefore, each wire is fully critical (C = 1), given that the probability of having a safe flight if any of the four wires are badly connected is zero. This structure is called a conjunctive structure. However, if we considered a hypothetical, and safer, previous

design of the rocket in which just one of the four wires had to be properly connected to ensure a safe flight, each wire's criticality would be reduced to ¼. This is called disjunctive-structure.

Pivotality, however, is assessed once the outcome is known and reflects how close an agent's action was to being counterfactually dependent on the outcome. So pivotality reflects the notion of counterfactual dependence included in Lewis's model of causation (Lewis, 1973, 1986) but with a "key" change: the counterfactual dependence is not evaluated in a dichotomous way (a cause may or may not be counterfactually dependent on the outcome) but as a graded measure. In particular, the counterfactual dependence is measured in terms of how many hypothetical interventions are required in a situation to make the outcome "Y" counterfactually dependent on the cause "X". Thus, the more hypothetical interventions are required, the less counterfactually dependent is the outcome "Y" on cause "X"; and therefore, the less responsible is "X" for "Y" (Lagnado et al., 2014, 2015). Pivotality is calculated from the number of changes required in the actual situation to make the agent's action pivotal for the outcome. That is, to make the outcome counterfactually dependent on agent A's action. Pivotality is computed with the following formula (Lagnado et al., 2015):

$$\text{Pivotality (A, 0, S)} = \frac{1}{(N+1)}$$

In the Vega rocket case, 0 is the rocket's accident, S describes the causal structure of the situation (all the wires must be properly connected to avoid the accident – conjunctive structure) and N the number of wires that should have been properly connected to change the outcome, for example, the wire A is pivotal for O (the accident). In our case, with four wires badly connected, three changes would have been needed to make the wire A pivotal for the rocket's accident. So wire A's pivotality equals 1/3. We have to mentally change the failures in the other three wires to make wire A a determinant in the result. However, in the hypothetical situation where just one wire had to be properly connected to avoid the accident (disjunctive structure), wire A would be fully pivotal to the outcome (P=1).

In the present work, we aim to explore whether the structural model of outcome responsibility can explain children's responsibility judgements about the components of

mechanical devices. That is, whether children's attribution of responsibility is mediated by the computation of both criticality and pivotality. To this end, we wanted to test whether pivotality judgements depend on counterfactual thinking, as the structural account posits (e.g., Lagnado et al., 2014, 2015). If this is so, children's pivotality judgements should be related to their counterfactual thinking abilities.

The development of counterfactual thinking

As we have seen, the structural account assumes that pivotality judgements involve counterfactual thinking. That is, they require envisaging how things could have turned out differently (Byrne, 2016). Until the first decade of the 21st century, some researchers found that primary schoolchildren (Burns et al., 2012; O'Connor et al., 2014; Weisberg & Beck, 2012) and even pre-schoolers (Beck et al., 2006; Guajardo et al., 2009; Roldán-Tapia et al., 2017) were able to understand and make correct inferences from counterfactual conditionals. However, recent findings suggest that children might provide correct responses to the counterfactual question not because they have understood the counterfactuality, but because of an artefact in the methodology testing the ability that masked some limitations in children (Rafetseder et al. 2013, 2021; Gómez-Sánchez et al., 2020). For example, a correct interpretation of the counterfactual "if the wire had not been broken, the accident would not have happened" implies that there is a conjectured situation (the wire was not broken and the accident did not happen) and a presupposed fact (the wire was broken and the accident happened). When children are informed that "the wire was not broken", they conclude that "the accident did not happen". This response was interpreted as proof of children's counterfactual reasoning ability. However, the same conclusion can be drawn without reasoning counterfactually. Actually, Rafetseder et al. (2010) proposed that children understood the counterfactual as a biconditional "if the wire was broken, the accident happened" but also "if the wire was not broken, the accident didn't happen".

Several studies have tried to solve this question by including "discriminative conditions" where a biconditional interpretation of a counterfactual conditional led to a different and wrong answer. For example, Rafetseder et al. (2013, 2021) presented children (5 to 10 years old), adolescents (13 to 15 years old) and adults with counterfactual conditionals (e.g., "If A had happened, then B would have happened") that included a piece of contextual information that cancelled the causal relation between antecedent and consequent ("B happened anyway"). This piece of contextual information has the same effect as using just a semifactual expression "Even if A had happened, B

would have happened". Semifactual conditionals are a subtype of counterfactual conditionals that seem to cancel the causal link between antecedent and consequent (Rodriguez Rosique, 2001; Schwenter, 2001). The conclusion in both cases is that "B happened". This manipulation is based on the fact that a basic conditional interpretation of the counterfactual expression without attending to the contextual information would lead to a wrong answer ("B did not happen"). Thus, it would allow us to distinguish responses given with counterfactual reasoning from those given using the biconditional representation.

As expected, pre-schoolers responded correctly to counterfactual questions for basic conditional reasoning problems without the contextual information, which was consistent with previous studies, but only adolescents and adults could do so in a discriminative condition with the contextual information, which requires more complex counterfactual thinking. Another piece of evidence in favour of children's biconditional interpretation of counterfactual expressions was provided by Moreno-Ríos and García-Madruga (2002). These authors presented children (7 to 11 years old), pre-adolescents (13 to 14 years old) and adults with semifactual conditionals. These authors found that only pre-adolescents made inferences similar to those made by adults. More recently, Gómez-Sánchez et al. (2020) explored children's interpretation of counterfactual expressions, using semifactual conditionals that also included some contextual information that could annul the relation established in the statement. The authors tested children aged from 7 to 12 years and adults. Their results suggested that children do not interpret semifactual expressions in the same way as adults until the end of primary school, around age 12. Taken together, the results from these studies (Gómez- Sánchez et al., 2020; Moreno-Ríos & García-Madruga; Raftseder et al., 2013, 2021) suggest that children's counterfactual thinking abilities, and particularly thinking with semifactuals, improves during the school years. Thus, if the ability to think counterfactually is still developing in childhood and the attribution of responsibility depends on it, changes in the attribution of responsibility could be expected during childhood. This possibility has not been tested, as far as we know.

A task to test the structural model of outcome responsibility

To test the structural account, Lagnado and colleagues (2014, 2015) developed the dot-click game task where adult participants were asked to click on a dot that appeared on a computer screen a specific number of times during a default period. Each participant played with three hypothetical players. Players' criticality was manipulated by varying

the causal structures of the team's challenges. Three different causal structures were included: conjunctive, disjunctive and a mixture of the two. In the conjunctive structures, all the players needed to succeed in their task for the team to win. By contrast, in the disjunctive structures, the team won if at least one of the players succeeded in their task. In the mixed structures, two players (A and B) had to succeed and also at least one of the remaining two players (C or D). Participants were presented with four different challenges on the screen and were asked to estimate "How critical is Player A for the team's outcome in each challenge?" (Lagnado et al., 2014). Results showed that participants rated Player A's contribution as more critical for the outcome in the conjunctive challenges (where his/her success was necessary for the team to win, C = 1) than in the disjunctive challenges (where the player's success was not necessary for the team to win, C = 1/2).

In the second phase of the experiment, participants viewed on the screen the results of nine challenges and were asked to indicate "How responsible was Player A for the team's results in the different situations?" (Lagnado et al., 2014). Player A's pivotality was manipulated by varying the number of players who succeeded in their task. Thus, there were cases where Player A was fully pivotal to the team's loss (P = 1) and other cases where his/her pivotality was reduced by increasing the number of players who also failed in their task (an additional player, P = ½: two additional players, P = 1/3). Results showed that participants held Player A more responsible for the team's loss in the challenges where Player A was pivotal to the outcome (P = 1) than in those where Player A's pivotality was lower (P = 1/2 and P = 1/3). Furthermore, results showed that pivotality judgements were also influenced by Player A's criticality. In cases where Player A was pivotal for the team's loss (P = 1), participants attributed more responsibility to Player A when s/he was fully critical to the team's victory (C=1) than when s/he was not (C = ½ and C = 1/3). More specifically, results showed that when pivotality was held constant, Player A's responsibility increased with criticality. These results support the structural account of responsibility attribution, showing that individuals take into account both criticality and pivotality when they judge how responsible an agent is for a joint outcome.

The structural account has been also tested in the context of abstract mechanical devices. Lagnado and colleagues (2015) mentioned an unpublished study in which participants were presented with the same causal structures and the same outcome patterns as described previously. However, instead of agents, they asked participants for several components of a machine. The participants' task was to estimate the extent to

which the different components of a machine caused the machine to pass or fail the trial. Results demonstrate that the structural account can also explain individuals' responsibility judgements about abstract mechanical devices. This suggests that adults attribute responsibility to human agents and mechanical devices in a similar way. It is not clear whether children attribute responsibility to mechanical devices in the same way as adults. In Experiment 1, we aimed to replicate Lagnado and colleagues' (2014) previous findings with adults, using our task designed to test children. In Experiment 2 we used the same task with children to test whether the structural model of responsibility can account for children's responsibility attribution about mechanical devices and whether the ability develops and is related to counterfactual reasoning.

Experiment 1

In order to explore whether the structural account can explain children's responsibility judgements about mechanical devices, we created a new task based on how a plug-in battery works. We tested adults to see whether the results were comparable to those obtained previously with other tasks. In this task, the battery had four wires: green, blue and two special wires, orange and yellow. The battery worked if the green, blue and at least one of the two special wires, either yellow or orange, worked correctly. Thus, criticality was manipulated by asking participants for a wire whose contribution was fully critical for the battery's functioning (the green wire, C = 1), or asking them for a wire whose contribution was not fully critical (the yellow wire, C = ½). Pivotality was manipulated by varying the number of wires that worked correctly when the battery was not working. In particular, there were trials where the target wire (e.g. the green wire) was fully pivotal to the battery's malfunctioning (P = 1) because it was the only wire that did not work properly. In other trials, the wire was not fully pivotal (P = ½) because a second wire (e.g. the blue wire) did not work correctly either. The participants' task was to estimate how critical the target wire's contribution was to a positive outcome (battery functioning) and, how responsible the target wire was for the negative outcome (battery malfunctioning). For both judgements, participants provided their answers on a non-numbered scale ranging from 0 to 10, whose endpoints were labelled "not at all" and "very much".

Based on previous findings (Lagnado et al., 2014), we expected participants to rate the target wire as more critical when its contribution was fully critical for the positive outcome (C = 1) than when it was not (C = ½). Similarly, we expected participants to

hold the target wire more responsible when its action was fully pivotal for the negative outcome (P = 1) than when it was partially pivotal (P = ½). We also expected a joint effect of the criticality and pivotality parameters, so that participants' responsibility judgments when they were asked about pivotality would be influenced by the target wire's criticality. In particular, we expected that when the wire's action was pivotal for the negative outcome, participants would hold the target wire more responsible when its contribution was also fully critical (C = 1) than when it was less critical (C = ½).

Finally, the structural model has some implications for how the two parameters are tested and the order of criticality and pivotality problems. Criticality judgements are made without considering the outcome, and therefore no influence of pivotality computation is expected on criticality. If the two parameters are independent, the order of presentation of trials would not influence the magnitude of the two parameters.

Method

Participants

A total of forty-seven students (10 males and 37 females) from the University of Granada (age: *M* = 21.20; *SD* = 3.71) took part in Experiment 1. They participated in exchange for course credits. Twenty participants completed criticality trials first, while the remaining eighteen participants started by completing pivotality trials. For this study, participants read and filled out consent forms complying with the University Research Ethics Committee guidelines (*Comité de ética de investigación humana de la Universidad de Granada* specifically approved for this study: 1068/CEIH/2020).

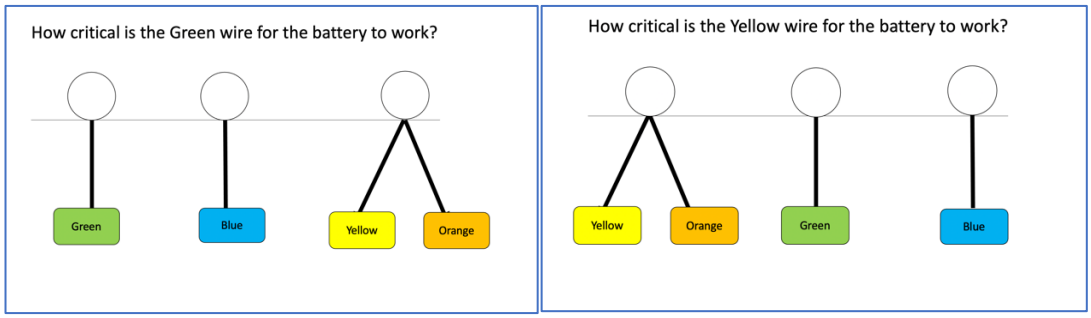
Materials

Eight slideshow presentations were prepared. Each presentation included six criticality judgements and six pivotality judgements. In half of the presentations, the criticality trials appeared first while in the other half they appeared after those for pivotality.

In the criticality trials, participants were presented with an image of the four wires (see Figure 1). Each wire was identified with a colour: green, blue, yellow, and orange. The two latter wires were referred to as "special wires". In four out of the six criticality trials, we asked participants for a wire whose contribution was fully critical for a positive outcome (e.g. the green wire, C = 1), whereas in the remaining two trials we asked them for a special wire whose contribution was not fully critical (e.g. the yellow wire, C = ½).

In both cases, the participants' task was to estimate how decisive the target wire's action (either green or yellow) was for the battery to work.

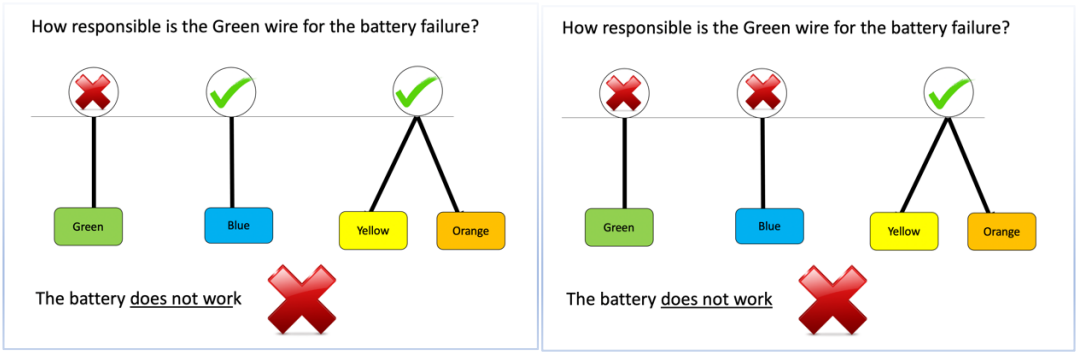
Figure 1. Screenshot of the two criticality value conditions by asking participants about the leftmost wire. The criticality in the left panel is 1 (green wire) and in the right panel ½ (yellow wire).



In the pivotality trials, participants viewed the same image as in the criticality ones, but in addition, they were shown the outcome: the wires that had worked correctly and those that had not. Specifically, when a wire worked properly, the wire turned green and a green tick appeared above it. However, when the wire did not work, the wire turned red and a red "X" appeared above it (see Figure 2). As can be seen in Figure 2, in both panels the leftmost cable has a constant criticality of 1. But in the left panel, its pivotality is 1, because if the wire had worked, the battery would have worked. It is maximally pivotal because nothing has to be changed. In contrast, in the right-hand panel, even if the green wire had worked, the battery would not have worked because the blue wire did not work. Thus, an additional change would have been needed to make the green wire pivotal to the outcome; therefore, in these cases pivotality is ½.

In two out of the six pivotality trials, the wire's action was fully pivotal (P = 1) and fully critical (C = 1) to the battery's malfunctioning. In the other two trials, the wire was also fully pivotal to the outcome (P = 1) but its contribution was not fully critical (C = ½). In the remaining two trials, the wire was fully critical (C = 1) but its action was not fully pivotal to the outcome (P = ½). The number of conditions was set at the minimum to test the independence between criticality and pivotality and thus reduce as much as possible the number of cases to be answered.

Figure 2. Screenshot of the two pivotality value conditions by asking participants about the leftmost wire. The pivotality in the left panel is 1 (green wire) and in the right panel ½ (green wire).



In all the presentations, we included a "ball task" before the experimental trials. The purpose of this task was to familiarise participants (especially children, in the next experiment) with the meaning of the scale. The Ball task included three trials. In these trials, participants viewed three urns that included balls of two different colours, with different proportions of them in each. Participants were asked to estimate how possible it was that a blindfolded child would pull out of the urn a ball with a specific colour. After this task, the experimental trials (either criticality or pivotality) appeared.

Participants provided their answers in a booklet containing 15 non-numbered scales (ranging from 0 to 10) whose endpoints were labelled "not at all" and "very much". In three of them, they provided their answers for the Ball task and in the remaining twelve, for the Battery task. Each scale was preceded by the trial number (e.g., Question 1A). This number as well as the trial's question (either criticality or pivotality) also appeared on the screen.

Procedure

Participants were tested in a quiet room, in groups of three to five. We started each experimental session by explaining the meaning of the scale. After that, and to ensure that participants understood the use of the scale, they completed the Ball task. Following this, the experimental trials started.

Regardless of whether participants started by answering criticality or pivotality trials, we told them: "*Chinese researchers have developed a special plug-in battery. It is special because it operates for 72 hours per charge and it takes just 20 minutes to load. However, researchers continue testing because they do not get to run the battery. This battery has four wires: a green wire, a blue wire and two special wires, one yellow and the other*

orange. The battery runs if the green and blue wires work and at least one of the special wires, either the yellow or the orange. For that reason, the yellow and the orange wires are special: only one of them needs to work to run the battery as long as the green and blue wires also work". Once the experimenter was sure that participants had understood the instructions (how many wires had to work to run the battery), they completed the experimental trials.

Before completing the criticality trials, we explained to participants what "critical" ("decisive" in Spanish) means. This explanation was included to ensure that children, in particular (in the next experiment), understood the meaning of "critical". Nevertheless, we also provided this explanation to adults. We used a well-known Spanish cooking show to illustrate what critical means. In this show, three judges decide who is the best cook among several candidates. We told participants that "critical" refers to "*the 'power' that someone has to make that thing happen or not*". After that, we asked them to envisage that in this TV show there was only one judge and we asked them how critical this judge's vote was to decide the winner. Participants provided verbal answers and we told them: "*in this case, the vote of Chicote (the judge) was critical because he had the power to decide who would be the winner*". Next we asked participants whether Chicote's vote would be more or less critical in the actual situation (in which there were three judges) than in the hypothetical situation mentioned previously. If a child claimed that Chicote's vote would be more critical in the actual situation than the hypothetical one, they received a further explanation. Subsequently, they completed the criticality trials.

Pivotality trials began with the following instructions: "*Now I will show you the results of some trials by Chinese researchers*. *Specifically, I will tell you which wires worked in each of these trials. Your task is to estimate how responsible one wire is for the battery's malfunctioning*. Here again, the trials appeared one after another, and participants were asked to rate how responsible the target wire was for the battery's malfunctioning.

Results and discussion

Criticality judgements

Participants' ratings of criticality trials were submitted to a 2 (value: ½, 1) x 2 (order: criticality first, criticality second) mixed ANOVA. Results are shown in Table 1. The analysis showed main effects for value, $F(1, 36) = 166.25$, $p < .001$, $\eta 2 = .82$, and for order, $F(1, 36) = 6.78$, $p < .05$, $\eta 2 = .15$. Participants rated the wire's contribution as more critical when it was fully critical for the battery operation than when it was not. In

both cases, participants provided higher ratings when they completed the block of criticality trials first than when they completed them after the pivotality ones (6.42 vs 5.29). The interaction was not significant, $F(1, 36) = 2.51$, $p = .121$, $\eta 2 = .06$, and therefore, the magnitude of the criticality parameter was not influenced by the order effect. It is possible that only participants who had problems from the beginning with information about outcomes, would expect receiving information about the outcome, and would later be uncertain about responsibility, when they were not informed of them. This could explain the reduction of criticality judgments in pivotality-first block trials.

Pivotality judgements

The same analysis was performed for participants' ratings in the pivotality trials. The mean ratings are shown in Table 1. The analysis revealed a main effect for value, $F(1, 36) = 78.43$, $p < .001$, $\eta 2 = .68$: participants held the target wire more responsible when its action was fully pivotal for the negative outcome than when it was not. Neither order, $F(1, 36) < 0.01$, $p = .977$, $\eta 2 < .001$, nor the interaction, $F(1, 36) = 1.71$, $p = .286$, $\eta 2 = .03$, were significant.

Table 1. Mean ratings (SDs) for criticality and pivotality judgements (from 0 to 10) for the two values.

	Value ½	**Value 1**
Criticality	2.70 (1.91)	8.39 (2.15)
Pivotality	6.88 (1.84)	9.68 (0.74)

In order to test whether criticality had an independent effect on pivotality, we compared pivotality judgments with the two values of criticality, keeping the pivotality value constant (1). We also included the order of the tasks to test the potential influence of the previous explicit evaluation of criticality on pivotality judgements. Data were submitted to a 2 (value: ½, 1) x 2 (order: criticality first, criticality second) mixed ANOVA. The analysis revealed a main effect of value, $F(1, 36) = 132.73$, $p < .001$, $\eta 2 = .78$: participants held the target wire more responsible when it was also fully critical for the battery's functioning (C = 1) than when it was not (C = ½), $M = 9.68$, $SD = 0.74$ and $M = 5.53$, $SD = 2.09$, respectively. No other significant effects were found.

In Experiment 1, we tested adults with a new task designed to test children's attributions of responsibility, replicating Lagnado and colleagues' (2014) main findings. In particular, participants rated the wire's contribution as more critical when it was fully critical for the battery's functioning than when it was not. Similarly, participants held the

target wire more responsible when its action was fully pivotal to the battery's malfunctioning than when it was not. Finally, we found that criticality judgements influenced pivotality judgements. Specifically, we found that participants rated the target wire as more responsible when its contribution was fully critical for the battery's functioning than when it was not, although in both cases it was pivotal to the battery's malfunctioning.

Experiment 2

In this experiment, we tested whether children's responsibility judgements about mechanical devices were dependent on 1) how critical the wire's contribution is perceived to be in order to run the battery and 2) how close the functioning of a wire was to be fully pivotal for the battery's malfunctioning. Furthermore, we aimed to test whether, as the structural account posits (Lagnado et al., 2014, 2015), there is a relation between counterfactual thinking and pivotality judgements. To this end, we also asked children to complete a counterfactual reasoning questionnaire. Prior works have suggested that the ability to reason counterfactually continues to develop during mid-childhood (Gómez-Sánchez et al., 2020; Moreno-Ríos & García-Madruga, 2002; Rafetseder et al., 2013, 2021). This kind of thinking allows us to consider not only a different outcome that could have happened (counterfactual thinking) but also how the same outcome could have happened in different situations (semifactual thinking). Based on this, we hypothesised that if pivotality judgements rely on counterfactual thinking, then children with better counterfactual (and semifactual) reasoning abilities should be more sensitive in their judgments of the wire's pivotality. Sensitivity to pivotality was measured by the difference in children's ratings in two conditions: when the wire was fully pivotal to the outcome (P = 1) and when it was not (P = ½). Thus, a child would be fully sensitive to pivotality if s/he gave a rating of 10 in the trials where the target wire was pivotal to the outcome (P = 1) and a rating of 5 in the trials where the wire was not fully pivotal to the outcome (P = ½). The smaller the difference between responses to the two pivotality conditions, the lower the child's sensitivity to pivotality. Therefore, we predicted that older children should be more sensitive than younger children, given their better counterfactual reasoning abilities. Regarding the criticality component, no differences were predicted between older and younger children. We expected them to give higher ratings in trials where the wire's contribution was fully critical to the positive outcome (C = 1) than in trials where it was not (C = ½). The structural account states that the two

parameters are independent. Therefore, with the same value of criticality, different conditions of pivotality would lead participants to hold different values of responsibility judgements. The same effect of order was expected as for adults: higher rates for criticality when criticality trials were shown first, but without changes in the magnitude of the criticality parameter. No differences were expected for pivotality.

Method

Participants

One hundred and sixty-nine schoolchildren (83 girls and 86 boys) aged between 8 and 13 years (M = 9.85; SD = 1.25) completed the Battery task: seventy-nine children from the second level of Spanish primary school (2nd PS-Level) that includes 3rd and 4th grades (M = 8.81; SD = 0.70) and ninety children from the third level of Spanish primary School (3rd PS-level) that includes 5th and 6th grades (M = 10.76; SD = 0.87). One hundred and sixty-one of them (M = 9.89; SD = 1.24) also completed the counterfactual reasoning questionnaire: seventy-three from second PS-level (M = 8.85; SD = 0.70) and eighty-eight from third PS-level (M = 10.76; SD = 0.87). All of them were native speakers of Spanish and were enrolled in state schools in Granada, Spain. The sample was composed of all the students from second PS-level (composed of children in third and four grades) and third PS-level (composed of children in fifth and sixth grades) from two schools that accepted the invitation to participate in the study. They participated only if their parents had given written consent for this study, complying with the ethical protocol of the University Ethics Committee (*Comité de ética de investigación humana de la Universidad de Granada* specifically approved for this study: 1068/CEIH/2020).

Materials

The materials used for the Battery task were the same as those described in Experiment 1. However, in this experiment, we also asked participants to complete a counterfactual reasoning questionnaire. The questionnaire included a short story and four questions: an inferential question, an epistemic status question, a comprehension question and a semifactual question. There were two different versions of the counterfactual reasoning questionnaire. The two versions differed only in the contextual information (the short story). For example, in one of the versions, children read the following: "*María is on the phone with her husband Juan. María is in a shop because she wants to buy a new TV. She goes up to a shelf and sees a TV for sale. She says to*

Juan: "If somebody had pressed the remote control…" Just at that moment, the phone goes dead." The four questions then appeared.

1) In the inferential question, participants were asked: How do you think the conversation would have continued if the phone line had not gone dead? If somebody had pressed the remote control, then the TV… They had to choose between three answers a)…would have come on, b)… would not have come on, and c)… It's impossible to know.
2) In the Epistemic status question, they were asked: When Maria said "If somebody had pressed the remote control…", what do you think was really going on? The three options for response were: a) Somebody pressed the remote control, b) Nobody pressed the remote control, and c) There was no remote control.
3) The control comprehension question was: Where was María when she was talking on the phone? And the three answer options were: a) In her home, b) In a shop, and c) On the beach
4) Finally, the semifactual question was: Imagine that you are in a park and you hear somebody say: "I know what really happened and even if Ana had trained hard, she would not have won the competition." What really happened? a) Ana won the competition, b) Ana did not win the competition, c) It's impossible to know. The questions appeared in the same order as above.

Procedure

The procedure for the Battery task was the same as that described in Experiment 1. The counterfactual reasoning questionnaire was applied in groups of twenty to twenty-five children in their classroom. The experimenter asked the children to read the story carefully and answer the questions that appeared below. The children were encouraged to ask any questions they might have while completing the questionnaire. When a child asked some question related to the content and not to the procedure for response, the text in the story was read again, taking care not to give any clues about the answer.

Results and discussion

Criticality judgements

Children's responses to criticality trials were submitted to a 2 (value: ½, 1) x 2 (order: criticality first, criticality second) x 2 (PS-level: second, third) mixed ANOVA, with the first variable manipulated within participants and the other two between groups.

Results are shown in Table 2. The analysis revealed a main effect for the three factors: children rated the wire's contribution as more critical when it was fully critical for the battery's functioning than when it was not $F(1, 165) = 191.55$, $p < .001$, $\eta 2 = .53$. Judgement values were higher when trials of criticality were presented first $F(1, 165) = 7.08$, $p < .01$, $\eta 2 = .04$ and when they were given by younger children $F(1, 165) = 4.31$, $p < .05$, $\eta 2 = .02$. There was also a significant three-way interaction between these factors, $F(1, 165) = 4.66$, $p < .05$, $\eta 2 = .03$. To analyse the interaction, we computed the effect of criticality and PS-level for each order of presentation. Results showed that the criticality effect was present in both cases: when children completed criticality trials first (9.10 vs 5.19), $F(1, 79) = 135.38$, $p < .001$, $\eta 2 = .63$, and when they started with the pivotality trials (7.98 vs 4.81), $F(1, 86) = 69.11$, $p < .001$, $\eta 2 = .44$, although the magnitude was higher in the first case. The interaction of criticality with PS-level was also significant, but only when criticality trials were presented first, $F(1, 79) = 5.45$, $p < .05$, $\eta 2 = .06$. Although children from both second and third PS-levels were sensitive to criticality value, $F(1, 38) = 39.87$, $p < .001$, $\eta 2 = .51$ and $F(1, 167) = 105.85$, $p < .001$, $\eta 2 = .72$, older children were more sensitive to criticality value than younger ones. When they completed criticality trials after pivotality ones, there was just a main effect of PS-level, $F(1, 86) = 9.89$, $p = .002$, $\eta 2 = .10$. Again, there was no change in the magnitude of the criticality parameter with the order factor.

Table 2. Mean ratings (SDs) for the two criticality values per Primary School level.

	C = ½	**C = 1**
Second PS-level	5.39 (3.20)	8.71 (1.77)
Third PS-level	4.56 (2.66)	8.28 (2.28)
Total	4.95 (2.95)	8.48 (2.06)

Pivotality judgements

In order to explore whether children were sensitive to pivotality, we performed an analysis including the children's responses only to trials where the wire criticality was equal to 1 but its pivotality differed (1 or ½). Their responses to these trials were submitted to a 2 (value: ½, 1) x 2 (order: pivotality first, pivotality second) x 2 (PS-level: second, third) mixed ANOVA, with the first variable manipulated within participants and the others two between groups. The results are displayed in Table 3. Results showed main effects for value: children held the target wire more responsible when its action was fully pivotal to the battery's malfunctioning than when it was not $F(1, 165) = 49.67$, $p < .001$,

η2= .23, and for order: their ratings were higher when they completed criticality trials before those for pivotality (7.26 vs 6.34). The interaction between value and PS-level was also significant, $F(1, 165) = 7.54$, $p < .01$, η2= .07. Further analysis showed that although children of both second and third PS-levels provided higher ratings when the wire's action was fully pivotal to the outcome than when it was not, the effect was greater at the third PS-level (7.59 vs 6.10); $F(1, 89) = 56.94$, $p < .001$, η2= .39 than at the second (6.94 vs 6.59); $F(1, 78) = 8.03$, $p < .01$, η2= .09. No other significant interactions were found. This shows that older children are more sensitive to pivotality.

Table 3. Mean ratings (SDs) for the two pivotality values per PS-level with the same criticality conditions (left and centre columns) and the same pivotality values with two different criticality conditions (right and centre columns).

	P = ½ [C=1]	**P = 1 [C=1]**	**P = 1 [C = ½]**
Second PS-level	6.88 (2.42)	7.72 (2.46)	5.84 (2.43)
Third PS-level	6.52 (2.19)	8.41 (2.04)	5.58 (2.33)
Total	6.69 (2.30)	8.08 (2.26)	5.70 (2.37)

In order to explore whether there was an effect of criticality on pivotality judgements, we selected responsibility judgments of pivotality in the conditions with full pivotality (P =1) and different values of criticality. We submitted participants' ratings to a 2 (criticality-pivotality: C = ½ - P = 1, C = 1 – P = 1) x 2 (order: criticality first, criticality second) x 2 (PS-level: second, third) mixed ANOVA. Results are displayed in Table 3. The analysis revealed a strong main effect for the criticality-pivotality factor, $F(1, 165) = 130.06$, $p < .001$, η2= .44; differences in order did not reach significance, $F(1, 165) = 3.77$, $p = .054$, η2= .02. When pivotality was held constant, children attributed more responsibility to the target wire when its contribution was fully critical for the battery's functioning than when it was not. In addition, children's ratings were higher when they started the task answering the criticality trials than when they started it completing the pivotality ones, $M = 7.16$, $SD = 1.72$, and $M = 6.65$, $SD = 1.98$, respectively. The interaction between pivotality-criticality and PS-level was also significant, $F(1, 165) = 5.56$, $p < .05$, η2= .03. Further analysis showed that although at both second and third PS-levels there was an effect of criticality on pivotality judgements, the effect was greater at the third PS-level than at the second, $F(1, 167) = 109.31$, $p < .001$, η2= .55, and $F(1,78) = 34.56$, $p < .001$, η2= .30, respectively. No other significant interactions were found.

Counterfactual reasoning questionnaire and pivotality judgements.

Children's responses to the counterfactual reasoning questionnaire were coded with scores of 1 for a correct response and 0 for incorrect ones. We excluded from the sample completing the counterfactual reasoning questionnaire fifty-five participants who answered the control comprehension question incorrectly. Thus, the sample for this analysis consisted of one hundred and six children (51 from the second PS-level and 55 from the third PS-level). From children's answers to the three remaining questions (inferential, epistemic status and semifactual), we computed a global index of counterfactual reasoning (CF index) based on the average of the correct responses to each question, with higher scores indicating a better counterfactual reasoning performance. An independent sample T-test was conducted using the CF index and grouped by the children's PS-level. Results showed that older children (third PS-level; M = 0.74, SD = 0.31) gave more correct responses than younger ones (second PS-level; M = 0.56, SD = 0.33) $T(159) = 2.59$, p =.005, d=0.56.

In order to test whether there was a relation between children's counterfactual reasoning abilities and their pivotality judgements, we computed an index of sensitivity to pivotality. This index reflects the difference between children's ratings in fully pivotal trials (P = 1) and partially pivotal trials (P = ½). We analysed whether there was a relation between the sensitivity to pivotality index, the three counterfactual reasoning questions (inferential, epistemic status and semifactual), and the CF index. Results showed a positive correlation between the sensitivity to pivotality index and the semifactual question (r = 0.25; p < .01). No other relations were significant (inferential: r = 0.02, p = .840; epistemic status: r = -0.07, p = .442; CF index: r = 0.10; p = .300). Although the three questions all concern counterfactual thinking, they do so in different ways: the inferential question is an automatic response with the structure of a modus ponens (if A, B; A, therefore B) that is cancelled by the contextual information. The second question evaluates children's ability to detect what is real and what hypothetical in the meaning of a counterfactual. These two measures did not correlate with the sensitivity to pivotality. It is not surprising that the third, the epistemic question, correlates with sensitivity to pivotality. The question requires an understanding that the outcome is fixed even if the antecedent is mentally mutated. This operation of mentally undoing actions while the real outcome is fixed is employed in the computation of pivotality, as proposed by the structural model of outcome responsibility.

General Discussion

In these two experiments, we have examined whether the structural account (Lagnado et al., 2014, 2015) can explain children's responsibility judgements about mechanical devices. The structural account posits that responsibility judgements are affected by both how critical a wire is perceived to be (criticality component) and how close a wire is to being pivotal for a negative outcome (pivotality component). In Experiment 1, using a novel task adapted to test children, we replicated Lagnado and colleagues' (2014) previous findings with adults. In Experiment 2, we applied this task to children aged between 8 and 12 years. We showed that children's responsibility judgements about mechanical devices were affected by both how critical the wire's contribution was for the battery to work and by how close the wire's action was to being pivotal to changing the outcome. In particular, children rated the wire's contribution as more critical in trials that were fully critical than in those that were partially critical. Similarly, in the cases where the wire's contribution was fully critical, participants held the target wire more responsible when its action was fully pivotal to the battery's malfunctioning than when it was not. Furthermore, judgments of responsibility after the outcome information was given were influenced by the criticality values. In particular, in the cases where the wire's action was pivotal to the outcome (P =1), participants rated the wire as more responsible when its contribution was fully critical (C = 1) than when it was not (C = 0.5). Thus, our results show not only that children are sensitive to both criticality and pivotality parameters but also that neither criticality nor pivotality alone is sufficient to explain responsibility attributions.

We also aimed to explore whether, as the structural account posits (Lagnado et al., 2014, 2015), pivotality judgements depend on counterfactual thinking. Counterfactual thinking is adequately developed in adults but not in children. To this end, in Experiment 2, we asked children to complete a counterfactual reasoning questionnaire. We found some interesting developmental changes. First, that children's sensitivity to pivotality increased with school age. That is, older children differentiated better than younger ones the cases where the wire's action was pivotal to the outcome from those where it was not. This did not happen with criticality. Secondly, in accordance with previous studies (Gómez- Sánchez et al., 2020; Moreno-Ríos & García-Madruga, 2002; Rafetseder et al., 2013), we found children's counterfactual thinking abilities increased during mid-childhood. In particular, older children obtained higher ratings than younger ones in the counterfactual reasoning questionnaire. Finally, we found a positive relationship between

children's performance in semifactual questions and their sensitivity to pivotality. The results are consistent with a possible relation between counterfactual thinking and pivotality judgements, as the structural account posits (e.g., Lagnado et al., 2014).

Additionally, there was an effect of the order of presentation of criticality and pivotality trials in judgments of responsibility. The most interesting finding is that this did not influence the magnitude of criticality or pivotality judgments, and therefore, the effect is not inconsistent with the causal structural model. In criticality trials, just one piece of information was given: whether the target wire connection was necessary or sufficient for the battery operation; no information about the outcome was given. In pivotality trials, however, information was given on both. Adults gave higher ratings for criticality when the criticality trials were presented first (Experiment 1), and so did children (Experiment 2). It is possible that participants who received pivotality trials first might have been more cautious later, giving lower responsibility ratings when they had only the information about criticality, perhaps anticipating uncertainty about the outcome. However, unlike adults, children also gave higher ratings in pivotality judgments when they started the task with the criticality trials. There is a consistent plausible explanation: when participants answer about the criticality of an event, they have to analyse the potential role of the different agents before they know the actual results. Therefore, their judgments are not modulated by other factors such as the actual result. On the other hand, when they have to make a judgment of pivotality for a target wire, they not only consider the outcome information, they are also informed about the structure of the problem, and implicitly about the criticality of the wire, before considering the result and how it could be undone. Participants in the criticality first condition completed a set of trials in which they had to answer with explicit judgments about the criticality of the wire (based on the structure of the problem). On the other hand, in the pivotality first condition, they made the same analysis of the structure of the problem (with the criticality of the agent) but did not respond to criticality, and therefore the analysis was made implicitly. Previous studies on children's reasoning about devices have shown that directing children to notice explicitly the reason for their functioning improved their performance (e.g., Legare & Lombrozo, 2014; Pine & Messer, 2000). Unlike in the pivotality first condition, in the pivotality second condition, children had to make explicit responses to the criticality question. This previous experience could teach children to carry out an explicit analysis of the problem, enabling them to make more accurate judgments about pivotality. In the previous experiments, participants either received partial information (no information

about the outcome) and decided about the criticality, or they received complete information about the device and its operation but were only asked about pivotality. Results were consistent with the independence of the two parameters (criticality and pivotality).

Finally, during childhood, individuals increase their abilities to reason about causes. They show some difficulties in thinking about situations with multiple potential causes. In Siegler's (1978) classic balance problems as those described by Inhelder and Piaget (1958), only older children, and in some cases only adolescents, were able to solve some problems with devices. This kind of problem requires thinking about actual outcomes, considering previous evidence and anticipating how the outcome could have been in other cases. These mental operations include counterfactual and semifactual thinking to undo previous results, as well as the attribution of responsibility. As we have seen, the causal model of attribution provides a good framework to analyse the development of children’s thinking about mechanical devices, making specific predictions about the independence between criticality and pivotality and the relation between pivotality and reasoning with counterfactuals (and particularly with semifactuals). The model could also predict a fact that we have confirmed: older children make better judgments than younger children in pivotality but not in criticality.

Capítulo 6

Discusión General

En la presente tesis doctoral hemos investigado los cambios cognitivos en el desarrollo de la atribución y de conocimiento y la atribución de responsabilidad en escolares de entre 8 y 13 años y adultos. En lo que respecta a la atribución de conocimiento, hemos mostrado que el sesgo de transparencia ilusoria en la intención (Keysar, 1994) y el sesgo retrospectivo auditivo (Bernstein y col., 2012) esta presente en la infancia, al menos, desde los 8 años de edad. La magnitud del sesgo de transparencia ilusoria permaneció estable entre los 8 y los 13 años. En cambio, la magnitud del sesgo retrospectivo disminuyó durante este mismo período del ciclo evolutivo. En concreto, a diferencia de investigaciones previas (Bernstein y col., 2011; Pohl y col., 2010), encontramos que, alrededor de los 9 años de edad, el sesgo comienza a disminuir y continúa haciéndolo hasta, al menos, los 12 o 13 años de edad. Algunos autores (p.ej. Epley y col., 2004a; Nickerson, 1999; Royzman y col., 2003) han propuesto la existencia de un mecanismo común al sesgo de transparencia ilusoria y el sesgo retrospectivo en su componente de impresión de previsibilidad (véase, por ejemplo, Blank y col., 2008). Nuestros resultados sugieren que, que, al menos en la infancia, existen diferencias sensibles al desarrollo en los procesos psicológicos implicados en ambos sesgos egocéntricos.

En lo que respecta a la atribución de responsabilidad, hemos mostrado que el Modelo Estructural de Atribución de Responsabilidad (Lagnado y col., 2014, 2015) puede explicar las atribuciones de responsabilidad que realizan escolares y adultos a un componente de un dispositivo electrónico. Sin embargo, el curso evolutivo de los juicios de criticidad y pivotalidad difirió. En concreto, mientras que no hubo diferencias evolutivas en el grado de sensibilidad a la criticidad, la sensibilidad a la pivotalidad incrementó con la edad. El modelo de Lagnado y colaboradores (2014, 2015), defiende que los juicios de pivotalidad dependen exclusivamente del pensamiento contrafáctico. Al igual que investigaciones previas (p. ej., Gómez-Sánchez y col., 2020; Raftseder y col., 2013), encontramos que las habilidades de los escolares para razonar contrafácticamente incrementan con la edad. Además, encontramos una correlación positiva entre la capacidad para razonar semifácticamente y la sensibilidad a la pivotalidad. Estos resultados avalan la idea de que el factor de pivotalidad es puramente contrafáctico.

Por otra parte, resulta interesante señalar las similitudes encontradas a nivel evolutivo entre la atribución de responsabilidad y la atribución de conocimiento. Al igual que los juicios de criticidad, la magnitud del sesgo de transparencia ilusoria permaneció

estable entre los 8 y los 13 años de edad. En este trabajo se ha puesto a prueba en adultos una explicación alternativa al fenómeno de la transparencia ilusoria según la cual en el sesgo estarían operando procesos de razonamiento causal (Estudio 2). El curso evolutivo compartido por ambos podría considerarse una evidencia indirecta a favor de la implicación de procesos de razonamiento causal en el fenómeno de la transparencia ilusoria. Por otra parte, también pueden observarse ciertas similitudes entre el curso evolutivo de los juicios de pivotalidad y el curso del sesgo retrospectivo. Curiosamente, mientras que la sensibilidad a la pivotalidad aumenta con la edad, la magnitud del sesgo retrospectivo disminuye. De hecho, el momento evolutivo de ambos cambios coincide. En ambos casos son los/as niños/as mayores (2º ciclo, 5º y 6º de primaria) quienes muestran una mayor sensibilidad a la pivotalidad (Estudio 4, Experimento 2) y un menor sesgo retrospectivo (Estudio 3, Experimento 1 y 2). Estudios con adultos han demostrado que la magnitud del sesgo retrospectivo disminuye cuando se les pide a los participantes que consideren resultados alternativos (p.ej., Arkes y col., 1988; Nestler y col., 2008b). Esta capacidad para crear alternativas a la realidad depende del pensamiento contrafáctico (Byrne, 2016). Así, podría ocurrir que las mejores habilidades de razonamiento contrafáctico de los/as niños/as mayores les otorgase una mayor capacidad para contemplar alternativas a la realidad (mundos hipotéticos) y que, entre ellas, se encontrase la capacidad para atender a otras perspectivas distintas a la propia, lo que se traduciría en un menor egocentrismo. Además, a nivel conceptual sesgo retrospectivo y pivotalidad guardan un estrecho vínculo: ambos requieren analizar las probabilidades de ocurrencia de un resultado informado.

El sesgo de transparencia ilusoria

Como comentamos anteriormente, en este trabajo hemos demostrado que los escolares de entre 8 y 13 años exhiben el sesgo de transparencia ilusoria. A diferencia del sesgo, la habilidad para detectar la intención sarcástica del emisor del mensaje incrementó con la edad. Estas diferencias en el curso evolutivo de intención e interpretación (sesgo) no hacen sino evidenciar lo ya constatado por múltiples investigaciones evolutivas: la atribución de intención y la atribución de interpretación (creencias) a otras personas implican procesos diferentes (para más detalles, véase Saxe y col., 2004; Poulin-Dubois y Yott, 2018). Un claro ejemplo de esta disociación lo podemos encontrar en el momento evolutivo en el cada tipo de atribución aparece. Se ha visto que la habilidad para identificar la acción de un agente como intencional aparece entre los 9 y los 12 meses (Akhtar y Martinez-Sussmann, 2007; Baird y Astington, 2005; Brink y col., 2015; Csibra

y col., 1999; Gergely y col., 1995; Olineck y Poulin-Dubois, 2009; Saxe y col., 2004), mientras que la habilidad para identificar correctamente las creencias falsas de otras personas aparece en torno a los 4 años (Gopnik y Wellman, 1992; Perner, 1991; Wiesmann y col., 2020).

En concreto, los/as niños/as mayores (11-13 años) atribuyeron más sarcasmo a la intención del emisor cuando su experiencia era negativa que los pequeños. Este resultado es consistente con las investigaciones que muestran como, aunque la habilidad para comprender el sarcasmo aparece en torno a los 5 o 6 años edad, esta continúa desarrollándose a lo largo de la infancia (p.ej., Burnett, 2015, Glenwright y Pexman, 2010; Nilsen y col., 2011). Asimismo, las diferencias en

A diferencia de en investigaciones con adultos (Moreno-Ríos y col., 2011), el sesgo de transparencia ilusoria no desapareció cuando los escolares habían de predecir cómo un receptor accidental interpretaría el mensaje ambiguo. En adultos, la desaparición del efecto se ha interpretado como una evidencia a favor de la implicación de principios pragmáticos en el sesgo: no hay conocimiento compartido entre emisor y receptor y, por lo tanto, no hay razón para mantener los principios conversacionales (Moreno-Ríos y col., 2011). Como comentábamos en la introducción, tanto la habilidad para comprender el sarcasmo, como el conocimiento pragmático en general de los niños/as continúa desarrollándose a lo largo de la infancia media y tardía (p.ej., Burnett, 2015, Glenwright y Pexman, 2010; Nilsen y col., 2011). Así, podría ocurrir que la no desaparición del efecto en escolares se debiese a que éstos aun no han interiorizado los principios conversacionales contemplados en la Teoría Estándar del Uso del Lenguaje (p.ej. Clark y Marshall, 1981; Greene y col., 1981), y más específicamente, el principio de cooperación en la comunicación (Grice, 1975). Esta idea es consistente con los resultados que encontramos cuando le pedimos a los escolares que juzgasen si el receptor del mensaje podría conocer o no la experiencia del emisor. En este caso, los/as niños/as atribuyeron un mayor conocimiento de la experiencia al receptor "real" del mensaje que al accidental.

La pregunta sobre el conocimiento de conocimiento de la experiencia nos permitió poner a prueba dos de las explicaciones ofrecidas al sesgo, la explicación de la *construal* (Keysar, 1994, 2000) y la explicación pragmática (Gerrig y col., 2000). Ambas explicaciones coinciden en señalar que el sesgo es producto de una atribución errónea, sin embargo, difieren en qué es lo atribuido: una información adicional no expresada en el relato (explicación pragmática) o el mensaje reconstruido (*construal*). Por tanto, sólo

desde la *construal* se esperaría que se atribuyese conocimiento de la experiencia al un lector accidental, puesto que el mensaje ya informa de la experiencia. Pese a que los escolares atribuyeron una mayor capacidad para conocer la experiencia a un receptor real que a uno accidental, éstos también atribuyeron al receptor accidental capacidad para conocer la experiencia. Además, nótese que en la medida indirecta de atribución de conocimiento (atribución de sarcasmo al receptor, el sesgo) los escolares también atribuyeron al receptor accidental capacidad para conocer la experiencia (el sesgo no desapareció en esta condición. Estos resultados sugieren que, al menos en la infancia, tanto principios como operaciones de *construal* estarían actuando en sesgo de transparencia ilusoria. Por otra parte, las diferencias encontradas entre ambas medidas (directa e indirecta) podrían estar relacionadas con los dos sistemas de procesamiento de la información (p.ej., Kahneman, 2011; Evans. y Stanovich, 2013). Así, la medida directa podría requerir un procesamiento más profundo y exhaustivo de la información (Sistema II o Procesamiento Tipo II), mientras que la medida indirecta podría implicar un procesamiento más automático, menos profundo de la información (Sistema I o Procesamiento Tipo I). Evans y Stanovich (2013) definen como una de las falacias clave de las Teorías de Procesamiento Dual de la Información afirmar que el Procesamiento Tipo 1 es el responsable de los sesgos cognitivos y el Procesamiento Tipo II de las respuestas correctas. Sin embargo, en casos como el de transparencia ilusoria donde el sesgo (error en la toma de perspectiva) es evaluado a partir de otros fenómenos como puede ser la comprensión del sarcasmo, el Procesamiento Tipo I sí podría estar involucrado en las respuestas incorrectas en esta tarea. Con frecuencia en nuestro día a día hemos de interpretar mensajes sarcásticos. Habitualmente, la interpretación que realizamos de estos mensajes es automática (no nos detenemos a pensar si hay una incongruencia entre "lo que se dice" y "lo que realmente se quiere decir") y correcta. Así, aunque en la mayoría de las ocasiones el Procesamiento Tipo I arroja respuestas eficientes y adecuadas, en otras ocasiones nos conduce a error.

Otra de las explicaciones ofrecidas al sesgo, la explicación de la "transparencia de la intención" defiende que el sesgo se produce porque los lectores asumen que la intención del emisor es "transparente" para el receptor -el receptor puede "ver" la intención sarcástica del emisor- (Keysar, 1994). De producirse el sesgo por la "transparencia de la intención", se habría esperado una relación directa entre el sarcasmo atribuido a la intención y el atribuido a la interpretación; sin embargo, ni en niños (Estudio 1) ni en adultos (Estudio 2, Experimento 2 y 3) se observó esta correspondencia. De manera

sistemática, tanto en escolares (Estudio 1) como en adultos (Estudio 2, Experimento 2 y 3), el sarcasmo atribuido a la intención fue independiente del atribuido a la interpretación. En el estudio con niños/as, pese a que los mayores se mostraron más capaces de detectar la intención sarcástica del emisor, estos no la trasladaron con mayor frecuencia la intención del emisor a la interpretación del receptor, tal y como defiende esta explicación. Esta ausencia de correspondencia entre el sarcasmo atribuido a intención e interpretación también pudo observase en adultos. En todos los casos, el sarcasmo atribuido a la intención del emisor fue independiente de la magnitud del sesgo de transparencia ilusoria (Estudio 2, Experimento 2 y 3). Estos resultados pese a que no nos permiten descartar por completo la explicación de "transparencia de la intención" sí muestran que ésta resulta insuficiente para explicar el fenómeno de la transparencia ilusoria en totalidad.

En este trabajo también exploramos una explicación alternativa según la cual en el sesgo de transparencia ilusoria estarían actuando procesos de razonamiento causal. Para evaluar nuestra propuesta nos basamos en resultados de investigaciones sobre *creeping determinism*, uno de los componentes o manifestaciones del sesgo retrospectivo que depende de procesos de razonamiento causal (p. ej., Blank y Nestler, 2007; Nario y Brascombe, 1995; Nester y col., 2008a; Nestler y col., 2010; Roese y Olson, 1996; Oeberst y col., 2014). Nuestros resultados sugieren que, al igual que el *creeping determinism*, en el sesgo de transparencia ilusoria estarían implicados procesos de razonamiento causal. En primer lugar, el sesgo de transparencia ilusoria disminuyó cuando la información disponible hacía difícil a los/as participantes crear un modelo causal coherente sobre cómo ocurrieron las cosas (Estudio 2, Experimento 2) o cuando se debilita el nexo causal entre antecedente y consecuente (Estudio 2, Experimento 3, grupo semifáctico). En segundo lugar, , y de nuevo al igual que el *creeping determinism* (p. ej., Roese y Olson, 1996), el sesgo incrementó cuando el nexo causal entre antecedente y consecuente era fortalecido. Por último, el *creeping determinism* (el efecto del conocimiento del resultado) apareció cuando los/as participantes fueron informados de cómo el receptor del mensaje había interpretado éste (Estudio 2, Experimento 1). Estudios previos han informado de que el *creeping determinism* sólo aparece cuando los participantes detectan información de naturaleza causal en la descripción del evento (Yopnick y Kim, 2012). Estos resultados sugieren, tal y como defiende la hipótesis de la ilusión causal, que en el sesgo de transparencia ilusoria estarían actuando procesos de razonamiento causal. Así, podría ocurrir que, tal y como Turner y Schley (2016) sugieren en su propuesta de considerar los efectos de anclaje y ajuste como un antecedente seguido

de un consecuente, en algunas ocasiones como por ejemplo el sesgo de transparencia ilusoria, los denominados efectos de anclaje y ajuste fuesen producto de un razonamiento causal. De hecho, la explicación ofrecida por algunos autores para los efectos de anclaje y ajuste (véase por ejemplo, Mussweiler y Strack, 1999; Strack y Mussweiler, 1997), y la teoría explicativa de los efectos de *creeping determinism* (CMT, Blank y Nestler, 2007) guardan grandes similitudes.

El sesgo retrospectivo

Pese a que el sesgo retrospectivo ha sido ampliamente estudiado en adultos, son pocas las investigaciones que han explorado el sesgo en niños/as y los resultados que de ellas se desprenden contradictorios (Bernstein y col., 2011; Pohl y col., 2010; Pohl y col., 2018). En este trabajo hemos demostrado que el sesgo retrospectivo auditivo esta presente en escolares de entre 8 y 13 años de edad (Estudio 3) tanto en diseños hipotéticos (Experimento 1 y 2) como en diseños de memoria (Experimento 2). A diferencia de investigaciones previas (Bernstein y col., 2011), encontramos que el sesgo continúa disminuyendo después de los 6 años en los diseños hipotéticos. De hecho, el curso evolutivo encontrado en ambos diseños experimentales fue muy similar. En ambos casos fueron los/as niños/as más pequeños/as quienes mostraron un mayor sesgo. Además, también en ambos diseños, el sesgo comenzó a descender alrededor de los 9 años. Esta edad (9 años) parece ser un período especialmente sensible en el desarrollo, señalado por muchas investigaciones como un "punto de inflexión" en las respuestas egocéntricas de los menores (p. ej., Lagattuta y col., 2010; Mitchel y col., 1996; Pohl y col., 2018). A nivel cognitivo, durante la infancia media y tardía los menores van adquiriendo y/o perfeccionando habilidades relacionadas con la metacognición (véase, por ejemplo, Flavell y col., 1995; Flavel y col., 1997; Siegler, 1996) así como mejorando sus habilidades memoria (p.ej., Cowan, 1997; Goswami, 2020; Keil, 1989; Schneider y Pressley, 1997). Además, otras habilidades como es el caso del razonamiento contrafáctico, también experimentan una mejora sustancial durante este período del ciclo vital (Gómez- Sánchez y col., 2020; Markovits, 2014; Moreno-Ríos y García-Madruga, 2002; Rafetseder y col., 2013). Todos estas "ganancias" que a nivel cognitivo ocurren durante este período podrían estar contribuyendo a la reducción del egocentrismo en este momento del ciclo vital (9 años en adelante) encontrado en ésta y otras investigaciones.

En adultos, el sesgo retrospectivo en general (Harley y col., 2004; Pohl y Hell, 1996), y el sesgo retrospectivo auditivo en particular, se ha mostrado un efecto difícil de eliminar o, incluso, reducir. Sin embargo, en un estudio reciente, Bernstein y

colaboradores (2018) mostraron que el sesgo auditivo desaparecía cuando se les pedía a los participantes que tratasen de identificar el estímulo auditivo distorsionado antes de recibir información sobre la identidad de éste y realizar sus estimaciones sobre el número de personas que serían capaces de identificarlo. Siguiendo a Bernstein y colaboradores (2018) incluimos una condición (Experimento 1) en la que pedíamos a los escolares que realizarán dos estimaciones sobre una misma canción: la primera sin información sobre el contenido del mensaje "oculto" y la segunda después de haber sido informados de su contenido. El objetivo fue comprobar si, al igual que ocurre en adultos, el hecho de hacer conscientes a los menores de las dificultades que entraña identificar el mensaje "oculto" sin recibir información sobre su contenido, eliminaría (o al menos reduciría el sesgo). Sin embargo, los/as niños/as continuaron mostrando el sesgo en esta condición. De hecho, la magnitud del sesgo en esta condición fue muy similar al encontrado en las condiciones tradicionales. No obstante, existe una diferencia clave entre el estudio de Bernstein y colaboradores (2018) y el nuestro que podría explicar porqué el sesgo retrospectivo no desapareció. En nuestro estudio, a diferencia del citado, los participantes no tenían que identificar el mensaje "oculto". Este hecho pudo favorecer el sesgo: hace más difícil notar la dificultad de adivinar el mensaje "oculto" sin conocer su contenido. De nuevo aquí, tanto la persistencia del sesgo en escolares como la desaparición del efecto en adultos podría estar relacionado con los dos sistemas de procesamiento de la información (p.ej., Evans y Stanovich, 2013; Kahneman, 2011).

En lo que respecta a la naturaleza del sesgo retrospectivo auditivo, en este trabajo hemos mostrado que los procesos implicados en la emergencia del sesgo en la infancia difieren en función del tipo de diseño experimental empleado para su evaluación. En adultos, investigaciones previas habían vinculado no sólo al sesgo retrospectivo auditivo (Higham y col., 2017; Bernstein y col., 2018) sino también al sesgo retrospectivo visual (Harley y col., 2004; Bernstein y Harley, 2007) con procesos de fluencia en los diseños hipotéticos. Así, el sesgo en estos diseños sería producto de una atribución incorrecta de la fluencia (facilidad de procesamiento) generada por el conocimiento del resultado a las propias características del estímulo a juicio (*fluency missattributions*). Nuestros resultados sugieren que esta atribución incorrecta de la fluencia es el mecanismo responsable de la emergencia del sesgo también en la infancia. En cambio, en los diseños de memoria, al igual que ocurre en adultos (Higham y col., 2017), el error aquí sería producto de cómo la información es recuperada y reconstruida de la memoria. Higham y colaboradores (2017) han relacionado el sesgo retrospectivo auditivo en los diseños de

memoria con *misinformation effect* (Loftus et al., 1978; Pohl & Gawlik, 1995). En concreto, estos autores argumentan que en estos diseños el conocimiento de un resultado altera la representación mental (esquema mental) del juicio previo, lo que conduciría a una reconstrucción sesgada del juicio previo y, por tanto, al sesgo.

Por otra parte, hasta donde nosotros conocemos, no existen estudios que hayan explorado si el sesgo retrospectivo difiere en función de los estímulos empleados para su evaluación. En la introducción del presente trabajo vimos que entre los 9 y 12 años de edad los menores desarrollan una preferencia por atender a estímulos visuales frente a auditivos (Nava y Pavani, 2013). Esto nos llevó a pensar que quizás pudiese haber diferencias relativas a la magnitud y curso evolutivo del sesgo en función del tipo de estímulos empleados para su evaluación (auditivo o visuales). Sin embargo, en el Estudio 3 (experimento 1) encontramos que tanto la magnitud del sesgo como el curso evolutivo del mismo es independiente de la modalidad perceptiva. Tanto en la tarea auditiva como en la visual, los/as más pequeños/as fueron quienes mostraron un mayor sesgo. En ambos casos, el sesgo comenzó a descender alrededor de los 9 años de edad (4º de primaria). No obstante, resulta importante señalar que nuestra tarea se diferencia de otras tareas visuales empleadas para evaluar el sesgo (p. ej., Bernstein y col., 2011; Harley y col., 2004), el resultado, la información de la identidad del estímulo visual, es proporcionado por el experimentador. En las previas, el resultado es "autogenerado" por los/as participantes a partir de su propia percepción del estímulo visual que progresivamente se va clarificando. Así, el sesgo capturado en la tarea visual podría considerarse un sesgo "híbrido" procedente tanto de estímulos visuales (la imagen del personaje famoso) como de estímulos visuales (la información proporcionada por el experimentador sobre la identidad del personaje).

Atribución de responsabilidad

Otro de los objetivos de la presente tesis doctoral ha sido investigar si el modelo de atribución de responsabilidad desarrollado por Lagnado y colaboradores (2015) puede explicar los juicios de responsabilidad que realizan escolares de entre 8 y 13 años y adultos a componentes de un dispositivo electrónico. Para poner a prueba el modelo, creamos una tarea en la que los participantes habían de juzgar la criticidad de un cable para el funcionamiento de una batería, así como juzgar la pivotalidad de éste en el incorrecto funcionamiento de la batería. Nuestros resultados mostraron que el Modelo Estructural de Atribución de Responsabilidad desarrollado por Lagnado y colaboradores (2014, 2015) puede explicar las atribuciones de responsabilidad que escolares y adultos

realizan a un componente de un dispositivo electrónico. Tanto niños/as (Experimento 2) como adultos (Experimento 1) se mostraron sensibles al valor de criticidad. Ambos juzgaron como más importante la contribución del cable cuando el funcionamiento de este era completamente crítico para el funcionamiento de la batería (C = 1) que cuando no lo era. Este resultado es especialmente relevante en el caso de los adultos dado que en el capítulo donde Lagnado y colaboradores (2015) aluden a dicha investigación, únicamente informan de haber manipulado el factor de pivotalidad.

Al igual que en criticidad, tanto escolares como adultos se mostraron sensibles a la pivotalidad. En concreto, en los casos en los que el funcionamiento del cable era completamente crítico para el funcionamiento de la batería (C = 1), éstos atribuyeron una mayor responsabilidad al cable en la condición donde su mal funcionamiento había sido máximamente pivotal al incorrecto funcionamiento de la batería (P = 1) que en la condición donde su pivotalidad era menor (P = ½). Además, las atribuciones de responsabilidad (juicios de pivotalidad) tanto de niños/as como de adultos estuvieron determinados por ambos factores: criticidad y pivotalidad. Así, en las condiciones donde el funcionamiento del cable había sido plenamente pivotal al resultado (P = 1), los participantes atribuyeron una mayor responsabilidad al cable cuando el funcionamiento de éste además era completamente crítico (C = 1) que cuando no lo era (C = ½). Estos resultados demuestran no sólo que niños y adultos son sensibles al valor de criticidad y pivotalidad del componente de un dispositivo electrónico sino también que ambos factores son necesarios para explicar sus juicios de responsabilidad en el contexto de un dispositivo electrónico.

Por otra parte, Lagnado y colaboradores (2014, 2015) defienden que los juicios de pivotalidad dependen exclusivamente del pensamiento contrafáctico. Dos de nuestros resultados avalan esta hipótesis. En primer lugar, tanto juicios de pivotalidad como habilidades de razonamiento contrafáctico (respuestas correctas proporcionadas al cuestionario) mostraron el mismo curso evolutivo: ambos incrementaron con la edad. En segundo lugar, la sensiblidad a la pivotalidad correlacionó positivamente la precisión de los escolares en las inferencias semifácticas. La pregunta semifáctica es especialmente relevante en este contexto porque evalúa la capacidad de los/as niños/as para comprender que un mismo resultado (p. ej., el incorrecto funcionamiento de la batería) puede permanecer inmutable incluso si el antecedente es eliminado, es decir, su capacidad para comprender, por ejemplo, que “aunque el cable marrón hubiese funcionado, la batería no habría funcionado”. La dificultad para comprender este tipo de afirmaciones haría más

difícil para los menores distinguir la situación donde, por ejemplo, únicamente uno de los cables funciona incorrectamente de aquella donde dos (o tres) cables funcionan de modo incorrecto y, por ende, menos sensibles a la pivotalidad.

A diferencia de en los juicios de pivotalidad, en los juicios de criticidad no hubo diferencias evolutivas. Tanto niños/as mayores como pequeños/as se mostraron igualmente sensibles al valor de criticidad. Esta ausencia de diferencias es consistente con la idea de que el factor de criticidad depende del razonamiento causal y, consecuentemente, con los resultados de estudios evolutivos que muestran como en torno a los 4 años de edad los menores ya son capaces de operar con los principios básicos de razonamiento (Bullock y German, 1979; Shultz y Mendelson, 1975). Por el contrario, si hubo diferencias evolutivas en el efecto que el valor de criticidad tuvo sobre los juicios de pivotalidad. En concreto, en los casos en los que se mantenía la pivotalidad constante (P = 1), los juicios de los/as niños/as mayores se mostraron más influenciados por el valor de criticidad (1 o ½) que los juicios de los/as pequeños/as, aunque en ambos grupos de edad el valor de criticidad influenció los juicios de pivotalidad; o dicho de otro modo, en ambos casos las atribuciones de responsabilidad estuvieron determinadas por el valor de criticidad y el grado de pivotalidad del cable. En este sentido, Siegler (1976, 1978), haciendo uso de la tarea piagetiana de la balanza, demostró que las habilidades de los/as niños/as para razonar sobre múltiples causas y, más específicamente, para razonar sobre cómo múltiples causas (peso y distancia) pueden combinarse para alcanzar un resultado (equilibrio de una balanza) se desarrolla de manera paulatina entre los 7 y los 12 o 13 años. Así, el mayor efecto del nivel de criticidad en los juicios de pivotalidad mostrado por los/as niños/as mayores podría estar simplemente reflejando su mayor capacidad para "integrar" en sus juicios de responsabilidad ambos factores (criticidad y pivotalidad).

Conclusiones

Se tiende a pensar que el egocentrismo como característica de pensamiento infantil queda atrás una vez que un/a niño/a es capaz de predecir que Sally buscará su canica en la cesta, es decir, una vez que este/a supera la Tarea de la Falsa Creencia (Baron-Cohen y col., 1985). En este trabajo mostramos cómo algunas manifestaciones egocéntricas, como es el caso del sesgo retrospectivo, continúan disminuyendo mucho tiempo después de que los menores superen esta tarea (4 años). Otras, en cambio, parecen haber alcanzado "niveles adultos" un poco antes en el desarrollo. Así, a pesar de que los diferentes sesgos egocéntricos como las tareas de TOM reflejen una misma limitación: la tendencia a contaminar con el propio conocimiento los juicios que se emiten, lo cierto es

que, como hemos visto en este trabajo, cada forma o manifestación egocéntrica tiene sus propias características que hacen a los menores más o menos propensos a sucumbir a este tipo de errores. Por tanto, el egocentrismo no debería de ser tratado como un fenómeno único con manifestaciones y procesos comunes a todas las respuestas egocéntricas proporcionadas por adultos y niños/as, sino como un fenómeno con una base común y con características distintivas de cada forma y manifestación. Una forma a nuestro juicio útil de explorar las diferencias y semejanzas de cada manifestación egocéntrica es emplear tareas que eliminen la necesidad de ignorar el propio conocimiento pero que a nivel conceptual impliquen un mismo juicio como, por ejemplo, ocurre con en la tarea de sesgo retrospectivo y de atribución de responsabilidad aquí empleadas: ambas requieren estimar las probabilidades de ocurrencia de un resultado; sin embargo, tan sólo en la primera de ellas hay un conocimiento que ha de ser ignorado a fin de emitir un juicio preciso. Con este trabajo hemos pretendido analizar procesos egocéntricos tratando de trasladar los mecanismos explicativos utilizados en unos para comprobar si se pueden trasladar para explicar los otros. En particular, y desde diferentes niveles de análisis, lo hemos hecho con el Heurístico de Anclaje y Ajuste y la atribución causal. Hemos observado que existen diferencias notables, pero también semejanzas. A nuestro juicio esta vía nos aporta un modo de profundizar en el conocimiento del desarrollo de la cognición social.

REFERENCIAS

Akhtar, N., & Martínez-Sussmann, C. (2007). *Intentional communication.* In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (p. 201–220). The Guilford Press.

Amsel, E., Goodman, G., Savoie, D., & Clark, M. (1996). The development of reasoning about causal and noncausal influences on levers. *Child Development, 67*, 1624–1646. https://doi.org/10.1111/J.1467-8624.1996.TB01818.X

Arkes, H. R. (2013). The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science, 22*(5), 356-360. https://doi.org/10.1177/0963721413489988

Arkes, H.R., Faust, D., Guilmette, T.J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*, 305-307. http://dx.doi.org/10.1037/0021-9010.66.2.252

Arkes, H.R., Wortman, R.L., Saville, P.D., & Harkness, A.R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66*(2), 252-254. https://doi.org/10.1037/0021-9010.66.2.252

Ash, I.K. (2009). Surprise, memory, and retrospective judgment making: Testing cognitive reconstruction theories of hindsight bias effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 916-933. http://dx.doi.org/10.1037/a0015504

Astington, J. W., & Baird, J. A. (2005). *Introduction: Why Language Matters.* In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (p. 3–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0001

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a

"theory of mind"? Cognition, 21, 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Bayen, U. J., Pohl, R., & Erdfelder, E. (2007). Hindsight bias across the life span. *Social Cognition, 25*(1), 83-97. https://doi.org/10.1521/soco.2007.25.1.83

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413–426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Bernstein, D.M., Atance, C., Loftus, G.R. & Meltzoff, A. (2004). We saw it all along. Visual Hindsight Bias in Children and Adults. *Psychological Science*, *15*(4), 264-267. https://doi.org/10.1111/j.0963-7214.2004.00663.x

Bernstein, D.M., Atance, C., Meltzoff, A.N., & Loftus, G.R. (2007) Hindsight bias and developing theories of mind. *Child Development, 78*(4), 1374-1394. https://doi.org/10.1111/j.1467-8624.2007.01071.x

Bernstein, D.M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of Age. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*(2), 378-391. https://doi.org/10.1037/a0021971

Bernstein, D. M., & Harley, E. M. (2007). Fluency misattribution and visual hindsight bias. *Memory, 15*(5), 548–560. https://doi.org/10.1080/09658210701390701

Bernstein, D.M., Kumar, R., Masson, M. E., & Levitin, D. J. (2018). Fluency misattribution and auditory hindsight bias. *Memory & Cognition*, *46*(8), 1331-1343. https://doi.org/10.3758/s13421-018-0840-6

Bernstein, D.M., Wilson, A.M., Pernat, N.L.M., & Meilleur, L. R. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19*(4), 588-593. https://doi.org/10.3758/s13423-012-0268-0

Birch, S.A. J., & Bernstein, D. M. (2007). What can children tell us about hindsight bias: A fundamental constraint on perspective-taking. *Social Cognition, 25*(1), 98-113. https://doi.org/10.1521/soco.2007.25.1.98

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*(3), 283-286. http://dx.doi.org/10.1111/1467-9280.03436

Birch, S. A. J., Brosseau-Liard, P.E., Haddock, T. & Ghrear, S.E. (2017). A 'curse of knowledge' in the absence of knowledge? People misattribute fluency when judging how common knowledge is among their peers. *Cognition, 166*, 447-458. http://dx.doi.org/10.1016/j.cognition.2017.04.015

Blank, H., Diedenhofen, B. & Musch, J. (2015). Looking back on the London Olympics: Independent outcome and hindsight effects in decision evaluation. *British Journal of Social Psychology, 54*(4), 798-807. https://doi.org/10.1111/bjso.12116

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, *11*(4-5), 491–504. https://doi.org/10.1080/09658210244000513

Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition, 25*(1), 132-146. https://doi.org/10.1521/soco.2007.25.1.132

Blank, H., & Nestler, S. (2006). Perceiving events as both inevitable and unforeseeable in hindsight: The Leipzig candidacy for the Olympics. *British Journal of Social Psychology, 45*(1), 149-160. https://doi.org/10.1348/014466605X52326

Blank, H., Nestler, S., Von Collani, G., & Fischer, V. (2008). How many hindsight bias are there? *Cognition, 106*(3), 1408–1440. http://dx.doi.org/10.1016/j.cognition.2007.07.007

Blank, H., Musch, J., & Pohl, R. (2007). Hindsight bias: On being wise after the event. *Social cognition, 25*(1), 1-9. https://doi.org/10.1521/soco.2007.25.1.1

Brannon, S. M., & Gawronski, B. (in press). Cognitive consistency in social cognition. In M. Hogg (Ed.), *The Oxford encyclopedia of social psychology*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acrefore/9780190236557.013.314

Brink, K. A., Lane, J. D., & Wellman, H. M. (2015). Developmental pathways for social understanding: Linking social cognition to social contexts. *Frontiers in Psychology, 6,* Article 719. https://doi.org/10.3389/fpsyg.2015.00719

Bullock, M., & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development, 50*, 89-96. https://doi.org/10.2307/1129045

Burnett, D. L. (2015). Exploring the role of conventionality in children's interpretation of ironic remarks. *Journal of Child Language, 42*, 1267–1288. http://dx.doi.org/10.1017/S0305000914000798

Burns, P., Riggs, K. J., & Beck, S. R. (2012). Executive control and the experience of regret. *Journal of Experimental Child Psychology, 111*, 501-515. https://doi.org/10.1016/j.jecp.2011.10.003

Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology, 67,* 135-157. http://dx.doi.org/10.1146/annurev-psych-122414-033249

Calvillo, D.P. & Gomes, D.M. (2011). Surprise influences hindsight-foresight differences in temporal judgments of animated automobile accidents. *Psychon Bull Rev, 18*, 385-391. http://dx.doi.org/10.3758/s13423-011-0062-4

Camerer, C., Lowenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*, 1232–1254. http://dx.doi.org/10.1086/261651

Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality, 51*(4), 605–620. https://doi.org/10.1111/j.1467-6494.1983.tb00868.x

Cane, P. (2002). *Responsibility in law and morality*. Oxford, UK: Hart Publishing.

Capelli, C. A., Nakagawa, N. & Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development, 61*(6), 1824-1841. https://doi.org/10.2307/1130840

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. H. Joshe, B. Webber, & I. A. Sag (Eds), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research, 22*(1), 93–115. https://doi.org/10.1613/jair.1391

Collins, J. D., Hall, E. J., & Paul, L. A. (2004). *Causation and counterfactuals*. Cambridge, MA: MIT Press Cambridge.

Cowan, N. (1997). *The development of working memory.* In N. Cowan (Ed.), *Studies in developmental psychology. The development of memory in childhood* (p. 163–199). Psychology Press/Erlbaum (UK) Taylor & Francis.

Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: Children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbol, 15,* 29–45. http://dx.doi.org/10.1080/10926488.2000.9678863

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition, 72*(3), 237–267. https://doi.org/10.1016/S0010-0277(99)00039-6

Dawson, N. V., Arkes, H. R., Siciliano, C., Blinkhorn, R., Lakshmanan, M., & Petrelli, M. (1988). Hindsight bias: An impediment to accurate probability estimation in clinicopathologic conferences. *Medical Decision Making*, *8*(4), 259–264. https://doi.org/10.1177/0272989X8800800406

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development, 67*, 3071–3085. https://doi.org/10.1111/j.1467-8624.1996.tb01903.x

Drummey, A.B., & Newcombe, N. (2002). Developmental changes in source memory. *Developmental Science, 5*(4), 502–513. https://doi.org/10.1111/1467-7687.00243

Epley, N. (2014). *Mindwise: How we understand what others think, believe, feel, and want*. New York: Alfred A. Knopf.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustements are insufficient. *Psychological Science, 17*(4), 311-318. https://doi.org/10.1111/j.1467-9280.2006.01704.x

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004a). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327-329. http://dx.doi.org/10.1037/0022-3514.87.3.327

Epley, N. Morewedge, C. K., & Keysar, B. (2004b). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*(6), 760-768. https://doi.org/10.1016/j.jesp.2004.02.002

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 387-414. https://doi.org/10.1037/0278-7393.24.2.387

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. http://dx.doi.org/10.1177/1745691612460685

Filippova, E. & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*, 126-138. http://dx.doi.org/10.1111/j.1467-8624.2007.01115.x

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowlegde on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288-299. http://dx.doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*(2), 349–358. https://doi.org/10.1037/0096-1523.3.2.349

Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). The development of children's knowledge about attentional focus. *Developmental Psychology, 31*(4), 706–712. https://doi.org/10.1037/0012-1649.31.4.706

Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Development, 68* (1), 39–47. https://doi.org/10.2307/1131923

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Gerrig, R.J., Ohaeri, J.O., & Brennan, S. E. (2000). Illusory transparency revisted. *Discourse Processes, 29*(2), 137-159. http://dx.doi.org/10.1207/S15326950dp2902_3

Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition, 115*(1), 166–171. https://doi.org/10.1016/j.cognition.2009.12.011

Ghrear, S. E., Birch, S. A. J., & Bernstein, D. M. (2016). Outcome knowledge and false belief. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00118

Glenwright, M., & Pexman, P. M. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language, 37*, 429–451. https://doi.org/10.1017/S0305000909009520

Gómez-Sánchez, J., Ruiz-Ballesteros, J. A., & Moreno-Ríos, S. (2020). How children and adults keep track of real information when thinking counterfactually. *PLOS ONE*, *15*(12), e0242967. https://doi.org/10.1371/journal.pone.0242967

Gopnik, A., & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development, 62*(1), 98–110. https://doi.org/10.2307/1130707

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1-2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Gordo, C., & Moreno-Ríos, S. (2019). Children's Illusory Transparency of Intention: Construal versus Pragmatic View. *Discourse Processes, 56*(4), 331-344. https://doi.org/10.1080/0163853X.2018.1508264

Goswami, U.C. (2020). *Cognitive development and cognitive neuroscience: The learning brain*. London; New York: Routledge, Taylor & Francis Group

Gray, R., Beilock, S. L. & Carr, T. H. (2007). "As soon as the bat met the ball, I knew it was gone": Outcome prediction, hindsight bias, and the representation and control of action in expert and novice baseball players. *Psychonomic Bulletin & Review*, *14*(4), 669-675. https://doi.org/10.3758/BF03196819

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language, 33*, 511-526. https://doi.org/10.1006/jmla.1994.1024

Grice, P.H. (1975). Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), *Syntax and Semantics* (pp. 41-58). Academic Press, New York.

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfidence predictions about the self and others. *Journal of Personality and Social Psychology, 59*(6), 1128-1139. https://doi.org/10.1037/0022-3514.59.6.1128

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology*, *27*(3), 681–702. https://doi.org/10.1348/026151008X357886

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26*(2-3), 103–117. https://doi.org/10.1207/s15324834basp2602&3_1

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science, 56*(4), 843–887. https://doi.org/10.1093/bjps/axi147

Harley, E.M. (2007). Hindsight bias in legal decision making. *Social Cognition*, *25*(1), 48-63. https://doi.org/10.1521/soco.2007.25.1.48

Harley, E. M., Carlsen, K. A., & Loftus, G.R. (2004). The “saw-it-all- along” effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(5), 960-968. https://doi.org/10.1037/0278-7393.30.5.960

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes, 36*, 147–165. https://doi.org/10.1207/S15326950DP3603_1

Harris, P. L., Johnson, C. N., Hutton, D., & Andrews, G. (1989). Young children's understanding of mind and emotion. *Cognition and Emotion, 3*, 379-400. https://doi.org/10.1080/02699938908412713

Hart, H. L. A. (2008). *Punishment and responsibility*. Oxford, UK: Oxford University Press.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311–327. https://doi.org/10.1037/0033-2909.107.3.311

Higham, P. A., Neil, G. J., & Bernstein, D. M. (2017). Auditory hindsight bias: Fluency misattribution versus memory reconstruction. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1143–1159. https://doi.org/10.1037/xhp0000405

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied, 5*(2), 205–221. https://doi.org/10.1037/1076-898X.5.2.205

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 566–581.https://doi.org/10.1037/0278-7393.26.3.566

Inhelder, B., & Piaget, J. (1958). The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books.

Ivanko, S. L., & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Process, 35*(3), 241-279. https://doi.org/10.1207/S15326950DP3503_2

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114(1),* 3–28. https://doi.org/10.1037/0033-2909.114.1.3 PMID: 8346328

Kahneman, D. (2011), *Thinking, Fast and Slow*, NewYork, NY: Farrar, Strauss, Giroux.

Kamin, K.A., & Rachlinski, J.J. (1995). Ex post ≠ ex ante: Determining liability in hindsight. *Law and Human Behavior*, *19*(1), 89-104. https://doi.org/10.1007/BF01499075

Keil, F. C. (1989). *Concepts, Kinds and Cognitive Development*. Cambridge, MA: MIT Press.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35*(2), 157–175. https://doi.org/10.1006/jmla.1996.0009

Keysar, B. (1994). The illusory transparency of intention: linguistic perspective taking in text. *Cognitive Psychology, 26*, 165-208. https://doi.org/10.1006/cogp.1994.1006

Keysar, B. (2000). The illusory transparency of intention: Does Julia understand what Mark means because he means it? *Discourse Processes*, *29*, 161-172. https://doi.org/10.1207/S15326950dp2902_4

Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language, 34*, 89-109. https://doi.org/10.1006/jmla.1995.1005

Khemlani, S. & Johnson-Laird, P.N. (2012). The processes of inference. *Argument & Computation*, *4*(1), 1-17. https://doi.org/10.1080/19462166.2012.674060

Klahr, D. (2000). Exploring science: The cognition and development of discovery process es. Cambridge, MA: MIT Press.

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, *46*(6), 1417-1432. http://dx.doi.org/10.1037/a0021062.

Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*(2), 659–674. https://doi.org/10.1111/cdev.12154

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2014). Causal responsability and counterfactuals. *Cognitive Science, 37*(6), 1036-1073. https://doi.org/10.1111/cogs.12054

Lagnado, D. A. & Gerstenberg, T. (2015). A difference-making framework for intuitive judgments of responsibility. In *Oxford Studies in Agency and Responsibility* (pp. 213-241). Oxford University Press.

Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology, 20,* 163–198. https://doi.org/10.1016/S0732-118X(02)00007-7

Leary, M. R. (1981). The distorted nature of hindsight. *Journal of Social Psychology, 115*(1), 25–29. https://doi.org/10.1080/00224545.1981.9711984

Leary, M. R. (1982). Hindsight distortion and the 1980 Presidential election. *Personality and Social Psychology Bulletin, 8*(2), 257–263. https://doi.org/10.1177/0146167282082012

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. Journal of Experimental Child Psychology, 126, 198-212. https://doi.org/10.1016/j.jecp.2014.03.001

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556–567. http://dx.doi.org/10.1174/021037002762064055

Lewis, D. (1986). Causal explanation. *Philosophical Papers, 2*, 214–240. http://dx.doi.org/10.1093/0195036468.003.0007

Lindsay, D. S. (2008). Source monitoring. In Byrne, J. (Series Ed.) & Roediger III, H. L. (Vol. Ed.), *Learning and memory: A comprehensive reference*. Vol. 2. Cognitive psychology of memory (pp. 325-348). Oxford, England: Elsevier.

Mackie, J. (1974). *The Cement of the Universe*. Oxford: Oxford University Press.

Mante-Estacio, M.J. & Bernardo, A.B.I. (2014). Illusory transparency in bilinguals: Does language of text affect bilingual readers' perspective taking in reading? *Current Psychology, 34*(4), 744-752. https://doi.org/10.1007/s12144-014-9286-6

Markovits, H. (2014). On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. *Journal of Experimental Child Psychology*, *128*, 37–51. https://doi.org/10.1016/j.jecp.2014.07.001

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H. C., & Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia, 87*, 74-84. http://dx.doi.org/10.1016/j.neuropsychologia.2016.04.031

Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition, 59*, 1-21. http://dx.doi.org/10.1016/0010-0277(95)00683-4.

Mitchell, P., & Taylor, L. M. (1999). Shape constancy and theory of mind: Is there a link? *Cognition, 70*(2), 167–190. https://doi.org/10.1016/S0010-0277(99)00011-6

Moore, M. S. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford: Oxford University Press.

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster

than non-causal events, *PeerJ, 5*, e2932. http://dx.doi.org/10.7717/peerj.2932

Moreno-Rios, S., & García-Madruga, J. A. (2002). El desarrollo del razonamiento sobre lo que podría haber ocurrido: condicionales indicativos y subjuntivos. *Infancia y Aprendizaje*, *25*(3), 485–498. http://dx.doi.org/10.1174/021037002762064055

Moreno-Ríos, S, García Madruga, J. A., & Byrne, R.M.J. (2008). Inferences from semifactual "even if" conditionals. *Acta Psychologica, 128*(2), 197-209. https://doi.org/10.1016/j.actpsy.2007.12.008

Moreno-Ríos, S., Rodríguez-Menchen, M. A., & Rodríguez-Gualda, I. (2011). Bias in perspective-taking during reading: Adjusting the knowledge of characters. *Journal of Pragmatics, 43*(7), 1977-1986. doi.org/10.1016/j.pragma.2010.12.013

Muentener, P., & Bonawitz, E. (2017). *The development of causal reasoning.* In M. R. Waldmann (Ed.), *Oxford library of psychology. The Oxford handbook of causal reasoning* (p. 677–698). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2-to 6-year-old children. *Developmental Psychology, 12*(1), 85-86. https://doi.org/10.1037/0012-1649.12.1.85

Muentener, P., & Bonawitz, E. (2017). *The development of causal reasoning.* In M. R. Waldmann (Ed.), *Oxford library of psychology. The Oxford handbook of causal reasoning* (p. 677–698). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Musch, J., & Wagner, T. (2007). Did everybody know it all along? A review of individual differences in hindsight bias. *Social Cognition, 25*(1), 64–82. https://doi.org/10.1521/soco.2007.25.1.64

Mussweiler, T., Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*(2), 136-164. https://doi.org/10.1006/jesp.1998.1364

Nario, R.N., & Branscombe, N.R. (1995). Comparison Processes in hindsight bias and causal attribution. *Personality and Social Psychology Bulletin, 21*(12), 1244-1255. https://doi.org/10.1177/01461672952112001

Nava, E., and Pavani, F. (2013). Changes in sensory dominance during childhood: converging evidence from the colavita effect and the sound-induced flash illusion. *Child Development.* 84, 604–616. https://doi.org/10.1111/j.1467-8624.2012.01856.x

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1399-1413. http://dx.doi.org/10.1037/a0020449

Nestler, S., Blank, H., & von Collani, G. (2008a). Hindsight bias and causal attribution: A Causal Model Theory of Creeping Determinism. *Social Psychology, 39*(3), 182-188. http://dx.doi.org/10.1027/1864-9335.39.3.182

Nestler, S., Blank, H., & von Collani, G. (2008b). Hindsight bias doesn't always come easy: Causal models, cognitive effort, and creeping determinism. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*(5), 1043-1054. http://dx.doi.org/10.1037/0278-7393.34.5.1043

Nestler, S., & Egloff, B. (2009). Increased or reversed? The effect of surprise on hindsight bias depends on the hindsight component. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*(6), 1539-1544. http://dx.doi.org/10.1037/a0017006

Nestler, S., & von Collani, G. (2008a). Hindsight bias and the activation of counterfactual mind-sets. *Experimental Psychology, 55*(5), 342-349. http://dx.doi.org/10.1027/1618-3169.55.5.342

Nestler, S., & von Collani, G. (2008b). Hindsight bias, conjunctive explanations and causal attribution. *Social Cognition, 26*(4), 482-493. http://dx.doi.org/10.1521/soco.2008.26.4.482

Nickerson, R. S. (1999). How we know –and sometimes misjudge-what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 747-753. http://dx.doi.org/10.1037/0033-2909.125.6.737

Nilsen, E. S., Glenwright, M., & Huyder, V. (2011). Children and adults understand that verbal irony interpretation depends on listener knowledge. *Journal of Cognition and Development, 12*(3), 374-409. https://doi.org/10.1080/15248372.2010.544693

O'Connor, E., McCormack, T., & Feeney, A. (2014). Do children who experience regret make better decisions? A developmental study of the behavioral consequences of regret. *Child Development, 85*(5), 1995-2010. http://dx.doi.org/10.1111/cdev.12253

Oeberst, A., von der Beck, I., & Nestler, S. (2014). Reading about explanations enhances perceptions of inevitability and foreseeability: a cross-cultural study with Wikipedia articles. *Cognitive Processing, 15*, 343-349. https://doi.org/10.1007/s10339-014-0603-7

Olineck, K. M., & Poulin-Dubois, D. (2009). Infants' understanding of intention from 10 to 14 months: Interrelations among violation of expectancy and imitation tasks. *Infant Behavior & Development, 32*(4), 404–415. https://doi.org/10.1016/j.infbeh.2009.06.005

Perner, J. (1991). *Learning, development, and conceptual change. Understanding the representational mind.* The MIT Press.

Pexman, P. M. & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics*, *20*(2), 178-196. http://dx.doi.org/10.1016/j.jneuroling.2006.06.001

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., & Glenwright, M. R. (2011). Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(8), 1097–1112. http://dx.doi.org/10.1007/s10803-010-1131-7

Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory, 11*, 421-441. https://doi.org/10.1080/09658210244000603

Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction, 18*(1), 35–52. https://doi.org/10.1207/S1532690XCI1801_02

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition, 25*(1), 14–31. https://doi.org/10.1521/soco.2007.25.1.14

Pohl, R. F., Bayen, U. J., Arnold, N., Auer, T. & Martin, C. (2018). Age differences in processes underlying hindsight bias: A life-span study. *Journal of Cognition and Development, 19*(3), 278-300. https://doi.org/10.1080/15248372.2018.1476356

Pohl, R. F., Bayen, U. J., & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology, 46*(5), 1268-1282. https://doi.org/10.1037/a0020209

Pohl, R.F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology, 49*(4), 270-282. https://doi.org/10.1026/1618-3169.49.4.270

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory, 11*(4-5), 337-356. https://doi.org/ 10.1080/09658210244000487

Pohl, R. F., & Gawlik, B. (1995). Hindsight bias and the misinformation effect: Separating blended recollections from other recollection types. *Memory, 3*(1), 21-55. https://doi.org/ 10.1080/09658219508251495

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes, 67(1),* 49–58. https://doi.org/10.1006/obhd.1996.0064

Pohl, R. F., Schwarz, S., Sczesny, S., & Stahlberg, D. (2003). Hindsight bias in gustatory judgments. *Experimental Psychology, 50*(2), 107–115. https://doi.org/10.1026/1618-3169.50.2.107

Poulin-Dubois, D., & Yott, J. (2018). Probing the depth of infants' theory of mind: Disunity in performance across paradigms. *Developmental Science, 21*(4), e12600. https://doi.org/10.1111/desc.12600

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world." *Child Development*, *81*(1), 376–389. https://doi.org/10.1111/j.1467-8624.2009.01401.x

Rafetseder, E., O'Brien, C., Leahy, B., & Perner, J. (2021). Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective. *Journal of Experimental Child Psychology*, *204*, 105058. https://doi.org/10.1016/j.jecp.2020.105058

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

Ragni, M., Khemlani, S., & Johnson-Laird, P. N. (2014). The evaluation of the consistency of quantified assertions. *Memory and Cognition, 42*, 53-66. https://doi.org/10.3758/s13421-013-0349-y

Rai, R., Mitchell, P., Kadar, T., & Mackenzie, L. (2016). Adolescent Egocentrism and the Illusion of Transparency: Are adolescent as egocentric as we might think? *Current Psychology, 35*(3), 285-294. http://dx.doi.org/10.1007/s12144-014-9293-7.

Robinson, C. W., and Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401. http://dx.doi.org/10.1111/j.1467- 8624.2004.00747.x

Robinson, C. W., and Sloutsky, V. M. (2010). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology,107*, 351–358. https://doi.org/10.1016/j.jecp.2010.04.006

Rodríguez Rosique, S. (2001). Las construcciones condicionales concesivas en español. *Moenia*, *7*, 261–270. ISSN 1137-2346

Roese, N.J., & Olson, J.M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology, 32*(3), 197-227. https://doi.org/10.1006/jesp.1996.0010

Roldán-Tapia, M. D., Moreno-Ríos, S., & Cánovas-López, R. (2017). Thinking about social and nonsocial alternative possibilities in premature preschoolers. *Journal of Clinical and Experimental Neuropsychology*, *39*(8), 725–737. https://doi.org/10.1080/13803395.2016.1257703

Royzman, E. B., Cassidy, K. W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, *7*(1), 38-65. http://dx.doi.org/10.1037/1089-2680.7.1.38.

Ruíz-Ballesteros, J.A., & Moreno-Ríos, S. (2017). Concessive and semifactual interpretations during reasoning with multiple conditionals. *The Quarterly Journal of Experimental Psychology, 70*, 1140-1150. https://doi.org/10.1080/17470218.2016.1172098

Ruffman, T., Rustin, C., Garnham, W., & Parkin, A. J. (2001). Source monitoring and false memories in children: Relation to certainty and executive functioning. *Journal of Experimental Child Psychology*, *80*(2), 95-111. https://doi.org/10.1006/jecp.2001.2632

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Savitsky, K., Keysar, B., Epley, N., Carter, T., y Swanson, A. (2011). The closeness-communicaction bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269-273. http://dx.doi.org/10.1016/j.jesp.2010.09.005

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking Developmental Psychology and Functional Neuroimaging. *Annual Review of Psychology, 55*(1), 87-14. https://doi.org/10.1146/annurev.psych.55.090902.142044

Schkade, D.A., & Kilbourne, L.M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes, 49*(1), 105-123. https://doi.org/10.1016/0749-5978(91)90044-T

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Schwenter, S. A. (2001). Expectations and (in)sufficiency: Spanish como conditionals. *Linguistics*, *39*(4), 733–760. https://doi.org/https://doi.org/10.1515/ling.2001.030.

Sharman, S. J., Powell, M. B., & Roberts, K. P. (2011). Children's ability to estimate the frequency of single and repeated events. International Journal of Police Science & Management, *13*(3), 234-242. https://doi.org/10.1350/ijps.2011.13.3.243

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1), 1–51. https://doi.org/10.2307/1165893

Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development, 46*(2), 394–399. https://doi.org/10.2307/1128133

Siegler, R. S. (1978). Children's Thinking: What Develops? Hillsdale, NJ: Lawrence Erlbaum.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* Oxford University Press.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*(4), 481-520. https://doi.org/10.1016/0010-0285(76)90016-5

Sloutsky, V. M., and Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Dev.* 74, 822–833. https://doi.org/10.1111/1467-8624.00570

Sloutsky, V. M., and Robinson, C. W. (2008). The role of words and sounds in visual processing: from overshadowing to attentional tuning. *Cogn. Sci.* 32, 354–377. https://doi.org/10.1080/03640210701863495

Sluzenski, J., Newcombe, N. S., & Ottinger, W. (2004). Changes in reality monitoring and episodic memory in early childhood. *Developmental Science, 7*(2), 225–245. https://doi.org/10.1111/j.1467-7687.2004.00341.x

Ståhl, T., & Ellemers, N. (2016). Ironic effects of moral motivation: Why working toward a moral goal reduces subsequent perspective taking. *Social Cognition 34*(2),133-148. https://doi.org/10.1521/soco.2016.34.2.133

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*(3), 437–446. https://doi.org/10.1037/0022-3514.73.3.437

Suppes, P. (1970). *A probabilistic theory of causation*. Amsterdam: North-Holland.

Todd, A. R., Hanko, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. Psychological Science, 22(1), 134-141. https://doi.org/10.1177/0956797610392929

Turner, B.M., & Schley, D.R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology, 90*, 1-47. http://dx.doi.org/10.1016/j.cogpsych.2016.07.003

Tversky, A., & Kahneman, D. (1974). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293-315. https://doi.org/10.1037/0033-295X.90.4.293

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1130. http://dx.doi.org/10.1007/978-94-010-1834-0_8

Wasserman, D. Lempert, R., & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17*, 30-35. https://doi.org/10.1177/0146167291171005

Weingartner, K. M., & Klin, C. M. (2005). Perspective taking during reading: An on-line investigation of the illusory transparency of intention. *Memory & Cognition, 33*, 48–58. https://doi.org/10.3758/BF03195296

Weingartner, K. M., & Klin, C. M. (2009). Who knows what? Maintaining multiple perspectives during reading. *Scientific Studies of Reading, 13*(4), 275–294. https://doi.org/10.1080/10888430902851539

Weisberg, D. P., & Beck, S. R. (2012). The development of children's regret and

relief. *Cognition and Emotion, 26*(5), 820-835. https://doi.org/10.1080/02699931.2011.621933

Wilkening, F., & Anderson, N. H. (1991). Representation and diagnosis of knowledge structures in developmental psychology. In N. H. Anderson (Ed.), Contributions to information integration theory: Vol III Developmental (pp. 45-80). Hillsdale, NJ: Erlbaum

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgements and evaluations. *Psychological Bulletin, 116*(1), 117-142. https://doi.org/10.1037/0033-2909.116.1.117

Wiesmann, C. G., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others' thoughts in the developing brain. *Proc Natl Acad Sci U S A, 117*(12), 6928-6935. https://doi.org/10.1073/pnas.1916725117

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance, 4*(2), 345–353. https://doi.org/10.1037/0096-1523.4.2.345

Wu, R., Gopnik, A., Richardson, D., & Kirkham, N. (2011). Infants learn about objects from statistics and people. *Developmental Psychology, 47*, 1220–1229. https://doi.org/10.1037/a0024023

Yopchick, J.E., & Kim, N.S. (2012). Hindsight bias and causal reasoning: a minimalist approach. *Cognitive Processing, 13*, 63-72. https://doi.org/10.1007/s10339-011-0414-z

Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Causality and counterfactuals in group attributions. *Cognition, 125*(3), 429–440. https://doi.org/10.1016/j.cognition.2012.07.014

Zwick, R., Pieters, R., & Baumgartner, H. (1995). On the practical significance of hindsight bias: The case of the expectancy-disconfirmation model of consumer satisfaction. *Organizational Behavior and Human Decision Processes, 64*(1), 103–117. https://doi.org/10.1006/obhd.1995.1093