# Omniscience errors in mental state reasoning

**Branden J. Bio* and Sangeet Khemlani**
**{branden.bio.ctr, sangeet.khemlani}@nrl.navy.mil**
*NRC Postdoctoral Fellow
Navy Center for Applied Research in Artificial Intelligence
U.S. Naval Research Laboratory, Washington, DC 20375 USA

## Abstract

Young children make systematic mistakes when reasoning about what other agents know and believe -- mature mental state reasoning emerges around late childhood. We describe a novel class of errors that adult reasoners make when considering information about the mental states of others. Participants in two studies reasoned about common conditional reasoning inferences couched in terms of an agent's knowledge or belief, e.g., *Alia knows that if it's rainy then the café is closed; It's rainy. What follows?* They generated their responses using a novel sentence construction interface. Many participants spontaneously generated responses such as, *Alia knows that the café is closed.* This pattern reflects an "omniscience" error, i.e., one in which reasoners erroneously impute knowledge of a deductive consequence to an agent. We discuss the results in the context of recent proposals on epistemic inference.

**Keywords:** omniscince errors; mental states; epistemic reasoning; knowledge; belief

## Introduction

People without formal training in logic can reason about the mental states of others. You do so when planning a surprise party for a friend, for instance: you inform guests not to tell your friend about the party, and you make sure to maintain the secret as well, allowing your friend to remain unaware. In essence, you track the mental states of yourself, your friend, and the party guests.

The ability to reason about mental states develops through childhood: children have difficulty keeping track of mental states, and so they make systematic errors regarding them (Dalke, 1995; Flavell, 1999). One prominent error children make is that they have difficulty keeping track of other people's false beliefs -- they mistakenly think that other people have access to their knowledge, as revealed by the "Sally-Anne" task (Baron-Cohen, Leslie, & Frith, 1985). Along with difficulty understanding others' false beliefs, children also have trouble representing the perspectives of others (Surtees, Butterfill, & Apperly, 2012).

Deficits in mental state reasoning by children may come about as a consequence of limitations in lower level processes such as executive functioning (Kouklari, et al., 2017) and inhibitory control (Austin, Groppe, & Elsner, 2014; Sabbagh etal., 2006). Process limitations can explain why children, but not adults, make such reasoning errors.

Adults make different sorts of errors when reasoning about mental states. For instance, they often overestimate the ubiquity and transparency of their own beliefs relative to those of others. Birch and Bloom (2004; 2007) likewise document a "curse of knowledge" bias in which a person's knowledge of the consequence of some event compromises their ability to reason about other people's beliefs (see also Diamond & Kirkham, 2005; Royzmann, Cassidy, & Baron, 2003).

As yet, there exists no comprehensive theory that explains the processes and representations that underlie mature mental state reasoning. Investigations into adult reasoning about mental states show the extent to which reasoners compartmentalize their beliefs from others (Apperly, Samson, & Humphrey, 2009; Keysar, Lin, & Barr, 2003), and the neural substrates where such compartmentalization can occur (Bio, Guterstam, Pinsk, Wilterson, & Graziano, 2021). Behavioral and neurocognitive investigations suggest that the brain recruits machinery for representing one's own mental states to represent those of others (e.g., Bio, Webb, & Graziano, 2019; Kovács, Téglás, & Endress, 2010), and that people can hold others' mental states in mind as alternative possible configurations of the world.

Indeed, general accounts of human reasoning argue that people draw conclusions by considering possible states of the world (see, e.g., Carey et al., 2020; Johnson-Laird & Ragni, 2019; Phillips et al., 2019). The maintenance of multiple possibilities in memory can be difficult, and so reasoners may consider one possibility at a time or else coalesce, simplify, and reduce possibilities in other ways that facilitate inference (Johnson-Laird, 1983). These shortcuts reduce the amount of processing reasoners have to carry out, but they can yield systematic errors (see, e.g., Khemlani & Johnson-Laird, 2017). If mental state inference, like other sorts of inference, is based on constructing possibilities (cf. Jara-Ettinger & Rubio-Fernandez, 2021), then significant cognitive load should disrupt adults' abilities for tracking belief states (Schneider, Lam, Bayliss, & Dux, 2012).

The preceding cases illustrate ways in which belief states are difficult to encode and maintain. But, it remains unknown whether individuals make systemic errors specific to *reasoning* processes. That is, do people systematically make logically suboptimal mental state inferences? No studies have investigated the issue, so this paper aimed to address the discrepancy. It investigated the presence of *omniscience errors* in reasoning, such as the pattern of reasoning embodied in (1):

1. Devon knows that if Olga is a client, she's also a student.
   Olga is a client.
   * *Therefore, Devon knows that Olga is a student.*

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Because Devon *knows* that all the clients are students, and because *knows* is a factive verb that presupposes the truth of its complement, it follows that Olga is indeed a student – but it doesn't follow that Devon is aware of this fact. The pattern reflects an error in reasoning: it's possible that Devon has no idea whether Olga is a client, or it's possible that Devon erroneously thinks Olga isn't a client. An analogous error applies to inferences about belief:

2. Luz believes that if Olga is a client, she's also a student.
   Olga is a client.
   * *Therefore, Luz believes that Olga is a student.*

Perhaps (2) is more egregious than (1), because *believe* is not factive, i.e., Luz may be mistaken that all the clients are students, or Luz may believe the claim as a conjecture only. Reasoners who make the errors in (1) and (2) may do so by attributing undue omniscience to Devon and Luz, so we describe these patterns of reasoning as omniscience errors.

In this paper, we report investigations whether people make such omniscience errors. The results illustrate how people compartmentalize (or fail to compartmentalize) the mental states of the agents they read about from their own deductive inferences. The paper begins by reviewing epistemic logics and explaining how they are built to preclude the kind of omniscience errors above. It also reviews preceding work on omniscient thinking in adult reasoners. We present a novel methodology for eliciting omniscience errors, and report two studies using the task to test whether reasoners untrained in logic commit them. We conclude by discussing the theoretical ramifications of these errors.

## Omniscience in logic and language

A consensus in contemporary cognitive science is that humans do not reason by recourse to any symbolic logic (Khemlani, 2018; Johnson-Laird, 2010; Elqayam & Over, 2013; Oaksford & Chater, 2007; cf. Bringsjord & Govindarajulu, 2020). But various systems of logic, including probability logics, continue to serve as benchmarks for accurate reasoning, both for the development of psychological theory (Pietarinen, 2003; Pfeifer & Kleiter, 2009) as well as in artificial intelligence (Sutcliffe, 2017). A prominent example is the usage of epistemic logics to model valid reasoning about mental states (Bolander, 2014; van de Pol, van Rooij, & Szymanik, 2018; van Ditchmarsch & Labuschagne, 2007).

Theorists developed epistemic logics to capture the modal properties of operators for knowledge and belief (Fagin, Halpern, Moses, & Vardi, 1995; von Wright, 1951; Hintikka, 1962). They argued that to express that an individual knows something – *A knows P,* or $\boldsymbol{K}_A(P)$ – is to express that *P* is true in one or more situations consistent with *A's* mental state. A countable infinity of epistemic logics exist: each separate logic denotes a distinct set of axioms that describe what can and cannot follow. Here are two axioms embodied in the most frequently used epistemic logics, along with their English translations:

$\boldsymbol{K}_A(P) \rightarrow P$ **Axiom T**
*(If* A knows P*, then* P *is the case,*
*i.e., knowledge is factual.)*

$\boldsymbol{K}_A(P \rightarrow Q) \rightarrow (\boldsymbol{K}_A(P) \rightarrow \boldsymbol{K}_A(Q))$ **Axiom K**
*(If* A knows that if P then Q*, then whenever*
A knows P*, then it follows that* A knows Q *too.)*

Axiom T expresses the notion that *A knows P* is true whenever *P* is true in both *A's* mental states as well as the world at large. The two axioms, and indeed, most axioms in epistemic logic, describe what can be derived from an agent's state of knowledge. Very few axioms concern what can be derived from a state of belief, and so as a consequence, to say that *A believes P* is to express that *P* is true in *A's* mental states but not necessarily the possible states of the world.

Critics of epistemic logic worry that it presents an implausible description of human reasoning. Axiom K above, after all, suggests that agents have immediate access to the logical consequences of their knowledge – a form of "logical omniscience" (see Stalnaker, 1991) – and early theorists such as Hintikka acknowledged this property as a discrepancy between logic and natural language (1962, p. 30-31).

Nevertheless, such logics can be useful in justifying certain commonsense intuitions. Consider again problem (1) above. In epistemic logic, we might express (1) as follows:

1'. $\boldsymbol{K}_{Devon}(\text{client}(Olga) \rightarrow \text{student}(Olga))$
$\text{client}(Olga)$

Intuitions suggest that it is a mistake to conclude that Devon knows whether or not Olga's a student, because there's no reason to believe that Devon knows she's a client. A reasoner who draws such a conclusion has made a gross error of omniscience – they presume that Devon has much more knowledge about the situation than the premises suggest. The intuition accords with all systems of epistemic logic, including the most permissive calculi, which treat this inference: $\boldsymbol{K}_{Devon}(\text{student}(Olga))$ as invalid.

Do humans make such errors of omniscience? It is challenging to investigate, because presenting reasoners with a prompt such as, "Does it follow that Devon knows that Olga is a student?" may unduly bias their responses. Reasoners who would otherwise hesitate to draw such a conclusion might do so if prompted. Hence, we developed an interface that permits investigators to study reasoners' epistemic conclusions. Studies using the interface reveal the presence of such errors of omniscience.

## A quasi-generative sentence construction interface for studying mental state reasoning

Experimentalists can explicitly probe people's inferences using generative and evaluative methodologies, e.g., they can ask individuals, "What, if anything follows?" or "What do

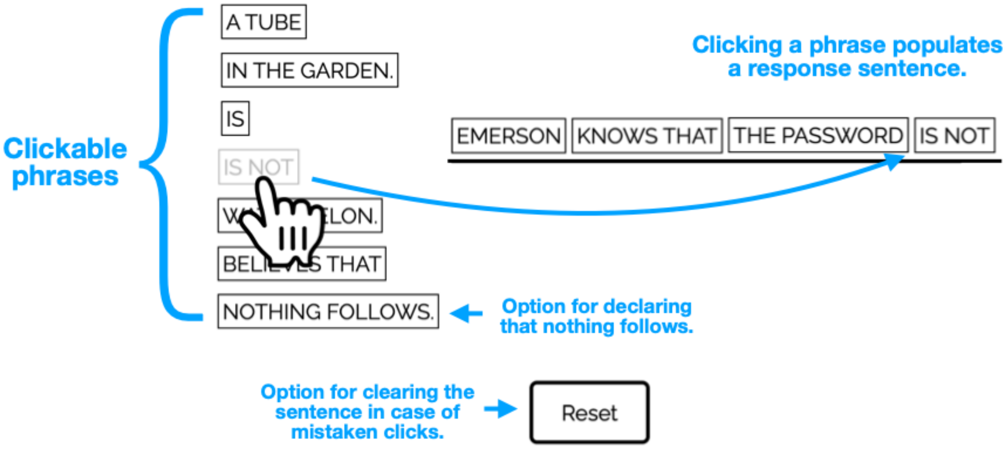
Clickable phrases
A TUBE
IN THE GARDEN.
IS
IS NOT
W LON.
BELI S THAT
NOTHING FOLLOWS.
Option for declaring that nothing follows.
Clicking a phrase populates a response sentence.
EMERSON
KNOWS THAT
THE PASSWORD
IS NOT
Option for clearing the sentence in case of mistaken clicks.
Reset

**Figure 1.** Sentence construction interface used in Experiments 1 and 2. Participants received a set of premises (see below) and responded to open-ended questions such as, "What, if anything, follows?" A set of buttons (left) corresponded to phrases that participants used to populate response sentences (right). One button in the interface permitted participants to declare that nothing followed, another permitted them to clear the response sentence and start the task over in case of mistakes.

you think happened?" or other such open-ended questions. They must then code participants' natural responses. The difficulty with adapting such tasks to study mental state reasoning is that participants may not spontaneously describe the inferences they make about mental states, and their inferences may be difficult to interpret. Evaluative, forced-choice approaches restrict the possibilities that participants can produce: if participants make errors when evaluating statements provided by experimenters, it is unclear whether they would ever make such errors in daily life.

We therefore developed a hybrid, quasi-generative methodology to study mental state reasoning and employed it in Experiments 1 and 2 below. Participants received open-ended prompts and interacted with an interface (see Figure 1) that permitted them to respond by clicking one or more buttons that corresponded to the various words needed to construct a full sentence. Each time they clicked a button, the corresponding word was added to the end of the sentence, and the button was removed from the options available. Participants could click a "reset" button in case they made an error. They could also click a button corresponding to "nothing follows". The approach allowed participants to consider all the various pieces of information in a given scenario, both relevant and irrelevant; and it minimized indirect effects of response options. The choice of words from which participants built their responses focused on inferences most relevant to the problems in Experiments 1 and 2, though it did not bias them to respond by inappropriately constructing sentences that described mental states. This method makes the probability of producing any coherent answer by chance incredibly small (<1%). The quasi-generative nature of the task permitted efficient coding of omniscience errors.

The instructions of Experiments 1 and 2 trained participants on sample practice trials that familiarized them with how to build their responses. Instructions showed how the same list of options could be used to make many different kinds of responses to the same problem, i.e., they explicitly encouraged participants to consider multiple response strategies. This sentence construction methodology was also flexible enough to allow for an attention verification within the same general problem design. Attention check trials were nearly identical to the problems in the study with the exception that participants were told to create the nonsensical sentence, "Believes that knows that nothing follows" rather than providing their own response. This provided a seamless transition between experimental problems and attention checks to verify participants' focus on the task.

## Experiment 1

Experiment 1 tested whether individuals make omniscience errors in reasoning about agents and their knowledge or belief about the world. The experiment provided participants with three sentences – one context sentence, a conditional statement about the mental state of an agent, and either a statement about the mental state of the agent (the *epistemic* problems) or a statement about information in the world (the *factual* problems). Here is an example of such a problem:

3. Ash notices something in the environment.
   Ash knows that if it's 1:30 pm then it is a dove.
   [*Ash knows that it is* / *It is*] 1:30 pm.

The epistemic problems provided mental state information in the second premise, e.g., *Ash knows that it is 1:30pm.* In these problems, it may be reasonable to infer that *Ash knows that it is a dove*. Half of the problems were epistemic, and the other half presented factual information, e.g., *It is 1:30pm*, which does not imply anything about Ash's mental state.

The experiment likewise manipulated the epistemic verbs on each trial: half used the verb *know* and the other half used the verb *believe*. The experiment used the same verb in both sentences on each problem, as shown in (3).

## Method

**Participants** 75 participants (mean age = 41.73 years; 35 females, 38 males, 2 prefer not to answer) performed the study using the Amazon Mechanical Turk online platform (see Paolacci, Chandler, & Ipeirotis, 2010, for a review). All participants reported being native English speakers. 16 participants failed attention checks; we excluded them from analysis and report analyses on the remaining 59 participants.

**Design, procedure, and materials.** Participants completed 10 trials in total, one at a time, using the quasi-generative online interface described in the previous section. The study included two attention check trials to verify participant engagement. These trials asked participants to select certain words from the word list to create a particular nonsense sentence. All other problems consisted of a context statement, a conditional statement, and either an epistemic statement or a factual one depending on the problem type. After 1 second, the sentence-construction interface appeared. The context statement was the same for each problem (e.g., "X notices something in the environment."). The conditional was of the form "X believes/knows that if P then Q" where $X$ takes the place of the name of an agent, $P$ a time of day, and $Q$ an animal. The experiment randomized the contents of each problem. The statement that followed the conditional either was of the form $P$ or else "X believes/knows that P". Half the problems used *believe* and half used *know* for each epistemic verb. The experiment therefore yielded a 2 (trial type: epistemic or factual) x 2 (verb type: *know* or *believe*) repeated-measures design.

When the quasi-generative interface appeared, it prompted participants to "click words to fill in the blank to indicate what, if anything follows". The interface presented a list of clickable buttons, which allowed participants to create complete sentences by selecting words from those provided. They received no feedback about the the sentence they constructed.

**Coding rubric.** Omniscience errors can occur only on factual trials; on those trials, we coded responses that contained the agent and either of the epistemic verbs as erroneous. These responses attributed a mental state (i.e., knowledge or belief) to the agent in the trial even though the trial stipulated only information about the state of the world, which agents may or may not have access to. Reasoners could make other sorts of errors, too: they could, e.g., provide a nonsense response; or else fail to generate a valid inference and erroneously conclude that nothing follows; or else respond in a way that was logically invalid regardless of any consideration of mental states. Table 1 lists a summary and examples of such errors in Experiment 1; the full coding rubric is available at https://osf.io/83ja6/.

## Results and discussion

Participants attributed epistemic states to agents even when it was not appropriate to do so. Figure 2 provides the proportion of omnisciences errors as a function of the epistemic verbs in the study. As the figure shows, participants overwhelmingly committed omniscience errors –they exhibited such errors for both the factive verb *know* (81% of trials; Wilcoxon test against chance, i.e., ~0%; $z$ = 6.91, p < .001, Cliff's $\delta$ = 0.86) and the non-factive verb *believe* (92% vs. 0%, Wilcoxon test, $z$ = 7.34, p < .001, Cliff's $\delta$ = 0.95). They committed omniscience errors more frequently for agents that held beliefs rather than knowledge (92% vs. 81%, Wilcoxon test, $z$ = 2.90, $p$ = .004, Cliff's $\delta$ = 0.15). 56 out of 59 participants made omniscience errors on at least one trial (binomial test, p < .001 using a conservative prior probability of .10). Table 1 provides a breakdown of all errors that participants commited across the study as a whole, both omniscience and otherwise; omniscience errors were more frequent than any other type of error.

Participants in Experiment 1 erroneously inferred both knowledge and belief on the parts of their agents. The study used contents such as *if it's 1:30pm then it is a dove*, i.e.,

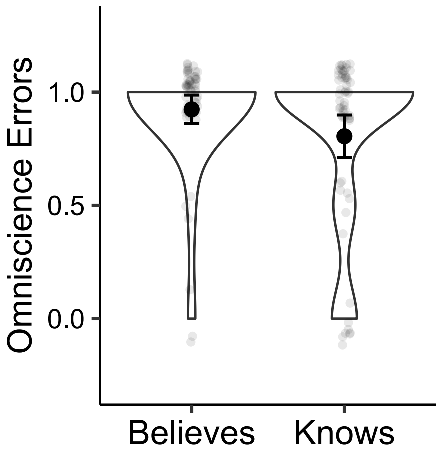

**Figure 2.** Violin plots of the proportion of omniscience errors for each epistemic verb in the factual condition for Experiment 1. Light gray circles denote individual participants' mean proportions of omniscience errors; dark black circles denote mean proportions across all participants; error bars denote 95% confidence intervals.

| Type of error | % | Example Trial | Example Answer |
|---|---|---|---|
| Omniscience errors | 86 | X believes that if it's 1pm then it is a badger. It's 1pm. | "X knows that it is a badger." |
| Nonsense response | 1 | X knows that if it's 5pm then it is a frog. It's not a frog. | "A frog." |
| "Nothing follows" errors | 3 | X knows that if it's 10am then it is a ladybug. It's 10am. | "Nothing follows." |
| Logical errors | 10 | X believes that if it's 9am then it is a newt. It's not a newt. | "It's 9am." |

**Table 1.** Examples of errors produced by participants in Experiment 1, the percentage of the total errors, example trials for which the error was relevant, and an example erroneous answer. Omniscience errors were the most frequent error that participants produced.

times of day and common animals. Participants may have assumed that reasoners have full access to the time of day and full knowledge about what doves look like, and so perhaps their omniscience errors reflect assumptions about such contents instead of patterns of epistemic reasoning. Experiment 2 sought to rule out this deflationary hypothesis. Another limitation of the study is that its problems consisted of both factual trials, which can yield omniscience errors, as well as epistemic trials, for which analogous conclusions are not errors but rather plausible inductive inferences. Because participants could only commit errors of omniscience on half of the trials of the study, the design was low-powered. Experiment 2 accordingly used a higher powered design to rule out the effects of content as a deflationary explanation for omniscience errors.

## Experiment 2

Experiment 2 tested whether participants were sensitive to information about agents' access to information. Omniscience errors could come about simply because participants infer that agents have access to information confirming or denying the clauses of a conditional. Problems in Experiment 2 accordingly made clear an agent's ability to access certain information. The experiment concerned a scavenger hunt scenario in which agents explored a building to uncover clues and find passwords. Information about the password depended on an object's presence at a particular location, e.g., here is one such conditional used in the study: "...if a globe is in the library then the password is pear." Some problems described agents who were in the same room as the object and who possessed knowledge about the conditional linking the object with the password. Here is an example of such a problem:

4. Ari is in the library. [access condition]
   The library is open and accessible.
   Ari knows that if a globe is in the library then the password is pear.
   A globe is in the library.

In this example, Ari has access to the library and knowledge that the globe's presence in the library implies that the password is "pear". Unlike in the previous study, Experiment 2 stipulated only factual information about the location of the object, that is, that the globe was in the library. Some participants may correctly deduce that the password is pear; others may infer incorrectly that Ari knows the password. To do so is to commit an omniscience error.

The other half of the problems described scenarios in which an agent had no access to information relevant to the conditional, e.g.:

5. Taylor is in the study. [no access condition]
   The office is locked and inaccessible.
   Taylor knows that if a map is in the office then the password is cherry.
   A map is in the office.

In (5), the agent isn't collocated with the relevant object, and the room that holds the relevant object is locked and inaccessible. The instructions of the study likewise make clear that no cameras or remote detection devices are present in the entire building. Hence, we refer to (5) as the *no-access* condition, whereas (4) presents the *access* condition.

Both (4) and (5) depict a modus ponens problem structure, i.e., one in which the truth of the *if*-clause is asserted. Half of the problems in Experiment 2 described modus ponens problems and the other half described modus tollens problems, which negated the *then*-clause, e.g.,

6. Sammy is in the planetarium.
   The closet is locked and inaccessible.
   Sammy knows that if a jar is in the closet then the password is pineapple.
   The password is not pineapple.

Experiment 2 sought to test whether participants are less likely to commit omniscience errors in the no-access condition; indeed, the no-access condition could eliminate such errors altogether. If errors persist, then they may be a robust and pervasive phenomenon of epistemic reasoning.

### Method

**Participants.** 63 participants (mean age = 37.16 years; 26 females, 34 males, 3 prefer not to answer) performed the study using the Amazon Mechanical Turk online platform. All participants self-reported as native English speakers. 9 subjects were excluded before analysis for failing attention check trials, yielding N = 54.

**Design, procedure, and materials.** Participants responded to 18 problems – 16 experimental and 2 attention checks. On each problem, participants read 4 sentences; the first specified which room an agent was located in; the second described either that same room, or else another room that held important information, and it stipulated whether the room was accessible to the agent or not. The third statement described a conditional (e.g., "...if a globe is in the library then the password is pear"), whose *if*-clause described an object within a room, and whose *then*-clause stipulated a password that the agent desired. The fourth statement either asserted the *if*-clause of the preceding conditional (yielding a putative modus ponens inference) or it negated the *then*-clause (yielding a putative modus tollens inference). Unlike Experiment 1, this last statement did not contain an epistemic verb in any condition. The experiment reflected a 2 x 2 x 2 repeated-measures design that manipulated the epistemic verb (*believe* vs. *know*), the problem structure (modus ponens vs. modus tollens), and the agent's access to information (access vs. no access).

### Results and discussion

Participants made omniscience errors in Experiment 2. They yielded such errors 54% of the time for problems that

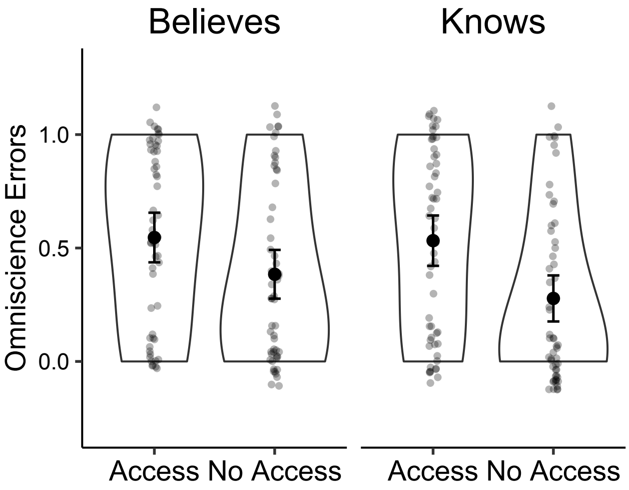

**Figure 3.** Violin plots of the proportions of omniscience errors for each epistemic verb and for both access conditions in Experiment 2. Light gray circles denote individual participants' mean proportions of omniscience errors; dark black circles denote mean proportions across all participants; error bars denote 95% confidence intervals.

described an agent who had access to information relevant to their mental state compared to 33% of the time when the agent did not have such access (Wilcoxon test, $z = 3.94$, $p < .001$, Cliff's $\delta = 0.29$). Participants were therefore sensitive to the accessibility of the information on each problem. Figure 3 shows the proportions of omniscience errors as a function of both the access condition and the verbs used in the study. The results revealed two overall patterns: first, participants were less likely to produce omniscience errors in Experiment 2 than Experiment 1; it may be that the manipulation of information access on some problems may have served as a cue for participants to consider the possibility – on *all* the problems – that the agents may not have full access to the information at hand. Second, as Figure 3 shows, even participants in the no-access condition yielded such errors more often than chance (e.g., ~0%; Wilcoxon test, $z = 5.58$, $p < .001$, Cliff's $\delta = 0.61$). As in Experiment 1, participants made more omniscience errors for *believe* than for *know* (47% vs. 41%, Wilcoxon test, $z = 2.55$, $p < .001$, Cliff's $\delta = 0.10$).

Experiment 2 shows that while omniscience errors can be moderated by making explicit agents have a clear separation from the information they are thinking about, participants still make these errors a large percentage of the time.

## General Discussion

Two experiments revealed systematic reasoning errors in adult reasoners. Participants considered problems such as:

Layla knows that if it's 7pm then it's a frog.
It's 7pm.
What, if anything, follows?

and concluded that *Layla knows that it's a frog* on 86% of factual trials in Experiment 1 and 44% of analogous trials in Experiment 2. Since the problem provided no information about whether Layla knows the time or not, it is a mistake to conclude that she knows any consequences of that fact. But, perhaps fairly, people may presume that Layla has constant access to information about what time it is: it is not unreasonable to suppose that she wears a watch or carries a cellphone. Experiment 2 sought to test whether people made omniscience errors even for scenarios that stipulated an agent's lack of access to relevant information. The manipulation reduced omniscience errors, but did not eliminate them. Participants in Experiment 2 continued to make omniscience errors more often than they committed any other error.

It may be that the errors we report came about only because of the task and design used in Experiments 1 and 2, that is, they do not reflect how people reason about knowledge and belief in general. We designed a novel quasi-generative task and interface (see Figure 1) so that participants could construct only those sentences that they deemed appropriate. The task explicitly permitted participants to answer that nothing followed, and the probability of generating any conclusion unintentionally was low. Attention check trials that looked like regular problems helped to select those participants who understood and were engaged in the task. Nevertheless, because the interface presented buttons that included epistemic sentence fragments attached to agents, e.g., ("Layla knows that..."), their presence may have served as a cue for participants to consider the mental states of the agent. Alternative user interfaces, e.g., those that include drop-down menus or predictive text, may help eliminate this concern.

If the interface we introduced adequately indexes people's patterns of epistemic reasoning, then it reveals that people systematically conflate information about the mental states of others with information they possess about the world. The result may suggest that the "curse of knowledge" studied by other researchers (Birch & Bloom, 2004; 2007) is a prosaic pattern of mental state reasoning.

Experiment 2 likewise suggests some ways to eliminate the error: it explicitly informed participants about the accessibility of certain pieces of knowledge. It may be that establishing that information, whether explicitly or implicitly, can help mitigate the error and promote better epistemic reasoning. Future studies will investigate this possibility.

In sum, we present evidence that people who are otherwise competent in reasoning about mental states make systematic errors by projecting knowledge into the minds of those who may not possess it.

## Acknowledgments

This work was supported by an NRC Research Associateship Award to BB and funding from the Naval Research Laboratory to SK. We also thank Kalyan Gupta, Danielle Paterno, and the Knexus Research Corporation for their help in data collection, and we are indebted to Bill Adams, Gordon Briggs, Magda Bugajska, Tony Harrison, Andrew Lovett, Maria Kon, and Greg Trafton for their critiques and comments.

# References

Apperly, I.A., Samson, D., Humphreys, G.W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, 45, 190-201.

Austin, G., Groppe, K., & Elsner, B. (2014). The reciprocal relationship between executive function and theory of mind in middle childhood: A 1-year longitudinal perspective. *Frontiers in Psychology, 5*, 1–11.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Bio, B. J., Guterstam, A., Pinsk, M., Wilterson, A. I., & Graziano, M. S. (2022). Right temporoparietal junction encodes inferred visual knowledge of others. *Neuropsychologia*, 171, 108243.

Bio, B. J., Webb, T. W., & Graziano, M. S. (2018). Projecting one's own spatial bias onto others during a theory-of-mind task. *Proceedings of the National Academy of Sciences*, 115(7), e1684-e1689.

Birch, S. A., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, 8, 255-260.

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

Bolander, T. (2018). Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic. In Jaakko Hintikka on knowledge and game-theoretical semantics (pp. 207-236). Springer, Cham.

Bringsjord, S., & Sundar Govindarajulu, N. (2020). Rectifying the mischaracterization of logic by mental model theorists. *Cognitive Science*, 44, e12898.

Carey, S., Leahy, B., Redshaw, J., & Suddendorf, T. (2020). Could it be so? the cognitive science of possibility. *Trends in Cognitive Sciences*, *24*, 3-4.

Dalke, D. E. (1995). Explaining young children's difficulty on the false belief task: Representational deficits or context-sensitive knowledge?. *British Journal of Developmental Psychology*, *13*, 209-222.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, 50, 21-45.

Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science*, 16, 291-297.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13, 331-338.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, 19, 249-265.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40, 760–768.

Fagin, R., Moses, Y., Halpern, J. Y., & Vardi, M. Y. (1995). Knowledge-based programs. In *Proceedings of the 14th Annual ACM Symposium on Principles of Distributed Computing*.

Harner, H., & Khemlani, S. (2022). Reasoning about *want*. *Cognitive Science,* 46.

Hintikka, K. J. J. (1962). *Knowledge and belief: An introduction to the logic of the two notions.* Ithaca: Cornell University Press.

Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). Quantitative mental state attributions in language understanding. *Science Advances,* 7, eabj0970.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107.

Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, 103950.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25-41.

Khemlani, S. S. (2018). Reasoning. In S. Thompson-Schill (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. Wiley & Sons.

Khemlani, S. S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and Machines*, 27, 11-35.

Khemlani, S., Wasylyshyn, C., Briggs, G., & Bello, P. (2018). Mental models and omissive causation. *Memory & Cognition*, 46.

Kouklari, E. C., Thompson, T., Monks, C. P., & Tsermentseli, S. (2017). Hot and cool executive function and its relation to theory of mind in children with and without autism spectrum disorder. *Journal of Cognition and Development*, *18*(4), 399-418.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830-1834.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford, UK: Oxford University Press.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5, 411-419.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125-137.

Pfeifer, N., & Kleiter, G. D. (2009). Mental probability logic. *Behavioral and Brain Sciences*, 32, 98-99.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, 23, 1026-1040.

Pietarinen, A. V. (2003). What do epistemic logic and cognitive science have to do with each other? *Cognitive Systems Research*, 4, 169-190.

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.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17*, 74–81.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, 23, 842-847.

Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, 425-440.

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, 30, 75-86.

Sutcliffe, G. (2017). The TPTP problem library and associated infrastructure. *Journal of Automated Reasoning*, 59, 483-502.

Van De Pol, I., Van Rooij, I., & Szymanik, J. (2018). Parameterized complexity of theory of mind reasoning in dynamic epistemic logic. *Journal of Logic, Language and Information*, 27.

Van Ditmarsch, H., & Labuschagne, W. (2007). My beliefs about your beliefs: a case study in theory of mind and epistemic logic. *Synthese*, 155, 191-209.

Von Wright, G. H. (1951). *An essay in modal logic.* Amsterdam: North Hollan.