Hindsight Bias and False-Belief Reasoning from Preschool to Old Age

Daniel M. Bernstein

Kwantlen Polytechnic University

Author note

Daniel M. Bernstein, Department of Psychology, Kwantlen Polytechnic University

This work was supported by grants from the Canada Research Chairs Program (950-232078) and the Social Sciences and Humanities Research Council of Canada (435-2015-0721). Special thanks to all the families that participated, and to Andre Aßfalg, Eric Mah, Kyle Matsuba, and Cristina Atance for providing helpful feedback on drafts of this work. The data are available at https://osf.io/d9ae5/?view_only=c483e01336ef4c15a6f86f175603c53e

Original submission ID: DEV-2020-3734

Abstract

Participants ranging in age from 3 to 98 years (N = 708; approximately 60% female; 49% Caucasian, 38% Asian; 12% Other ethnicities, 1% Indigenous; modal household income > $80,000) completed a battery of tasks involving verbal ability, executive function, and perspective-taking. Wherever possible, all participants completed the same version of a task. The current study tested hindsight bias and false-belief reasoning to determine how these constructs relate to each other across the child-to-adult lifespan. Participants of all ages showed robust hindsight bias and false-belief reasoning errors. Hindsight bias followed a U-shaped function, wherein preschoolers and older adults showed more hindsight bias than older children and younger adults. False-belief reasoning, conversely, was relatively constant from preschool to older adulthood. Hindsight bias did not correlate with false-belief reasoning. We conclude that hindsight bias and false-belief reasoning errors are robust but unrelated cognitive biases across the lifespan.

Keywords: lifespan, hindsight bias, curse of knowledge, theory of mind, mentalizing, false-belief reasoning

Hindsight Bias and False-Belief Reasoning From Preschool to Old Age

In December 2019, a strange new virus emerged that caused severe respiratory illness. Over the coming months, COVID-19 quickly spread throughout the world, hospitalizing and killing millions of people. We write this in the midst of COVID-19's devastation, while living in history (Brown et al., 2009). It remains to be seen how well individual countries or the world handled this pandemic. History will reveal the path taken, while suggesting alternate paths that could have or should have been taken. *Hindsight bias* imbues the world with a sense of clarity and meaning that did not exist while events were unfolding. It is hard to take a naïve perspective when we know the outcome to a problem or pandemic.

Perspective taking is a hallmark of human social cognition. Although we use social cognition throughout life, few studies use a lifespan approach to explore its development. Social cognition affects social competence and allows us to communicate and empathize with others through our assessment of what others know and feel (Decety & Jackson, 2004; Repacholi & Slaughter, 2003). Our own knowledge and feelings, though, can limit our ability to take another person's perspective, or know how another person feels.

The current work explores two types of social cognition – hindsight bias and false-belief reasoning – across the lifespan. Hindsight bias is the over-estimation of one's own prior or another person's naïve knowledge when one knows the outcome to a situation or problem (Fischhoff, 1975; Pohl & Erdfelder, 2017). *False-belief reasoning* is an aspect of what is called mentalizing or theory of mind, which is the understanding that other minds may differ from one's own mind and that people can hold mistaken beliefs about the world (Premack & Woodruff, 1978; Wellman, 1990). We regard both hindsight bias and false-belief reasoning errors as cognitive biases or systematic errors in thinking that involve ignoring knowledge to reason from a naïve perspective (Kahneman, 2011).

Some Background

In a classic hindsight-bias task, adults predict the outcome to an event. Upon learning the outcome, they try to recall their prediction. Most adults fail to ignore the outcome and say they “knew it all along” (Fischhoff, 1975; Wood, 1978). Hindsight bias, also called the curse of knowledge*,* occurs in many domains, including medicolegal and economic decisions, consumer satisfaction, sports and election outcomes (Arkes, 2013; Giroux et al., 2016; Hawkins & Hastie, 1990; Roese & Vohs, 2012). In each case, privileged knowledge biases judgments by making people think they knew the outcome before they learned it.

Hindsight bias conceptually resembles children’s failure to understand false beliefs. In a classic false-belief task, preschoolers watch a story in which Maxi places chocolate in the cupboard before leaving the room. While Maxi is away, Suzy moves the chocolate to a box. When Maxi returns, children report where he will look for the chocolate. Most 3-year-olds incorrectly say, “box.” By 4.5 years of age, most children appreciate that Maxi has a false belief about where the chocolate is and correctly say, “cupboard” (Baron-Cohen et al., 1985; Wimmer & Perner, 1983; Wellman et al., 2001).

Most hindsight bias and false-belief studies involve adults and children, respectively. Besides surface-level similarities, the two constructs conceptually relate because both adults and children fail to ignore what they know when estimating what another person knows (see Birch & Bloom, 2003; Massaro et al., 2014; Pinker, 2014; Royzman et al., 2003; Taylor, 1988).

Despite overlap between hindsight bias and false-belief reasoning, the two constructs differ. One difference is that hindsight bias involves uncertain judgments, while false-belief reasoning involves beliefs that we hold with certainty (Bernstein et al., 2007). In the classic false-belief task, children know that Maxi placed the chocolate in the cupboard before Suzi moved it. Introduce uncertainty to false-belief tasks, and we see older children and adults answering egocentrically much like children (Birch & Bloom, 2007). Differences in task structure between hindsight and false-belief tasks may obscure their conceptual link. Hindsight tasks tend to be continuous, and false-belief tasks tend to be categorical. One consequence of this task difference is that many researchers treat hindsight bias as a continuous construct and false-belief reasoning as a categorical construct.

Despite differences, there is reason to study links between hindsight bias and false-belief reasoning, and to do so across the lifespan. Lifespan studies inform theory and practice about the development of cognitive biases (Baltes, 1987; Siegler, 2005). Understanding how hindsight bias and false-belief reasoning relate across development can inform theories of cognitive development and social cognition. The current work's focus on much of the human lifespan can help answer questions about how hindsight bias and false-belief reasoning express at different time points in human development (Bernstein, 2018). There are many theoretical accounts of hindsight bias and false-belief reasoning (Blank et al., 2008; Perner & Roessler, 2012). Few, however, assume a lifespan perspective or link the two constructs (Mitchell et al., 1996; Royzman et al., 2003). There are two large and separate literatures on hindsight bias and false-belief reasoning. Relatively few developmental psychologists are interested in hindsight bias, and relatively few cognitive psychologists are interested in false-belief reasoning. By showing a link or lack thereof between hindsight bias and false-belief reasoning from a lifespan perspective, we will learn about general human cognitive development. If these are related constructs, researchers will benefit from fusing the literatures on hindsight bias and false-belief reasoning.

Conversely, if these are unrelated constructs, researchers will save time, effort, and resources building theories that treat these constructs separately.

As has been noted about hindsight bias, “A shortcoming of the existing literature is that…no common methodology to address developmental issues across the entire life span has been developed” (Bayen et al., 2007, p. 84). The same applies to theory of mind (cf. Lagattuta et al., 2014). To study development across a wide age range, especially one spanning childhood to adulthood, researchers often must use different tasks to measure the same construct in different age groups. This works if the different tasks have construct validity, or they measure what they should measure. Different tasks, however, can produce performance differences that have nothing to do with developmental differences. Ideally, it is better to hold task differences constant across age groups, thereby requiring that the same task be used to measure a given construct in all age groups tested. Some prior work on hindsight bias and false-belief reasoning has done just that.

In one study, 3- to 5-year-olds and adults completed a visual hindsight task in which they identified gradually clarifying images of degraded common objects on a computer. Sometimes observers knew in advance a clarifying object’s identity, and then estimated when a naïve peer would identify it. Children and adults showed robust hindsight bias by using their advance knowledge to overestimate their peers’ ability to identify the objects (Bernstein et al., 2004). Other cross-sectional and cross-experimental work using the visual hindsight task and other tasks reveals that hindsight bias follows a U-shaped pattern from preschool to old age such that preschoolers and older adults show more hindsight bias than do older children and younger adults (Bayen et al., 2007; Bernstein, Erdfelder et al., 2011).

Cross-experimental work using different tasks in different age groups shows the following with respect to false-belief reasoning: Possible emergence in infancy (Onishi & Baillergeon, 2005); steep increase from 3 to 5 years of age (Wellman et al., 2001); gradual increase in adolescence (Carpendale & Chandler, 1996); relative stability in younger adults, followed by gradual decline in older adults (Henry et al., 2013; see Miller, 2012). To address shortcomings from using different tasks to measure the same construct in different ages, researchers have developed the *Sandbox* task to measure false-belief reasoning in children and adults (Sommerville et al., 2013). In this task, participants watch an experimenter enact stories with two characters. First, Sally puts a toy dog in one location (Location 1 or L1) in a five-foot long Styrofoam-filled box, and then leaves the room. Next, Ann moves the dog elsewhere in the box (Location 2 or L2). Upon Sally's return, participants indicate where she will look for her dog (false-belief condition) or where she first put the dog (memory-control condition). The Sandbox task allows one to measure false-belief errors on a continuum as a bias toward L2 while controlling for memory for L1. The Sandbox task also introduces uncertainty into false-belief reasoning by making L1 and L2 just two locations within many possible locations.

In one study involving the Sandbox task, preschoolers and adults showed more bias in the false-belief condition than the memory-control condition. Thus, preschoolers and adults showed an egocentric bias by indicating that Sally would look for her toy dog somewhere near L2, a location that participants but not Sally knew the dog to be (Sommerville et al., 2013). Other work involving the Sandbox task has shown that false-belief reasoning remains relatively stable from preschool to older childhood, followed by a small but significant decline from younger to older adulthood (Bernstein et al., 2017; Bernstein, Thornton, & Sommerville, 2011). These authors argue that the Sandbox task controls for memory demands within the task or what they call task-specific memory, whereas other tasks do not. Controlling for task-specific memory, the authors contend, could explain the relatively flat developmental pattern in false-belief reasoning observed from preschool to adulthood that differs from the developmental pattern that emerges with other false-belief tasks (Bernstein et al., 2017).

The prior lifespan work involving the visual hindsight task and the Sandbox false-belief task has some limitations. Prior work on hindsight bias involved preschoolers through adolescents (Ages 3 – 15 years), younger adults (Ages 18-29 years), and a small sample of older adults (n = 22; Ages 61-95 years). Older adults were community-dwelling, but were unscreened for neurological problems, such as stroke and head trauma (Bernstein, Erdfelder et al., 2011). Prior work on false-belief reasoning involved preschoolers (Ages 3 and 5 years), adolescents (Ages 9-12 years), younger adults (Ages 17-25 years), and older adults (Ages 65-92 years). The latter work also combined previously-published data with new data in different age groups who received different task orders (Bernstein et al., 2017). The current study attempted to replicate both these prior studies, but used a single, large lifespan sample that completed the same tasks in the same order. The tasks that our participants completed were continuous measures that showed no ceiling or floor effects, thereby permitting us to measure variability in hindsight bias and false-belief reasoning across the lifespan. We also included measures of verbal ability and executive function, which prior work omitted. We screened older adults for various problems. The current study also permitted us to test the correlation between hindsight bias and false-belief reasoning across the lifespan, something that no prior study has done. Based on the assumption that hindsight bias and false-belief reasoning are related constructs, we hypothesized that:

1. Hindsight bias and false-belief reasoning errors will follow a U-shaped pattern of development from preschool to old age.

2. Hindsight bias and false-belief reasoning errors will correlate positively.

Method

Participants

Our full sample consisted of 708 participants, ranging in age from 2 years, 6 months to 98 years; however, the youngest participant to complete our two primary measures was 3 years, 0 months. Participants either came from the community or from a mid-sized North American university. Community participants received $25, while students received course credit. Testing occurred in a university laboratory, community centre, or the participant's home. We calculated age as a continuous variable by subtracting the date of testing from the participant's birthday. We also divided participants into five age groups to facilitate comparison to prior lifespan work on hindsight bias and false-belief reasoning (Bernstein, Erdfelder et al., 2011; Bernstein et al., 2017). Our groups consisted of 107 3.0- to 5.99-year-olds (M = 4.48 years, SD = 0.95; 47 female; 5 missing sex); 114 6.0- to 9.99-year-olds (M = 7.84 years, SD = 1.12; 58 female; 2 missing sex); 139 10.0- to 17.99-year-olds (M = 13.44 years, SD = 2.31; 62 female; 2 missing sex); 247 18.0- to 64.99-year-olds (M = 41.92 years, SD = 14.36; 179 female; 13 missing sex); and 101 65+ year-olds (M = 73.29 years, SD = 6.27; 73 female). Age grouping occurred on the year, and was not rounded up or down; so, a child aged 5.99 years was categorized as a 5-year-old.

A total of 558 participants indicated their ethnicity: 273 Caucasian; 214 Asian; 6 Indigenous; 65 Other. Modal annual household income for parents/guardians of child participants was over $80,000 (USD). We did not ask adult participants to report their household income. Adults older than 55 years of age completed the Mini Mental State Exam to control for mild cognitive impairment and overall mental health (Folstein, Folstein, & McHugh, 1975). All but 1 of these participants scored in the normal range (24 or better). Older adults were screened for various problems (e.g., diagnosis of major psychotic illness or dementia; stroke, head trauma; consumption of 3 or more alcoholic beverages per day). Thus, our older adults were generally healthy. All participants older than 16 years of age provided informed consent, while children provided verbal assent and their parents/guardians provided informed consent. This study received ethics approval from Kwantlen Polytechnic University's Research Ethics Board; protocol #2015-036; study title: Lifespan Cognition.

Measures

Participants completed a battery of measures involving verbal ability, executive function, and perspective-taking. The primary measures here tapped perspective-taking. Measures related to verbal ability and executive function served as potential covariates and mediators.

Verbal Ability

Kaufman Brief Intelligence Test—Second Edition (KBIT-2; Kaufman & Kaufman, 2005). The KBIT-2 is a measure of receptive verbal intellectual functioning (hereafter referred to as verbal ability) in which the experimenter says a word or asks a factual question and the participant points to one of four pictures depicting that word or fact. We administered the vocabulary section of this test to assess general verbal ability. The task took approximately 5 minutes to complete. We administered this task to all participants.

Executive Function

Inhibitory Control

Day/Night (Gerstadt, Hong, & Diamond, 1994). This task is a measure of inhibitory control. Children were instructed to say "Day" when they saw a card with a picture of the moon and stars on it, and "Night" when they saw a card with a picture of the sun on it. This task consisted of two practice trials with corrective feedback, followed by 16 test trials. Children received 1 point for each correct response and 0 for each incorrect response. Children's first response was always scored, even in those instances in which there was a self-correction. We then calculated a proportion correct score. The task took approximately 5 minutes to complete. We administered this task to children ages 3-5 years.

Stroop Color Word (Stroop, 1935). Participants indicated the color of a word printed on the computer monitor. Participants completed 48 randomized trials. On half the trials, the color and word were *congruent* (e.g., the word, green, appeared in green). On the remaining trials, the color and word were *incongruent* (e.g., the word, green, appeared in blue). Participants indicated the color of the word by pressing the corresponding color marked on the keyboard. We calculated an error score consisting of the difference between total number of errors on incongruent trials minus total number of errors on congruent trials (Stroop Error Difference). Positive scores resulted from more errors on incongruent than on congruent trials, indicating a lack of inhibitory control. We also calculated a reaction-time score consisting of the difference between average reaction times on correctly-answered incongruent trials minus correctly-answered congruent trials (Stroop RT difference). Positive scores here indicated faster reaction times for congruent words, indicating a lack of inhibitory control. The task took approximately 5 minutes to complete. We administered this task to all participants 6 years of age or older.

Working Memory

Digit Span (Wechsler, 1958). Participants repeated in order single digits between 1 and 9 that the experimenter said aloud. Trials began with two digits and increased by one digit until participants missed three consecutive trials. We recorded the highest level of success obtained, with a minimum of 0 up to a maximum of 10. The task took approximately 5 minutes to complete. We administered this task to all participants.

Perspective Taking

Hindsight Bias

Visual hindsight task (Bernstein et al., 2007; see also Bernstein, Erdfelder et al., 2011 for nearly identical materials and procedure as used here). This task is a measure of visual hindsight bias. We used three objects, measuring up to 5 in. (12.7 cm) long and 5 in. high: green and white airplane, orange and brown alligator, and red car. The experimenter placed each object on a platform eye level to the participant. This platform stood inside a plastic milk crate with the participant facing the crate's open side. A black piece of paper covered the back of the crate. Ten different laminated transparency sheets were placed in a three-ring binder, which sat atop the crate. The sheets hung in front of the toy, obscuring its appearance. Each sheet contained a unique set of black dots that covered 5% of the sheet's surface area. Behind 10 blank sheets, each object is easily identified; however, behind the 10 sheets used here, 50% of each object is occluded and impossible to identify. The experimenter introduced this task as the "hide it game". Participants sat in a chair located 15 inches from the plastic box. At the start of the visual hindsight task, the experimenter asked participants to put on a black apron to reduce glare from the transparency sheets. Participants had to identify the object correctly on two consecutive sheets (e.g., transparency sheet number 6 and 7) before proceeding to the next trial. We counted the first correct identification as the correct identification point.

Participants first completed a *baseline* condition in which they tried to identify each object as it gradually became clearer. The experimenter started by asking participants, "What does this look like to you?" The experimenter then removed one sheet at a time, and asked participants, "What does it look like to you now?" after each sheet had been removed. When participants correctly identified the object, the experimenter recorded how many sheets had been removed. This score ranged from 1 to 10, where 1 meant that participant identified object on the first screen, and 10 meant that participant identified object after all 10 screens had been removed.

In the *Hindsight* condition, participants saw the object at the start of each trial. The experimenter then covered the object with all 10 filter sheets and asked participants to indicate when someone named Joe or Molly for male or female participants, respectively, could identify the object. Participants 3 to 5 years old played the game with a doll, while older participants played with an imaginary character. Participants learned that Joe/Molly was the same age and just as smart as they were, but that Joe/Molly hadn't played this game before. For preschoolers, the experimenter kept Joe/Molly hidden from view at the start of each Hindsight trial to reassure participants that Joe/Molly could not see the object. The experimenter started by asking participants, "What does this look like to Joe/Molly?" The experimenter then removed one sheet at a time, and asked participants, "What does it look like to Joe/Molly now?" after each sheet had been removed. The experimenter recorded the point at which participants reported that Joe/Molly could identify the object, by recording the number of sheets that had been removed. As with the Baseline trials, this score ranged from 1 to 10. After the participant indicated this identification point for Joe/Molly on two consecutive screens, the next trial began. Participants completed the three baseline trials before completing the three hindsight trials. We used a single fixed order of presentation for the three objects for the baseline and hindsight trials: plane, alligator, car.

Following prior work, we calculated hindsight bias as the difference between the average identification point in the baseline condition minus the average identification point in the hindsight condition (Bernstein, Erdfelder et al., 2011). Average baseline and average hindsight scores were computed for any participant who had scores on at least 2 baseline or 2 hindsight trials, respectively. For example, a participant who identified objects on average at the seventh filter screen in the baseline condition and the fifth filter screen in the hindsight condition would obtain a hindsight score of 2. Thus, the greater the hindsight bias score, the more hindsight bias the participant showed. In this example, the participant required more visual information to correctly identify the object in the baseline condition than they required in the hindsight condition. The task took approximately 8 minutes to complete. We administered this task to all participants.

False-belief Reasoning

Sandbox task (Sommerville et al., 2013; see also Bernstein et al., 2017 for nearly identical materials and procedure as used here. See https://youtu.be/TpkjruhH8a4 for example). This task is a measure of false-belief reasoning. The Sandbox task involved a rectangular Styrofoam-filled box 60 inches long by 18 inches wide by 12 inches deep. The experimenter enacted four stories, one per trial, each involving two characters and a hidden object. The physical Sandbox represented different containers in each story (sandbox, freezer, bathtub, planter box). In each story, a protagonist placed a small object in the Sandbox at one location (L1) and then exited. While the protagonist was away, a second character moved the object to another location (L2) within the Sandbox. Participants received one critical question at the end of each trial: *False-belief* trials required participants to take the perspective of the protagonist, who had a false belief about the object's location (e.g., “Where will Sally look for the toy dog?”); *memory-control* trials required participants to remember the object's initial location (e.g., “Where did Sally put the toy dog before she left?”). In the middle of each trial, participants completed a 20-second visual search filler task before receiving the false-belief or memory-control question. L2 appeared to the left or right of L1, respectively, on half the trials. The distance between L1 and L2 was 14 inches on all trials, and L1 and L2 occupied different locations throughout the Sandbox container across trials. Objects were represented by props on each trial (e.g., toy dog; ice cream cone); story characters were not represented by props. We counterbalanced trial order such that half the participants in each age group completed two false-belief followed by two memory-control trials. The remaining participants completed two memory-control trials followed by two false-belief trials.

One reason that researchers use the memory-control condition in the Sandbox task is that this is the control condition that researchers developed in early work on false-belief reasoning. Specifically, this question ensures that the child has an accurate memory of the first location in standard change-of-location tasks (Baron-Cohen et al., 1985). The memory-control question continues to be one of the most-widely used control questions in child false-belief tasks (e.g., change-of-location and change-of-contents). Sommerville et al. (2013) based the Sandbox task, in part, on the standard change-of-location task used in developmental science. They included memory-control questions in the Sandbox task to assess whether participants could remember an object's original location when they know its current location.

Following prior work, we calculated bias scores as the distance in inches between L1 and the participant's response on each trial (Sommerville et al., 2013). Positive bias scores resulted when participants chose a location in the direction of L2. Negative bias scores resulted when participants chose a location away from L2. For example, assume that L1 equals 10 and L2 equals 24. If the participant's response was 18, then this would result in a bias score of 8. If the participant's response was 5, then this would result in a bias score of -5. We then averaged these bias scores for the false-belief and memory-control conditions, respectively. We computed averages only for participants who completed both trials in each belief condition (2 false belief, 2 memory control). Next, for each participant, we calculated a difference score (henceforth *egocentric bias*) by taking the difference between the average false-belief bias minus the average memory-control bias. A positive egocentric bias score (false-belief bias > memory-control bias) indicated difficulty suppressing one's own knowledge to reason about another person's false beliefs. Thus, the greater the egocentric bias score, the poorer false-belief reasoning the participant showed. The task took approximately 8 minutes to complete. We administered this task to all participants.

Procedure

Participants completed a battery of tasks individually with a student experimenter in a quiet room as part of a larger social cognition study. The current work reports on the subset of tasks listed under Measures. The experimenter told participants that they would be playing a set of games together. Participants completed the KBIT-2, visual hindsight task baseline trials, Sandbox, 2 tasks not reported here, Day/Night or Stroop, Digit Span Forward, visual hindsight task hindsight trials, and then 4 tasks not reported here. There was a delay of approximately 30 minutes between the visual hindsight baseline and hindsight trials. Participants took breaks as needed. Total testing time ranged from 50-90 minutes.

Results

The data are available at https://osf.io/d9ae5/?view_only=c483e01336ef4c15a6f86f175603c53e

We present results around our two hypotheses, but first we present zero-order correlations between our various measures to determine whether our sample behaved as expected in terms of verbal ability and executive function (see Table 1). These correlations also permitted us to determine which, if any, of our verbal ability and executive function measures should be controlled for in our main analyses. As expected, KBIT-2 scores correlated highly with age (r =

.732). As for our executive-function measures, Digit Span scores correlated with age (*r* = .401) and hindsight bias (*r* = -.193), but not with egocentric bias. In preschoolers, Day/Night scores correlated with participant age (*r* = .523), but not with hindsight bias or egocentric bias. For the rest of our sample, Stroop Error Difference correlated with participant age (*r* = .148), but not with hindsight bias or egocentric bias. Stroop RT Difference did not correlate with participant age, hindsight bias, or egocentric bias. Overall, the significant correlations between age, verbal ability, and executive function suggest that our sample performed as expected.

We conducted two overall sets of analyses on our two main tasks, visual hindsight and Sandbox. The first set of analyses involved categorical age instead of continuous age. We ran two separate one-way ANOVAs with hindsight bias and egocentric bias, respectively, as the dependent variable, and age group as the independent variable. We tested linear, quadratic, and cubic effects to provide an inclusive but not exhaustive set of tests of linear and non-linear functions. These analyses facilitated comparison to prior lifespan work involving hindsight bias and false-belief reasoning (Bernstein, Erdfelder et al., 2011; Bernstein et al., 2017). Next, we ran two multiple regressions using mean-centered age group and mean-centered age group that we then squared and cubed to predict hindsight bias (or egocentric bias), controlling for verbal ability. In the regression analysis predicting hindsight bias, we also controlled for working memory, because it was the only executive function that correlated with either hindsight bias or egocentric bias.

Our second overall set of analyses involved continuous age instead of categorical age. Unlike prior lifespan work involving the visual hindsight task and the Sandbox task, the current work allowed us to investigate the effects of continuous age on hindsight bias and egocentric bias across the lifespan. We did this in two ways. First, we investigated the role of verbal ability as a potential mediator of the link between continuous age and hindsight bias and between continuous age and egocentric bias. We performed two simple mediation analyses using the PROCESS macro in SPSS. The outcome variable was hindsight bias (or egocentric bias), and the predictor variable was continuous age. The mediator was KBIT-2 score. The second way that we investigated the effects of continuous age on hindsight bias and egocentric bias across the lifespan was by running two Bayesian regressions in JASP (JASP Team, 2020). Hindsight bias (or egocentric bias) served as the dependent variable, and continuous age served as the independent variable. As with the ANOVAs, we tested linear, quadratic, and cubic age effects. For these analyses, we mean-centered age and then squared and cubed those values. We also performed Bayesian analyses on our categorical age data, but do not report those here for brevity sake. The Bayesian results of our ANOVAs were consistent with the Bayesian regressions. For Bayesian analyses, we used JASP's default priors, where prior probabilities (p(M)) were lower for more complex models.

Conventional null-hypothesis significance tests provide a *p*-value indicating statistical significance, but Bayesian hypothesis tests offer a numerical value called the Bayes Factor (BF). The BF expresses evidence in favor of the alternative model over the null model, typically denoted as H1 (evidence for the alternative model) versus H0 (evidence for the null model). A BF of 1 suggests no evidence for either hypothesis; a BF above 1 indicates increasing evidence for H1 over H0; a BF below 1 indicates increasing evidence for H0 over H1 (Dienes, 2014). Thus, compared to the conventional significance test, the BF can express evidence in favor of a null hypothesis instead of being inconclusive when $p > .05$.

Does Hindsight Bias Follow a U-Shaped Pattern of Development?—Yes, It Does

For the visual hindsight task, we had complete data for 678 participants. Table 2 shows the average identification point for the baseline and hindsight conditions in the visual hindsight task by age group. Worth noting in Table 2 is that preschoolers and older adults required the removal of more filter screens to identify objects in Baseline compared to older children and younger adults. Figure 1 shows the difference between average hindsight and baseline identification points by age group. This difference score was our measure of hindsight bias. One advantage of collecting data from a large lifespan sample is that one can observe developmental patterns using age as a continuous variable. Figure 2 shows hindsight bias as a function of age in years.

All age groups showed robust and significant hindsight bias (see Figures 1 and 2). This indicates that participants had trouble ignoring their own knowledge when reasoning about a naïve character's knowledge. Consistent with our first hypothesis, a one-way ANOVA with categorical age as the independent variable and hindsight bias as the dependent variable revealed a U-shaped pattern in hindsight bias: $F(4, 673) = 26.09$, $p < .001$, partial eta-squared = .134 for the combined term; $F = 23.00$, $p < .001$ for the unweighted linear term; $F = 81.45$, $p < .001$ for the unweighted quadratic term and $F = 3.78$, $p = .023$ for the deviation term; $F = 7.52$, $p = .006$ for the unweighted cubic term and $F < 1.0$ for the deviation term (see Figure 1). This data pattern shows significant linear, quadratic, and cubic effects, but only the quadratic deviation accounted for significant variance over its prior term in the model. The overall data pattern replicates prior work (Bernstein, Erdfelder et al., 2011). Following up on the significant age group effects, using Bonferroni correction and $p < .0062$ as the criterion, we observed that preschoolers showed more hindsight bias than all other age groups (all p values < .001). We also found that 10- to 17-year-olds showed less hindsight bias than older adults ($p = .001$). We next explored change in hindsight bias in preschoolers only, the age during which one typically observes dramatic improvement in false-belief reasoning. Hindsight bias was unrelated to age in years and months, a continuous variable (r = .084). The latter result is surprising, given the marked improvement that one typically observes in false-belief reasoning as preschoolers age.

Next, we conducted a multiple regression to control for verbal ability and working memory. We entered KBIT-2, then digit span, mean-centered age group, mean-centered age group squared, and mean-centered age group cubed. KBIT-2 accounted for significant variance in hindsight bias (*R-squared* = .045, p < .001). Digit span did not add significantly to the model (change to *R-squared* = .005, p = .053). Age group added significantly to the model (change to *R-squared* = .006, p = .036), as did age group squared (change to *R-squared* = .069, p < .001), and age group cubed (change to *R-squared* = .012, p = .002). Thus, the quadratic and cubic nature of the relation between age group and hindsight bias held after controlling for verbal ability and working memory.

A stricter test of any quadratic effect requires simultaneously observing both a significantly decreasing line in an interval of the independent variable and a significantly increasing line in another interval of the independent variable (Simonsohn, 2018). Our data met both criteria using age group (categorical) and participant age (continuous), lending further support to the U-shape pattern that we observed. Figures 1 and 2 show that visual hindsight bias stops decreasing and starts increasing between 10-17 years of age.

Finally, we performed two different mediation analyses using continuous age instead of categorical age. In the first analysis, we tested whether verbal ability mediated the link between age and hindsight bias. The total effect was not significant [Effect = -.001, 95% CI (-.005, .004)]. The direct effect of age on hindsight bias was significant [Effect = .019, 95% CI (.013,

.025)], and so was the indirect effect of age on hindsight bias mediated by verbal ability [Effect = -.020, 95% C.I. (-.025, -.014)]. Similar results emerged when using age squared instead of age as the predictor variable. Thus, verbal ability mediated the link between age and hindsight bias and between age squared and hindsight bias. Given that we did not hypothesize this effect and that age and verbal ability were highly correlated, we are hesitant to interpret this result. Future work should replicate this result and investigate it further.

In the second analysis using continuous age instead of categorical age, we ran a Bayesian regression. This resulted in extreme evidence for H1 (BF > 100) involving a combination of age, age squared, and age cubed functions. The Bayesian regression also revealed extreme evidence for H1 involving a combination of age and age squared, and age squared and age cubed functions (see Table 3). All three pieces of extreme evidence for H1 involved age squared. These results confirm that the relation between age and hindsight bias is non-linear, something easily seen in Figures 1 and 2. The overall age by hindsight-bias data pattern reveals that young children and older adults show more hindsight bias than do older children and younger adults.

Does False-Belief Reasoning Follow a U-Shaped Pattern of Development?—No, It Does Not

For the Sandbox task, we had complete data for 683 participants. Table 4 shows the average bias score for the memory-control and false-belief conditions in the Sandbox task by age group regardless of whether participants completed both conditions. Worth noting in Table 4 is that the average memory-control and false-belief bias scores are larger for preschoolers than older children and adults. This is to be expected from what we know about false-belief reasoning in preschoolers versus older children and adults. Not shown directly in Table 4, but unexpected, is that the difference between memory-control and false-belief bias scores is significantly greater than zero and similar in magnitude in all age groups. This indicates that all age groups were egocentric, and egocentric to a similar extent. Figure 3 shows the difference between the average false-belief and memory-control bias scores by age group. This difference score was our measure of egocentric bias. Figure 4 shows egocentric bias as a function of age in years.

Appendix A shows the trial-level performance on the Sandbox task. Of note in the histograms is that participants showed more bias on each of the false-belief trials than they showed on the memory-control trials. Also of note is that participants' bias scores tended to be bimodal, meaning that participants mainly chose either L1 (the correct location) or L2 (the incorrect location). That said, there were still many bias scores that fell between L1 and L2. It is these intermediate bias scores that reflect the continuous nature of false belief reasoning, as captured by the Sandbox task.

All age groups showed robust and significant egocentric bias (see Figure 3). This indicates that participants had more trouble inhibiting their own knowledge when reasoning about characters' false beliefs than they did remembering the objects' initial hiding locations. A one-way ANOVA with categorical age as the independent variable and egocentric bias as the dependent variable revealed small differences in egocentric bias across the lifespan: $F(4, 678) = 2.01$, $p = .091$ for the combined term; $F = 6.93$, $p = .009$ for the unweighted linear term; $F < 1.0$ for the unweighted and deviation quadratic and cubic terms (see Figure 3). This data pattern partially replicates prior work that showed little difference in false-belief reasoning on the Sandbox task from preschool to younger adulthood, followed by a small but significant decline in false-belief reasoning from younger to older adulthood (Bernstein et al., 2017). Following up on the linear effect, using Bonferroni correction and $p < .0062$ as the criterion, we observed no age differences in egocentric bias (all p values > .14).[1] We next explored change in egocentric bias in preschoolers, the age during which one typically observes dramatic improvement in false-belief reasoning. In preschoolers, egocentric bias was unrelated to age in years and months, a continuous variable (r = -.004).[2] Thus, although all age groups showed egocentric bias, all age groups also showed similar amounts of this bias (see Figures 3 and 4).

Next, we conducted a multiple regression to control for verbal ability. We entered KBIT-2, then mean-centered age group, mean-centered age group squared, and mean-centered age group cubed. KBIT-2 accounted for significant variance in egocentric bias *(R-squared* = .010, p = .008). None of the age group variables added significantly to the model (all *ps* > .587). Thus, only verbal ability contributed significant variance to egocentric bias.

Finally, we performed two different mediation analyses using continuous age instead of categorical age. In the first analysis, we tested whether verbal ability mediated the link between age and egocentric bias. The total effect was significant [Effect = -.028, 95% CI (-.052, -.050)]. The direct effect of age on egocentric bias was not significant [Effect = -.011, 95% CI (-.045, .024)], and neither was the indirect effect of age on egocentric bias [Effect = -.018, 95% C.I. (-.050, .014)]. Thus, verbal ability did not mediate the link between age and egocentric bias.

In the second analysis using continuous age instead of categorical age, we ran a Bayesian regression. This resulted in anecdotal evidence for H1 (BF = 1.941) involving age as a predictor of egocentric bias. All other models provided moderate to strong evidence for H0 (see Table 5). These results confirm that egocentric bias was largely age invariant, something easily seen in Figures 3 and 4.[3]

Does Hindsight Bias Relate to False-belief Reasoning?—No, It Does Not

In total, 664 participants completed both the visual hindsight task and the Sandbox task. Internal consistency was good for the visual hindsight task based on the 3 baseline items (Cronbach's alpha = .86) and the three hindsight items (Cronbach's alpha = .86). Internal consistency was poor for the Sandbox task based on the two memory-control items (Cronbach's alpha = .51) and the two false-belief items (Cronbach's alpha = .52). Across all ages, hindsight bias did not correlate with egocentric bias, our measure of false-belief reasoning (r = -.004).[4] Within each age group, this correlation remained null: Ages 3-5, r = -.043; Ages 6-9, r = -.098; Ages 10-17, r = -.044; Ages 18-64, r = .019; Ages 65+, r = -.002). Thus, hindsight bias was unrelated to false-belief reasoning.

Discussion

The current study used a large lifespan sample to determine how hindsight bias and false-belief reasoning relate to age and to each other. Addressing our two hypotheses, we found that: (1) Hindsight bias followed a U-shaped function from preschool to older adulthood; false-belief reasoning, conversely, was relatively constant from preschool to older adulthood. (2) Hindsight bias did not correlate with false-belief reasoning.

Regarding our first hypothesis, hindsight bias was significant in all age groups, but greater in preschoolers than in all other age groups. Also, hindsight bias was greater in older adults than in 10- to 17-year-olds. This U-shaped data pattern conceptually replicates prior work that used a smaller age range with smaller samples, often using general-knowledge tasks focusing on childhood or adulthood (e.g., Bayen et al., 2006; Bernstein, Erdfelder et. al., 2011; Ghrear et al., 2020; 2021; Groß & Pachur, 2019; Pohl et al., 2010; 2018). False-belief reasoning errors were significant in all age groups, but did not differ with age. The statistically stable lifespan false-belief function that we observed differs from other work showing an increase in false-belief reasoning from 3 to 7 years of age, followed by a decline in false-belief reasoning from younger to older adulthood (Bernstein, Sommerville, & Thornton, 2011; Bernstein et al., 2017; Mahy et al., 2017; see also Wacker et al., 2017). Overall, the current work addresses shortcomings of prior work by sampling across the entire child-to-adult lifespan to yield data patterns for hindsight bias and false-belief reasoning, while assessing verbal ability and executive function. A final point about our first hypothesis is that young children showed more variability than older children and adults on both the visual hindsight task and the Sandbox task (see Figure 1-4). This suggests that processes besides those directly involved in hindsight bias and false-belief reasoning affect young children's responses.

Regarding our second hypothesis, the null correlation between hindsight bias and false-belief reasoning that we observed indicates that these two constructs are unrelated. Researchers have noted why we might expect hindsight bias to correlate with false-belief reasoning (e.g., Taylor, 1988; Birch & Bernstein, 2007; Mitchell et al., 1996; Royzman et al., 2003). Few studies have demonstrated this link directly, though. In one study, preschoolers completed a battery of false-belief tasks and hindsight tasks, including the visual hindsight task used here. In that study, hindsight bias correlated positively with false-belief reasoning, even after controlling for age, verbal ability, and executive function (Bernstein et al., 2007). In another study, 3- to 13-year-olds completed computer versions of the visual hindsight task and Sandbox task used here. No correlation emerged between hindsight bias and false-belief reasoning (Begeer et al., 2016). Combining the current data with the mixed and limited correlational data on hindsight bias and false-belief reasoning, our best estimate at this point is that hindsight bias is unrelated to false-belief reasoning.

Implications of Our Results

Work on false-belief reasoning specifically and theory of mind generally continues in many directions, with separate focuses on different developmental periods in humans and non-humans (e.g., Krupenye & Call, 2019; Marchetti et al., 2018; see also Ishiguro, 2021 for robots developing theory of mind). Our results on the Sandbox task reveal surprisingly stable false-belief reasoning from preschool to old age. This result challenges most assumptions about false-belief reasoning across the lifespan. As noted previously, the literature shows a non-linear pattern of development featuring dramatic improvements in false-belief reasoning from young preschoolers to school-age children, followed by gradual improvements in adolescence, and declines from younger to older adulthood. This developmental picture differs markedly from the developmental pattern that we observed here (see Figure 4).

Our findings have implications for social and interpersonal relations and decision making across the child-to-adult lifespan. Some researchers distinguish between having and using a tool, such as theory of mind (Keysar et al., 2003). These authors argue that while adults have a mature theory of mind, they do not always use this ability effectively to interpret others' actions. Every age group in the current study showed robust hindsight bias and false-belief reasoning errors (egocentric bias). This result indicates that preschoolers through older adults have trouble ignoring privileged knowledge to reason from a naïve perspective. This pervasive trouble can adversely affect social relations and everyday decisions when we fail to appreciate that others do not share our knowledge, customs, or beliefs (e.g., Arkes, 2013; Henry et al., 2013; Mulvey et al., 2016). There is a rich literature on heuristics and biases (Tversky & Kahneman, 1974); some of that work involves the lifespan (e.g., Strough et al., 2011; Weller et al., 2011). Researchers often include hindsight bias in lists of everyday cognitive biases that affect decision making; however, researchers tend not to include egocentric bias in that list. Lifespan studies of egocentric bias could add to our understanding of how everyday heuristics and biases affect judgment and decision making in childhood and adulthood.

Limitations and future directions

Ultimately, one’s conclusions are only as good as one’s measures. Compared to the vast literatures on hindsight bias and false-belief reasoning, relatively few studies have investigated the psychometric properties of hindsight bias and false-belief reasoning tasks. There are some notable exceptions, mostly in theory of mind (e.g., Blank et al., 2008; Bosco et al., 2016; Chen et al., 2017; Gallant et al., 2020; Haywood & Homer, 2017; Osterhaus et al., 2016; Warnell & Redcay, 2019; Wellman & Liu, 2004). Some of this work has demonstrated adequate reliability and validity of various measures, while other work has shown the opposite.

We only measured one aspect of reliability in the current work, and that was internal consistency. The visual hindsight task had good internal consistency, while the Sandbox task did not. We did not measure the validity of either task. Some prior work has demonstrated the construct validity of the visual hindsight task by showing that it correlates with more standard verbal hindsight tasks such as knowledge of trivia (Bernstein, Erdfelder et al., 2011). Prior work has also demonstrated the construct validity of the Sandbox task by showing that it correlates with more standard false-belief tasks in children such as the change-of-location task (Sommerville et al., 2013; Mahy et al., 2017; cf. Samuel et al., 2018a; 2018b).

As we noted in the Introduction, it is better to use the same task rather than different tasks to measure a given construct (e.g., hindsight bias) across age groups. However, even with this approach, it is likely that age differences on such a task reflect myriad differences in social settings and dynamics. These differences include experimenter-participant rapport, experimenter credibility, participant engagement and motivation. In the current work, we used only one task to measure hindsight bias, and one task to measure false-belief reasoning.

Some researchers have questioned whether hindsight bias or theory of mind are unitary constructs (e.g., Blank et al., 2008; Heyes, 2014; Quesque & Rosetti, 2020; Schaafsma et al., 2015; Warnell & Redcay, 2019). For example, researchers have posited three hindsight bias components: memory distortions, foreseeability, and inevitability (Blank et al., 2008; see also Roese & Vohs, 2012). How we tested hindsight bias aligns best with the foreseeability component; however, other work involving visual and auditory hindsight bias uses a memory design in which people try to recall the point at which they first identified a clarifying stimulus (Harley et al., 2004; Higham et al., 2017). Using such a memory design aligns best with the memory distortions component of hindsight bias. Thus, the current hindsight bias results may not generalize to other hindsight tasks; also, each hindsight component may follow a different developmental pattern. The same logic applies to our false-belief reasoning results. Future work should include multiple measures of hindsight bias and false-belief reasoning, while paying close attention to their psychometric properties (cf., Borsboom, 2006).

Some researchers have wondered what the Sandbox task measures (Samuel et al., 2018a; 2018b). In one study, adult participants showed more bias in the false-belief than the memory-control condition of the Sandbox task, the same data pattern that we observed here (Samuel et al., 2018a). However, in a follow-up experiment, the researchers found no difference between the false-belief and a false-photo condition. The authors concluded that adult egocentric bias in the Sandbox task indicates a general difficulty reasoning about false representations rather than a specific difficulty reasoning about false beliefs. Notably, the authors used an attention-check condition developed by Sommerville and colleagues (2013) to control against participants developing a strategy of always choosing location 1, the correct response to memory-control and false-belief questions. In the attention-check condition, the protagonist places an object in one location and then watches another character move that object to a second location. Thus, the protagonist has a true rather than a false belief about the object's current location. Unlike

Sommerville and colleagues who did not analyze data in the attention-check condition, Samuel and colleagues used this condition to exclude participants who failed to indicate a response location closer to the second location than the first location.

In a follow-up study, Samuel and colleagues (2018b) again excluded participants who failed this attention-check question in the Sandbox task. In this latter study, the authors found no difference in adults between the false-belief and the memory-control conditions. Together, this pair of studies question the Sandbox task's overall utility. Unlike these prior studies, we did not use an attention-check condition in the current study. In essence, this attention-check question controls against participant strategies that would produce a null effect between the memory-control and false-belief condition. However, the fact that we observed a significant difference between these conditions in all age groups in the current study suggests that we did not need this attention-check condition. Future work should continue to explore what exactly the Sandbox task measures.

A possible limitation of the current work is the use of forward digit span as a measure of working memory. Some have argued that backward digit span is a better measure of working memory than is forward digit span (Reynolds, 1997). We pilot tested both versions; however, we found that our youngest children had trouble following the backward digit span instructions. We, thus, opted to use forward digit span, which children had no trouble following.

A consideration in the current work is the impact of response bias on our findings. Our two primary tasks permitted us to explore hindsight bias and egocentric bias. However, these biases did not permit us to explore more basic response biases that arise in other tasks. Some research seeks to determine whether response bias is a cognitive trait that can be used to predict how people will respond in various tasks, including recognition memory, eyewitness identification, and false memory (Kantner & Lindsay, 2012; 2014). Future work should explore whether certain age groups are more prone to respond in a biased manner across different tasks.

Finally, the current work employed a cross-sectional design, which should not be used to study developmental change. Thus, any age effects observed here could be due to cohort effects (e.g., historical context of one's youth) rather than true age effects. Longitudinal designs, though costly and hard to employ, are needed to identify developmental change. This is not feasible when tracking development from preschool to old age; however, there are different experimental designs that can be used to study developmental change over large age ranges, including growth models and sequential cohort models (e.g., Milojev & Sibley, 2017). Future work should use such modeling approaches to track development in hindsight bias and false-belief reasoning across the lifespan. Future work should also use formal modeling approaches, such as multinomial processing tree models, to assess hindsight bias and false-belief reasoning across the lifespan (e.g., Bernstein, Erdfelder et al., 2011; Erdfelder et al., 2009).

Conclusion

This study showed that hindsight bias and false-belief reasoning follow different developmental patterns from preschool to old age (see Figures 1-4). All ages exhibited both hindsight bias and false-belief reasoning errors, the latter in the form of egocentric bias. Despite robust errors in all ages, hindsight bias was U-shaped and false-belief reasoning was age invariant. The fact that hindsight bias and false-belief reasoning show different developmental patterns and do not correlate leads us to conclude that these are distinct constructs.

References

Arkes, H.R. (2013). The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science*, 22, 356-360.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611-626.

Baron-Cohen, S, Leslie, A.M., & Frith, U, (1985). Does the autistic child have a "theory of mind?" *Cognition*, *21*, 37-46.

Bayen, U. J., Erdfelder, E., Bearden, J. N., & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1003–1018. doi:10.1037/0278-7393.32.5.1003

Bayen, U.J., Pohl, R.F., Erdfelder, E., & Auer, T.S. (2007). Hindsight bias across the lifespan. *Social Cognition*, *25*, 83-97.

Begeer, S., Bernstein, D.M., Aßfalg, A., Azdad, H., Glasbergen, T., Wierda, M, & Koot, H.M. (2016). Equal egocentric bias in school-age children with and without Autism Spectrum Disorder. *Journal of Experimental Child Psychology*, *144*, 15-26. doi:10.1016/j.jecp.2015.10.018

Bernstein, D.M. (2015-2020). Lifespan hindsight bias and false belief [unpublished raw data]. https://osf.io/d9ae5/?view_only=c483e01336ef4c15a6f86f175603c53e

Bernstein, D.M. (2018). Lifespan theory of mind: A call for broader perspectives and more integration. *Zeitschrift für Psychologie, 226*, 85-86. https://doi.org/10.1027/2151-2604/a000324

Bernstein, D.M., Atance, C., Loftus, G.R., & Meltzoff, A.N. (2004). We saw it all along: Visual hindsight bias in children and adults. *Psychological Science*, *15*, 264-267.

Bernstein, D.M., Atance, C., Meltzoff, A.N., & Loftus, G.R. (2007). Hindsight bias and developing theories of mind. *Child Development*, *78*, 1374-1394.

Bernstein, D.M., Coolin, A., Fischer, A., Thornton, W.L., & Sommerville, J.A. (2017). False-belief reasoning from 3 to 92 years of age. *PLoS ONE,* 12: e0185345. https://doi.org/10.1371/journal.pone.0185345

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, & Cognition, 37*, 378-391.

Bernstein, D.M., Thornton, W.L., & Sommerville, J.A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false-belief task. *Experimental Aging Research*, 37, 481-502.

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*, 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*, 283-286.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18,* 382-386.

Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*, 1408-1440.

Bosco, F. M., Gabbatore, I., Tirassa, M., & Testa, S. (2016). Psychometric properties of the Theory of Mind Assessment Scale in a sample of adolescents and adults. *Frontiers in Psychology,* 7, 566. doi: 10.3389/fpsyg.2016.00566

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika 71*, 425-440. https://doi.org/10.1007/s11336-006-1447-6

Brown, N. R., Lee. P. J., Krslak, M., Conrad, F. G., Hansen, T., Havelka, J. & Reddon, J. (2009). Living in history: How war, terrorism, and natural disaster affect the organization of autobiographical memory. *Psychological Science, 20*, 399-405.

Carpendale, J.I., & Chandler, M.J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development*, 67, 1686-1706.

Chen, K.W., Lee, S.C., Chiang, H.Y., Syu, Y.C., Yu, X.X., & Hsieh, C.L. (2017). Psychometric properties of three measures assessing advanced theory of mind: Evidence from people with schizophrenia. *Psychiatry Research*, *257*, 490–496. https://doi.org/10.1016/j.psychres.2017.08.026

Decety, J., & Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3, 71-100.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2014.00781

Erdfelder, E., Auer, T.S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie, 217*, 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental status". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189-198. Doi: 10.1016/0022-3956(75)90026-6.

Gallant, C.M.M, Lavis, L., & Mahy, C.E.V. (2020). Developing an understanding of others' emotional states: Relations among affective theory of mind and empathy measures in early childhood. *British Journal of Developmental Psychology*, *38*, 151-166. https://doi.org/10.1111/bjdp.12322

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3-7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129-153.

Ghrear S, Baimel A, Haddock T, & Birch S.A.J. (2021). Are the classic false belief tasks cursed? Young children are just as likely as older children to pass a false belief task when they are not required to overcome the curse of knowledge. *PLoS ONE, 16,* e0244141. https://doi.org/10.1371/journal.pone.0244141

Ghrear, S., Fung, K., Haddock, T. & Birch, S.A. (2020). Only familiar information is a "curse": Children's ability to predict what their peers know. *Child Development*. doi:10.1111/cdev.13437

Giroux, M. E., Coburn, P. I., Harley, E. M., Connolly, D. A. & Bernstein, D. M. (2016). Hindsight bias and law. *Zeitschrift für Psychologie*, *224*, 190–203. doi: 10.1027/2151-2604/a000253

Groß, J., & Pachur, T. (2019). Age differences in hindsight bias: A meta-analysis. *Psychology and Aging, 34*, 294–310. https://doi.org/10.1037/pag0000329

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311-327.

Hayward, E.O., & Homer, B.D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, *35*, 454-462.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging, 28*, 826-839. https://doi.org/10.1037/a0030677

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science,* 9, 131–143.

Higham, P.A., Neil, G.J. & Bernstein, D.M. (2017) Auditory hindsight bias: Fluency misattribution versus memory reconstruction. *Journal of Experimental Psychology: Human Perception & Performance, 43,* 1144–1159. doi:10.1037/xhp0000405.

Ishiguro, K. (2021). *Klara and the Sun*. New York: Alfred A. Knopf

JASP Team (2020). (Version 0.13.1) [Computer software]. https://jasp-stats.org/

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

Kantner, J., & Lindsay, D.S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40*, 1163-1177. https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D.S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review, 21*, 1272-1280. doi: 10.3758/s13423-014-0608-3.

Kaufman, A.S., & Kaufman, N.L. (2005). *Kaufman Tests of Educational Achievement, Second Edition, Brief Form*. Bloomington, MN: Pearson, Inc.

Krupenye, C., & Call, J. (2019). Theory of mind in animals: Current and future directions. *WIREs Cognitive Science,* https://doi.org/10.1002/wcs.1503

Keysar, B., Lin, S., & Barr, D.J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25-41. DOI: 10.1016/S0010-0277(03)00064-7

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*, 659-674. DOI: 10.1111/cdev.12154.

Mahy, C.E.V., Bernstein, D.M., Gerrard, L.D., & Atance, C.M (2017). Testing the validity of a continuous false belief task in three- to seven-year old children. *Journal of Experimental Child Psychology*, *160*, 50-66. Doi: 10.1016/j.jecp.2017.03.010

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie, 226*, 98–109. doi: 10.1027/2151-2604/a000326

Massaro, D., Castelli, I., Sanvito, L, & Marchetti, A. (2014). The 'I knew it all along' phenomenon: second-order false belief understanding and the curse of knowledge in primary school children. *European Journal of Psychology of Education, 29*, 311-326. https://doi.org/10.1007/s10212-013-0200-7

Miller, S.A. (2012). *Theory of mind: Beyond the preschool years*. New York, NY: Psychology Press.

Milojev, P., & Sibley, C. G. (2017). Normative personality trait development in adulthood: A 6-year cohort-sequential growth model. *Journal of Personality and Social Psychology, 112*, 510–526. https://doi.org/10.1037/pspp0000121

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.

Mulvey, K.L., Rizzo, M.T. & Killen, M. (2016), Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science, 19*, 999-1010. https://doi.org/10.1111/desc.12345

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–259. doi: 10.1126/science.1107621

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development*, *87*, 1971-1991.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences*, *16*, 519-525.

Pinker, S. (2014, September 26). The source of bad writing: The 'curse of knowledge' leads writers to assume their readers know everything they know. *Wall Street Journal.* Retrieved from http://online.wsj.com/articles/the-cause-of-bad-writing-1411660188.

Pohl, R.F., Bayen, U.J., Arnold, N., Auer, T-S., & Martin, C. (2018) Age differences in processes underlying hindsight bias: A life-span study. *Journal of Cognition and Development*, *19*, 278-300, DOI: 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*, 1268–1282. https://doi.org/10.1037/a0020209

Pohl, R. F., & Erdfelder, E. (2017). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (p. 424–445). Routledge/Taylor & Francis Group.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*, 384–396. https://doi.org/10.1177/1745691619896607

Repacholi, B.M., & Slaughter, V. (Eds.). (2003). *Individual differences in theory of mind: Implications for typical and atypical development*. Hove, UK: Psychology Press.

Reynolds, C.R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12*, 29-40. https://doi.org/10.1016/S0887-6177(96)00015-7

Roese, N.J., & Vohs, K.D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*, 411-426.

Royzman, E.B., Cassidy, K.W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, *7*, 38-65.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266. https://doi.org/10.1037/a0018729

Samuel, S., Legg, E.W., Lurz, R., & Clayton, N.S. (2018a). Egocentric bias across mental and non-mental representations in the Sandbox Task. *Quarterly Journal of Experimental Psychology, 71*, 2395-2410. doi.org/10.1177/1747021817742367

Samuel S., Legg E.W., Lurz R., & Clayton N.S. (2018b). The unreliability of egocentric bias across self–other and memory–belief distinctions in the Sandbox Task. *Royal Society Open Science. 5*: 181355. http://dx.doi.org/10.1098/rsos.181355

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Science, 19*, 65-72.

Siegler, R.S. (2005). Children's learning. *American Psychologist*, *60*, 769-778.

Simonsohn, U. (2018). Two-lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science, 1,* 538-555. doi:10.1177/2515245918805755

Sommerville, J.A., Bernstein, D.M., & Meltzoff, A.N. (2013). Measuring false belief in centimeters: Adults and children fail to suppress privileged knowledge on a novel change-of-location task. *Child Development*, *84*, 1846-1854.

Stroop, J.R. (1935). Studies of interference in verbal serial reactions. *Journal of Experimental Psychology, 18*, 643-662.

Strough, J., Karns, T.E., & Schlosnagle, L. (2011). Decision-making heuristics and biases across the life span. *Annals of the New York Academy of Sciences*, 1235, 57-74.

Taylor, M. (1988). Conceptual perspective taking: Children's ability to distinguish what they know from what they see. *Child Development, 59*, 703-718.

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

Wacker, R., Bölte, S., & Dziobek, I. (2017). Women know better what other women think and feel: Gender effects on mindreading across the adult life span. *Frontiers in Psychology*, 8, 1324. doi: 10.3389/fpsyg.2017.01324.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*, 35–57. https://doi.org/10.3758/s13423-017-1343-3

Warnell, K.R., & Redcay, E (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*: 103997 10.1016/j.cognition.2019.06.009

Weller, J.A., Levin, I.P., & Denberg, N.L. (2011). Trajectory of risky decision making for potential gains and losses from age 5 to 85. *Journal of Behavioral Decision Making*, *24*, 331-344.

Wellman, H.M. (1990). *The child's theory of mind*. Cambridge, MA., MIT Press.

Wellman, H.M., Cross, D, & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655-684.

Wellman, H.M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wechsler, D. (1958). *The measurement of adult intelligence (4th Ed.).* Baltimore: Williams and Wilkins.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 345-353.

Endnotes

1. Given the surprising nature of the small effect that age had on egocentric bias, we wondered whether another variable might have masked a larger age effect. We, therefore, considered whether task order affected egocentric bias in our sample. Recall that we administered two different orders of the Sandbox task. A univariate ANOVA on egocentric bias, testing the effects of age group and task order, revealed an interaction (see Supplementary Materials). The interaction resulted from the two youngest age groups (3- to 5-year-olds, 6- to 9-year-olds) showing significantly more egocentric bias when they completed the false-belief trials after completing the memory-control trials. Adults (18- to 64-year-olds, 65+ year-olds) showed the reverse pattern. Because task order and its interaction with age group were not part of our initial research question, we do not consider these variables further.
2. Partial correlation, controlling for Sandbox task order (r = .110, n.s.).
3. A reviewer wondered how the removal of completely wrong answers on both false-belief trials would affect our data pattern. Only 2 / 683 participants chose exactly L2 on both false-belief trials. Removing these participants did not affect the data pattern. Given how rare it was for participants to choose exactly L2 on both false-belief trials, we next relaxed our criterion for capturing completely wrong answers. We did this by removing participants whose responses fell within 10% of the boundary on each side of L2. Recall that L2 was 14 inches away from L1. This 10% boundary corresponded to scores between 12.6 and 15.4 inches. In all, 12 / 683 participants met this, more liberal, criterion. Again, removing these participants did not affect the data pattern.
4. Partial correlation, controlling for Sandbox task order (r = -.002).

Table 1

Correlations between variables of interest.

	KBIT-2	Digit Span	Day / Night	Stroop Error Difference	Stroop RT Difference	Hindsight Bias	Egocentric Bias
Age	.732**	.401**	.523**	.148**	.053	-.022	-.093*
KBIT-2		.637**	.531**	.073	.043	-.224**	-.102**
Digit Span			.463**	.013	.041	-.193**	-.044
Day / Night						.025	.046
Stroop Error Difference					.465**	-.014	-.008
Stroop RT Difference						-.042	.018
Hindsight Bias							-.004
Egocentric Bias							

Note. KBIT-2 = Kaufman Brief Intelligence Test; Day / Night task administered only to children 3 – 5 years of age; Stroop task administered to participants older than age 5.

*p < .05; **p < .01

Table 2

Baseline and Hindsight mean (and standard error of the mean) number of filter screens in place for each age group in the Visual Hindsight task.

Age Group	Baseline	Hindsight
3-5 yrs [*n* = 89]	7.639 (0.151)	5.371 (0.182)
6-9 yrs [*n* = 111]	6.836 (0.135)	6.051 (0.163)
10-17 yrs [*n* = 137]	6.700 (0.122)	6.179 (0.147)
18-64 yrs [*n* = 239]	6.280 (0.092)	5.492 (0.111)
65+ yrs [*n* = 102]	7.206 (0.141)	5.966 (0.170)

Table 3

Bayesian regression model comparison involving mean-centered age, mean-centered age squared (Age^2), and mean-centered age cubed (Age^3) as predictors of egocentric bias.

Models	BF_{10}
Null model	1.000
$Age^2 + Age^3$	6.89e10
$Age + Age^2 + Age^3$	1.21e10
$Age + Age^2$	6.63e4
Age^2	6.14e1
Age	0.10
Age^3	0.10
$Age + Age^3$	0.01

Note. All models are compared to the Null model. BF_{10} > 100 denotes extreme evidence for the alternative hypothesis (H1); BF_{10} = 30 - 100 denotes very strong evidence for H1; BF_{10} = 10 - 30 denotes strong evidence for H1; BF_{10} = 3 - 10 denotes moderate evidence for H1; BF_{10} = 1- 3 denotes anecdotal evidence for H1; BF_{10} = 1 denotes equal evidence for H1 and the null model (H0); BF_{10} = 1/3 – 1 denotes anecdotal evidence for H0; BF_{10} = 1/30 – 1/10 denotes strong evidence for H0; BF_{10} = 1/100 – 1/30 denotes very strong evidence for H0; BF_{10} < 1/00 denotes extreme evidence for H0 (Wagenmakers et al., 2018).

Table 4

Mean memory-control and false-belief (and standard error of the mean) values for each age group in the Sandbox task.

Age Group	Memory Control	False Belief
3-5 yrs [*n* = 94]	7.864 (0.576)	11.731 (0.662)
6-9 yrs [*n* = 114]	3.419 (0.523)	5.888 (0.601)
10-17 yrs [*n* = 138]	1.126 (0.475)	3.4764 (0.547)
18-64 yrs [*n* = 235]	1.531 (0.364)	3.010 (0.419)
65+ yrs [*n* = 102]	1.723 (0.553)	2.922 (0.636)

Table 5

Bayesian regression model comparison involving mean-centered age, mean-centered age squared (Age^2), and mean-centered age cubed (Age^3) as predictors of egocentric bias.

Models	BF_{10}
Null model	1.000
Age	1.491
$Age^2 + Age^3$	0.690
Age^3	0.639
$Age + Age^2 + Age^3$	0.135
$Age + Age^2$	0.296
$Age + Age^3$	0.192
Age^2	0.137

Note. All models are compared to the Null model. $BF_{10} > 100$ denotes extreme evidence for the alternative hypothesis (H1); $BF_{10} = 30 - 100$ denotes very strong evidence for H1; $BF_{10} = 10 - 30$ denotes strong evidence for H1; $BF_{10} = 3 - 10$ denotes moderate evidence for H1; $BF_{10} = 1- 3$ denotes anecdotal evidence for H1; $BF_{10} = 1$ denotes equal evidence for H1 and the null model (H0); $BF_{10} = 1/3 – 1$ denotes anecdotal evidence for H0; $BF_{10} = 1/30 – 1/10$ denotes strong evidence for H0; $BF_{10} = 1/100 – 1/30$ denotes very strong evidence for H0; $BF_{10} < 1/00$ denotes extreme evidence for H0 (Wagenmakers et al., 2018).

Figure 1

Average hindsight bias as a function of age group on the visual hindsight task. Scores above zero indicate hindsight bias. Error bars are 95% confidence intervals.

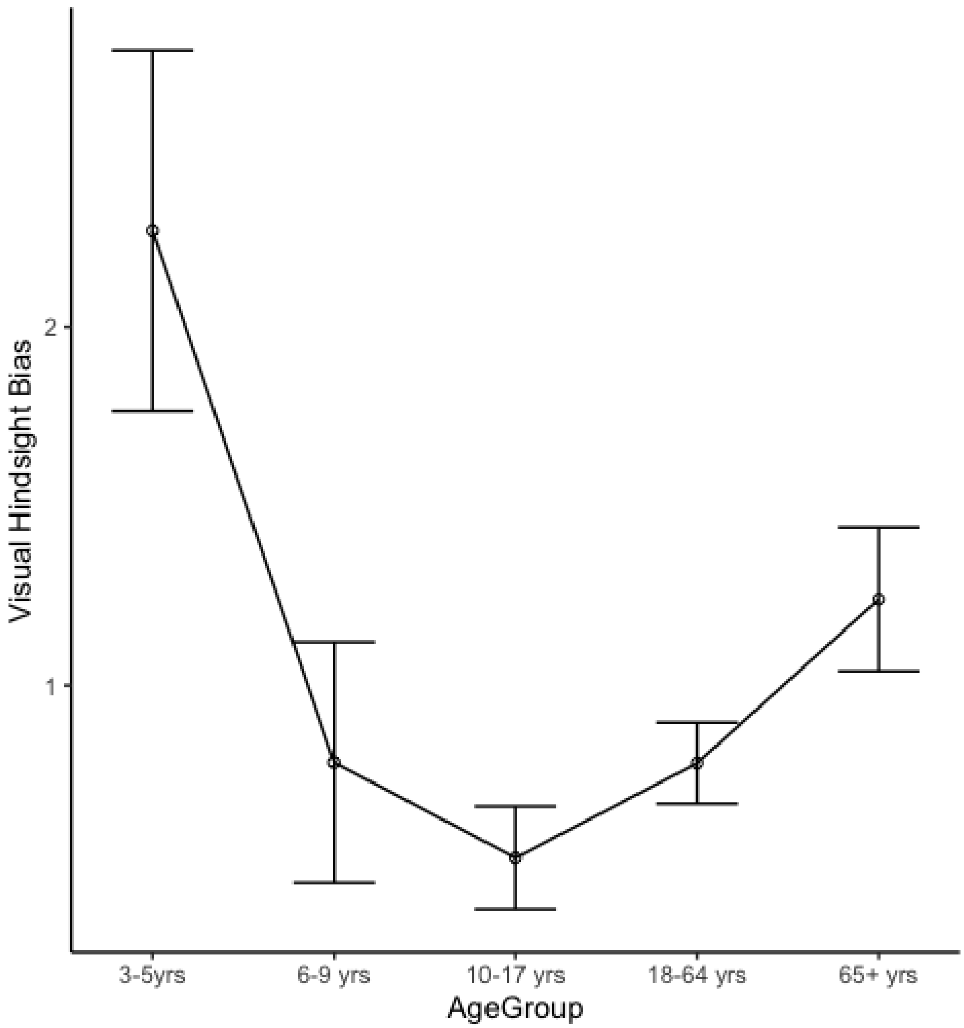


Figure 2

Average hindsight bias as a function of age in years on the visual hindsight task. Scores above zero indicate hindsight bias. Grey band represents 95% confidence interval on the fit line (the estimated mean at each age) based on the linear model.

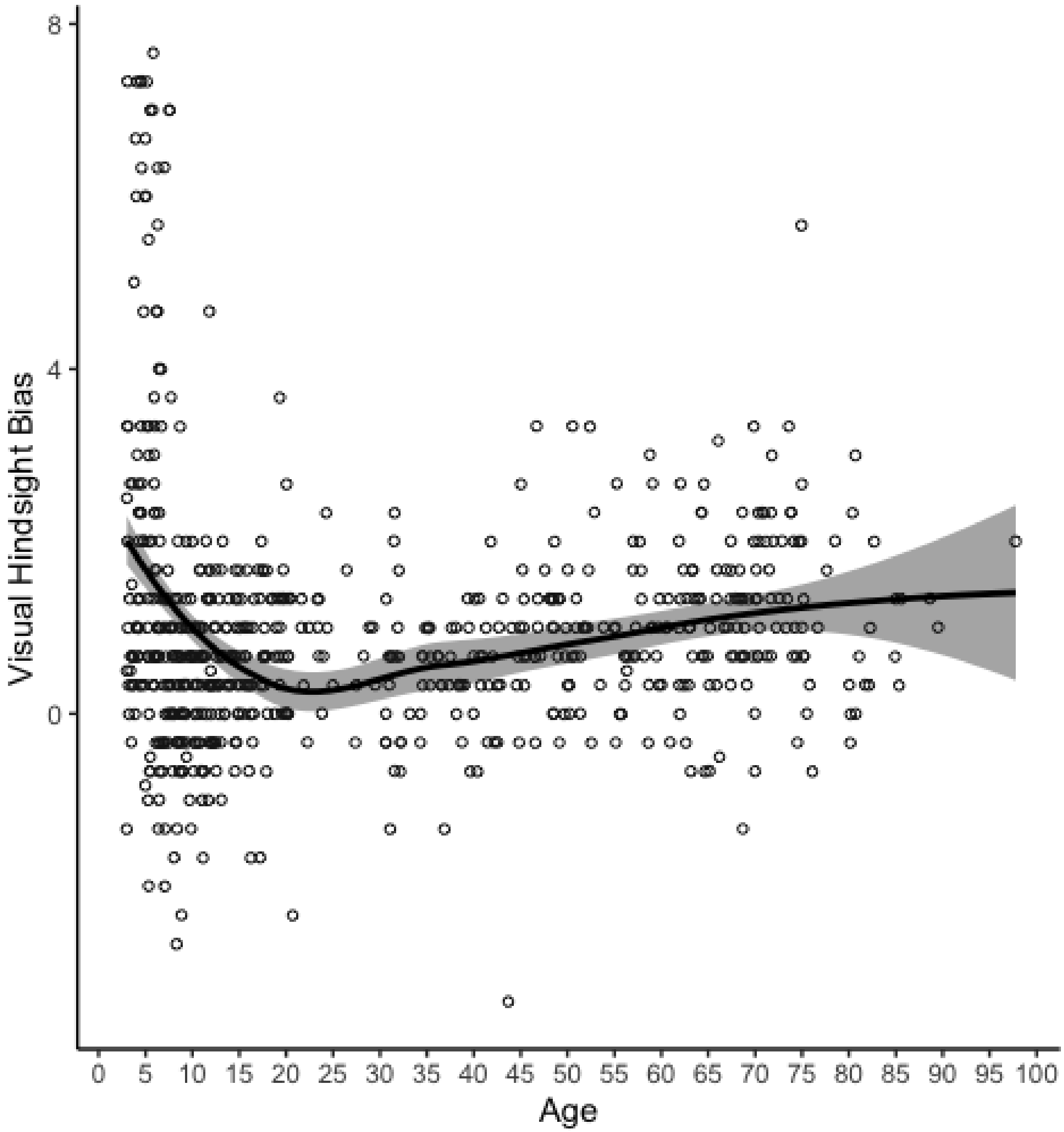


Figure 3

Average egocentric bias as a function of age group on the Sandbox task. Egocentric bias is an inverse measure of false-belief reasoning where scores above zero indicate trouble ignoring privileged knowledge when reasoning about another person's false belief. Error bars are 95% confidence intervals.

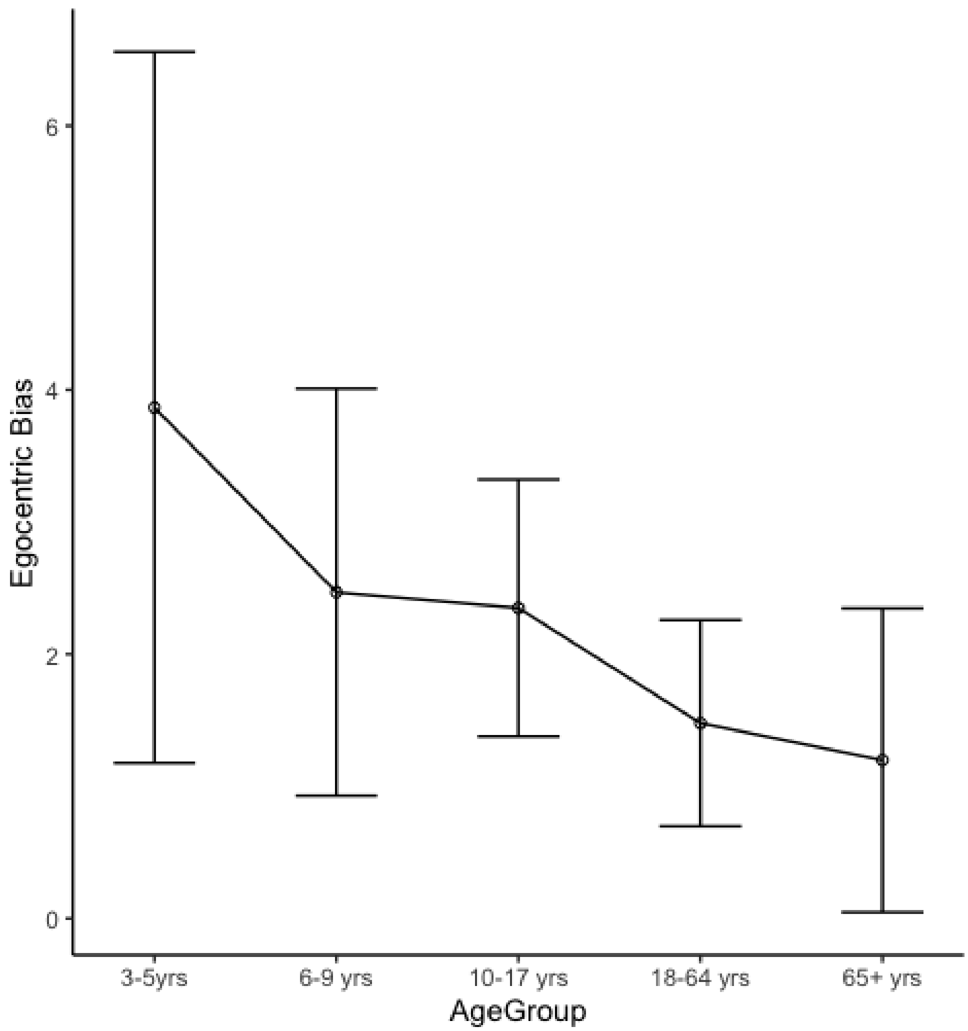
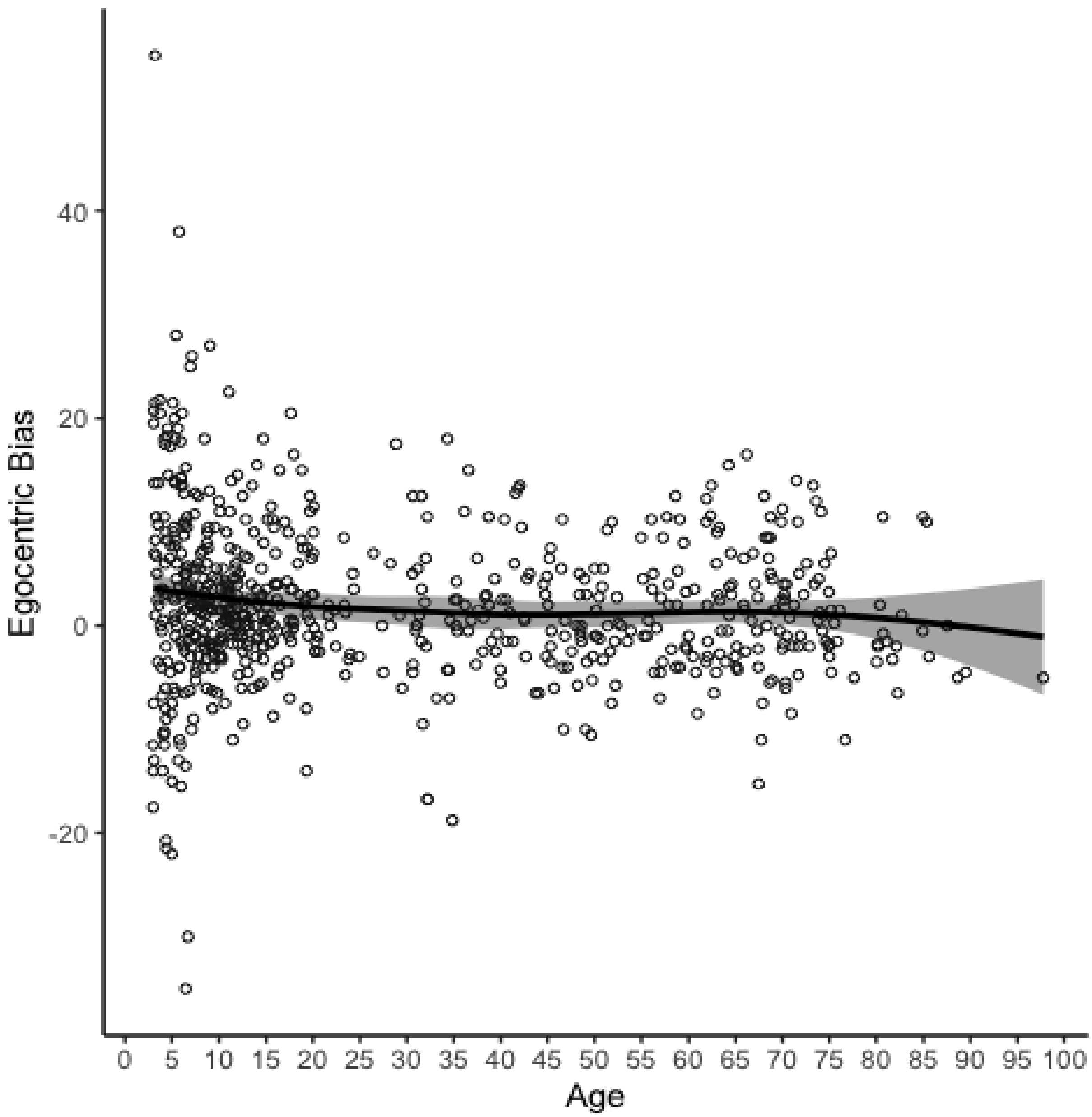


Figure 4

Average egocentric bias as a function of age in years on the Sandbox task. Egocentric bias is a measure of false-belief reasoning errors, where scores above zero indicate trouble ignoring privileged knowledge when reasoning about another person's false belief. Grey band represents 95% confidence interval on the fit line (the estimated mean at each age) based on the linear model.



Appendix A

Memory-control and false-belief bias on each trial, collapsed across age, on the Sandbox task. Scores of 0 indicate perfect performance. Positive scores indicate a bias toward Location 2, which appeared 14 inches away from Location 1.