Factive mindreading reflects the optimal use of limited cognitive resources

Tadeg Quillien*[1] and Max Taylor-Davies[2]

[1]Department of Psychology
[2]School of Informatics
[1,2]University of Edinburgh

January 23, 2026

*Corresponding author. Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ. tadeg.quillien@gmail.com.

Abstract

The capacity to represent the mental states of other individuals, known as Mindreading or Theory of Mind, is key to successful social prediction. We suggest that cognitive systems for mindreading are resource-rational: they are optimized for generating good predictions about the behavior of other individuals, while not exceeding the computational capacity of the mindreader. We explore this hypothesis in a simple formal model where we derive cognitive strategies that excel at social prediction while minimizing cognitive effort. We find that it is often optimal for resource-limited mindreaders to keep track of the facts that another agent also knows, instead of explicitly representing the content of the agent's beliefs. When evaluated in mindreading tasks, simulated agents that use this 'factive' strategy tend to make mistakes in the same cases as non-human primates and young human children. Even agents that use more sophisticated strategies avoid representing beliefs unless necessary. Our results elucidate the computational principles underlying efficient social prediction, and explain many of the successes and failures of human and non-human mindreading from first principles.

Keywords: theory of mind, knowledge, false belief, information theory, resource rationality, social cognition

1 Introduction

Predicting the behavior of other individuals is a key adaptive challenge for most organisms. The challenge of social prediction has been a key driver of the evolution of Theory of Mind, or ‘mindreading’: the ability to represent the latent mental states of others. This capacity has been extensively studied across multiple domains, including its evolutionary origins, development in children, neural mechanisms, and conceptual structure [1–8].

Researchers have recently started to develop formal theories of mindreading at the computational level [e.g. 7, 9, 10]. These theories typically use a *normative* modeling approach; in a normative approach, one compares participants’ behavior to the predictions of rational models that solve an information-processing problem in the optimal way [11–13]. Specifically, recent theories of mindreading assume that mindreaders have an internal causal model of the mental states of other agents, and can update this model in an approximately Bayesian way [7]. This approach has been successful for explaining the successes of mindreading in human adults and children across many tasks [7, 10, 14].

However, a normative approach is less well suited to explaining the patterns of systematic mistakes that participants—especially younger children and non-human animals—make in mindreading tasks [1, 15]. Consider for example the following setting: a character named Sally puts a ball in a basket, and then goes away. While Sally is away, Anne removes the ball from the basket and puts it in a box. When Sally comes back, where will she look for the ball? This task requires the participant to predict the behavior of an agent whose belief (the ball is in the basket) does not match reality (the ball is now in the box). Human children younger than four typically answer incorrectly ([15, 16], but see [3, 17]). Variants of the task adapted for non-human animals show that non-human primates also struggle to represent false beliefs ([18–20], but see [21–24]). These findings are difficult to explain in formal models that assume that mindreaders engage in approximately optimal computations [7, 10]. Several computational models have been developed to explain difficulties in mindreading [25–29], but these models typically can explain a limited range of findings, and often make relatively strong assumptions about cognitive architecture.

In this paper, we develop a computational approach that can account for both the successes and limitations of mindreading, building from first principles. Like many existing models, we take a normative approach, asking how a well-designed cognitive system for mindreading would work. However, we also consider how this cognitive system would deal with limitations in computational resources. From this perspective, systematic mistakes can be understood as resulting from cognitive 'shortcuts' that save computation [30].

Formally, we conduct a *resource-rational analysis* of mindreading in a simple model of social prediction. In a resource-rational analysis, researchers seek to derive the optimal policy for solving an information-processing problem, under the constraint that this policy has to be executed by an agent with limited computational resources [30–32]. Here we consider a large space of possible policies for social prediction, and find the policies that optimize predictive performance under computational resource constraints. This process allows us to study social prediction policies that have been 'designed' by a normative optimization process, rather than hand-coded by a researcher.

To preview our results, we find that the optimal policy for mindreading depends on the amount of computational resources available to the mindreader. Mindreaders with high resources use a *meta-representational* strategy: they represent the beliefs of the other agent [33, 34]. Mindreaders with more limited resources adopt a *factive* strategy, simply tracking what the other agent knows [1, 35–37]. Importantly, mindreaders using a factive strategy make systematic mistakes in the same situations as non-human primates, young human children, and human adults under cognitive load ([16, 20, 38, 39], for review see [35, 40]). This similarity suggests that mistakes in mindreading tasks stem from the use of cognitive strategies that are optimally designed to save computation at the expense of accuracy [30].

The difference between meta-representational and factive mindreading can be understood as follows. Observers that use meta-representations are constructing a model of the way another agent models its environment. For example, if Alice and Bob are in a room, Bob's world model might contain the information that 'there is an apple on the table', and Alice can meta-represent that 'Bob thinks: "there is an apple on the table"', see Figure 1A. This approach is very flexible, because it can accomodate cases where Bob has a different belief than Alice. For example if Bob mistakenly thinks that there is an orange on the table, Alice can meta-represent 'Bob thinks: "there is an orange on the table"' while simultaneously representing in her own world model that the fruit on the table is an apple. At the same time, meta-representation can be extremely costly in computational terms [41]. Consider just the memory demands: storing another individual's complete model of the world could in principle require as much memory as your own model of the world.

Factive mindreading is a simpler strategy. Consider again Alice and Bob, in the same room: they share much of their knowledge about their environment, such as seeing an apple on the table. Alice can store 'there is an apple on the table' in her own world model, and add a simple tag noting that Bob also has access to this fact (Figure 1 Right;

[37]). This simpler strategy is called 'factive' because it represents relations between the other agent and true facts about the world [35, 42].

Factive mindreading is less flexible than meta-representation: it only allows Alice to represent Bob as knowing or ignoring a fact. So, Alice cannot represent Bob as having a different belief than hers. But by avoiding duplicate representations of the shared environment, factive mindreading comes at a significantly lower computational cost. Here we show that factive mindreading can in fact be the optimal cognitive policy for an organism with limited computational resources. This result provides a normative explanation for a wide range of empirical findings about human and non-human mindreading.

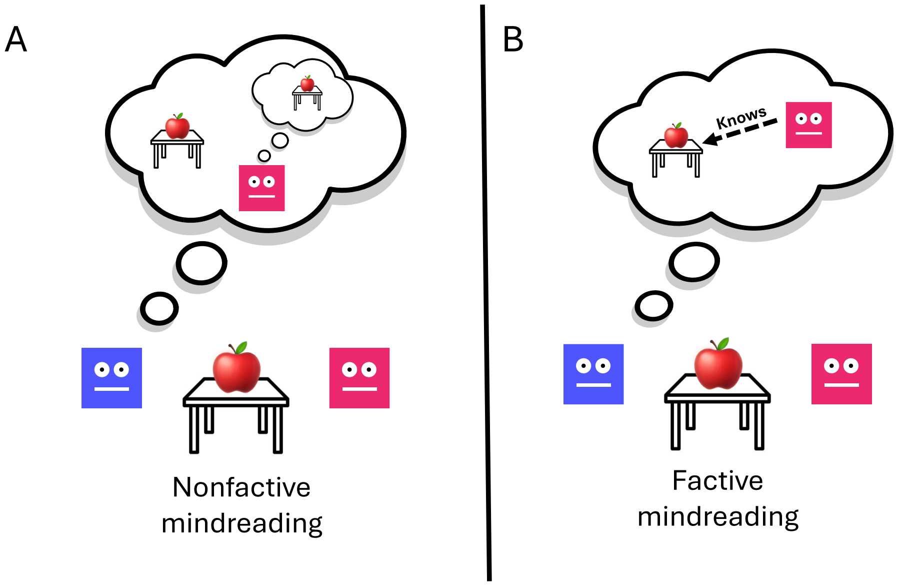


Figure 1: Difference between nonfactive and factive mindreading (adapted from [37]). **A**: The mindreader (blue) represents a fact (the apple is on the table) in its primary representation of the world, and also represents the other agent (pink) as representing that there is an apple on the table (a meta-representation). **B**: A factive mindreader simply tracks whether the other individual has epistemic access to a fact in the mindreader's world model.

Our computational approach allows us to model social cognition without relying on folk-psychological concepts. Following previous work, we still use some folk-psychological language *for ease of exposition*. Specifically, for convenience we say that factive mindreaders represent what other agents *know* instead of what they *believe* [40].

For our purposes, 'knowledge' denotes two important properties: agents can only know things that are true, and accidentally true beliefs do not count as knowledge [42, 43]. In contrast, the content of an agent's belief is a proposition like 'the apple is on the table'.

We operationalize computational limitations in information-theoretic terms, as a bound on how much information the observer is able to extract from the environment. The advantage of this approach is that it allows us to remain agnostic with respect to particular implementation or substrate details—since information-theoretic constraints can be interpreted in multiple ways, such as limitations on inference or memory [44]. Information-theoretic principles have been widely used in models of resource-rational cognition [45–61]. They offer a principled way to model cognitive resource limitations in the abstract, without making strong assumptions about cognitive architecture [44].

2 Modeling framework

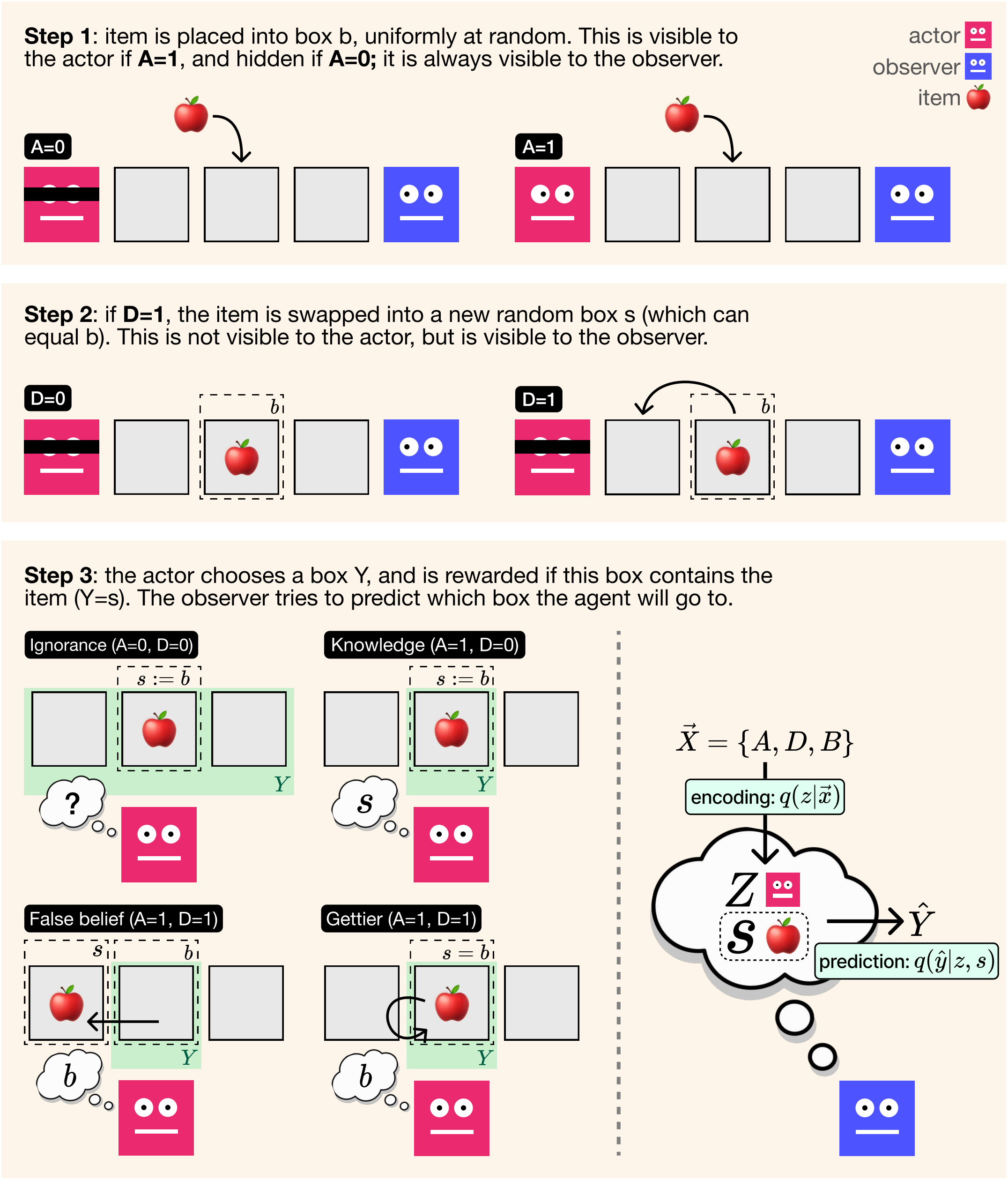
We consider an *observer* who has to predict the behavior Y of an *actor*—for example the observer must predict where the actor will look for an apple. The observer has access to a stream of data $\vec{X}$ from the world, some of which is relevant to predicting the actor's behavior (information about the actor's location, gaze direction, etc). An observer with limited cognitive resources cannot process in detail all the information contained in the incoming sensory data, so they need to construct a compressed representation Z, that they will then use to predict Y. Ideally, Z extracts the information in $\vec{X}$ that is most relevant to the task of predicting the other individual's behavior (Figure 2 lower-right).

This problem can be formalized using the information bottleneck [62], a framework closely related to rate-distortion theory [63]. In an information bottleneck problem, we seek to construct an optimal encoder from $\vec{X}$ to Z. Formally, an encoder is a conditional probability distribution $q(z|\vec{x})$ that specifies the probability that the observer will form the representation $Z = z$ given that the state of the world is $\vec{X} = \vec{x}$, for all possible values of $\vec{x}$ and z.

The computational capacity of the observer is defined as an upper bound on the mutual information between $\vec{X}$ and Z:

$$I(\vec{X}; Z) = \sum_{\vec{x},z} \mathrm{Pr}(\vec{x}, z) \log \frac{\mathrm{Pr}(\vec{x}, z)}{\mathrm{Pr}(\vec{x})\mathrm{Pr}(z)} \tag{1}$$

where $Pr(\vec{x}, z) = q(z|\vec{x})Pr(\vec{x})$. Intuitively, this value quantifies the amount of information that compressed representation Z can 'preserve' about the input data $\vec{X}$. Given this upper bound on mutual information, the goal is to find an encoder that, on average, yields the representation Z that is most useful for predicting the actor behavior Y.



Step 1: item is placed into box b, uniformly at random. This is visible to the actor if **A=1**, and hidden if **A=0;** it is always visible to the observer.

Step 2: if **D=1**, the item is swapped into a new random box s (which can equal b). This is not visible to the actor, but is visible to the observer.

Step 3: the actor chooses a box Y, and is rewarded if this box contains the item (Y=s). The observer tries to predict which box the agent will go to.

Figure 2: Social prediction task and theoretical framework. Steps 1 to 3 describe the dynamics of the task. In Step 3, green shading indicates the box(es) that the actor is most likely to go to in each case; thought bubbles represent where the actor thinks the item is. Lower right: information bottleneck model. $\vec{X}$ represents information in the world relevant to mindreading, such as what the actor (pink) can and cannot see. The observer (blue) constructs a compressed representation Z on the basis of $\vec{X}$, and also has access to additional representation S which reflects the world state (i.e. the true item location). The observer then uses Z and S to make a prediction $\hat{Y}$ about the actor's box choice Y.

Crucially, we assume that the observer also has a representation S of the state of the physical world, because this representation is generally useful even outside of the context of social prediction. For example, the observer tracks the true location of the apple because they may want to eat it themselves. We assume that the information-theoretic costs of building representation S have already been paid by the observer, so that it can effectively be re-used for free in the social prediction task. We can then re-frame the task as that of predicting Y from Z and S *jointly* (see Figure 2 lower-right), with the usefulness of Z quantified as the additional predictive power that it gives the observer about Y, given that the observer already represents S. This quantity is operationalized as a conditional mutual information:

$$I(Y;Z|S) = I(Y;Z,S) - I(Y;S) \tag{2}$$

In sum, we are looking for the optimal encoder:

$$q_C(z|\vec{x})^\star = \underset{q}{\text{arg max}}\; I(Y;Z|S) \tag{3}$$
$$\text{subject to } I(\vec{X};Z) \leq C$$

where C is the upper bound on the amount of information the observer can extract from $\vec{X}$. Given the conditionalization on S, our problem is an instance of the *conditional* information bottleneck, and we solve it using a variant of the Blahut-Arimoto algorithm [62, 64, 65] given by [66], see Supplementary Information.

Importantly, we are not arguing that resource-rational mindreaders are solving Equation 3 themselves. Instead, resource-rational analysis takes the perspective that the constrained optimization problem has been approximately solved over time by evolutionary, developmental or learning processes, and that the observer is simply executing the resulting policy [30]. For simplicity we focus on the cognitive costs involved in constructing representation Z, but not in the costs involved in deriving a prediction from Z (following e.g. [51, 67]; but see Supplementary Information for preliminary analysis of decoding costs). Therefore we assume that the observer predicts behavior Y with the Bayes-optimal decoder $q(\hat{y}|z,s)$.

2.1 Task

We study the resource-rational mindreading problem in a simple task in which the observer must predict the behavior of an actor.

2.1.1 Actor's task.

The actor faces N opaque boxes. One of these boxes B is selected uniformly at random, and a valuable item (such as the apple in our earlier example) is placed into box B. The actor will have to choose a box and gets reward r if it picks the box containing the item, and 0 otherwise, see Figure 2.

With some probability $Pr(A)$, the actor has perceptual access and can see in which box the item is initially being placed (i.e. it can see which box is selected as B). Otherwise (with probability $1 - Pr(A)$), the actor is ignorant and receives no information about the item's location.

With probability $Pr(D)$, we then switch the item to a box S, selected uniformly at random (this can be the original box B), *always* outside of the actor's awareness. This 'Deceiver' event ($D = 1$) implies that any belief that the actor has formed might now be false.[1]

We use S to denote the final location of the item; if the item did not get switched we simply have $S = B$.

Following recent computational approaches to mindreading [7, 9, 10, 29], we assume that the actor is approximately rational and seeks to maximize expected reward given the information it has access to (see Supplementary Information). This means that the actor chooses a box uniformly at random if it did not see where the item was placed ($A = 0$). Otherwise ($A = 1$), it goes to the box where it last saw the item (box b), although it may sometimes choose a different box by mistake.

2.1.2 Observer's task.

The observer already has a representation of current location S of the item, and can extract the value of A, D and B as input data from the environment; i.e. we have $\vec{X} = \{A, D, B\}$. The observer's goal is to accurately predict where the actor will go, that is, to accurately estimate the probability of each choice.

This simple setting allows us to explore different situations traditionally studied in Theory of Mind research, including tasks where the observer must predict the behavior of an actor with knowledge ($A = 1$, $D = 0$), false belief ($A = 1$, $D = 1$, $s \neq b$), accidentally true belief ($A = 1$, $D = 1$, $s = b$), and ignorance ($A = 0$). Intuitively, variable B represents the 'content' of the actor's belief (assuming that $A = 1$), while A and D determine whether the actor knows the item location (specifically, the actor has knowledge if $A = 1$ and $D = 0$).

[1] 'Deception' is not meant to imply that there is another agent who is strategically acting to manipulate the actor's beliefs.

Below we derive the resource-rational observer policies for this task, using the framework outlined in the previous section, and investigate their properties. We call the combination of parameters $Pr(A)$, $Pr(D)$ and N the *social ecology*; in addition to these, we also vary the computational capacity C of the observer. In a given simulation, the values of parameters $Pr(A)$, $Pr(D)$, N and C are fixed, and the resource-rational policy is optimized for its expected performance across all possible settings of A, D, B and S (the probability of each setting is determined by the social ecology). Note that different social ecologies can in principle favor different resource-rational policies; in this sense resource-rational policies are *ecologically rational* [68, 69].

We predict that an observer that can only dedicate limited resources to constructing Z should focus these limited resources on encoding information that is least likely to be redundant with S. From this perspective, encoding the content B of the actor's belief ('the apple is in box 3') can be wasteful, because this information is typically already in the observer's own representation S. Instead, the observer can encode the value of A and D: whether the actor 'knows' the location of the apple. We therefore define as *factive* a policy that i) extracts little or no information about B, ii) extracts relatively more information about A and D. We measure the information extracted about a variable as the mutual information between the variable and compressed representation Z. Code for implementing our model is available on the Open Science Framework.

3 Results

We find that factive mindreading emerges as the optimal cognitive strategy across a non-trivial portion of the parameter space, see Figure 3B. In many social ecologies, resource-rational observers with low computational capacity extract substantially more information about the knowledge-relevant variables A and D than the belief-relevant variable B. Figure 3A illustrates this pattern for one example social ecology: observers with low computational capacity extract information about A and D but not about B, which is only represented by observers above a certain capacity threshold.

We can explore the representations formed by factive observers by visualizing the mapping from $\vec{X}$ to Z to $\hat{Y}$ in observers that extract no information about B. A detailed example is given in the Supplementary Information (Figure S3). Here we give a high-level overview of this content, showing an idealized depiction of the mapping performed by factive vs meta-representational observers (Figure 4). We find that factive observers have a representation Z that can be in only two possible states: the observer either represents the actor as being Ignorant (whenever $A = 0$ or $D = 1$) or Knowledgeable (whenever $A = 1$ and $D = 0$). Correspondingly, the observer predicts that an Ignorant

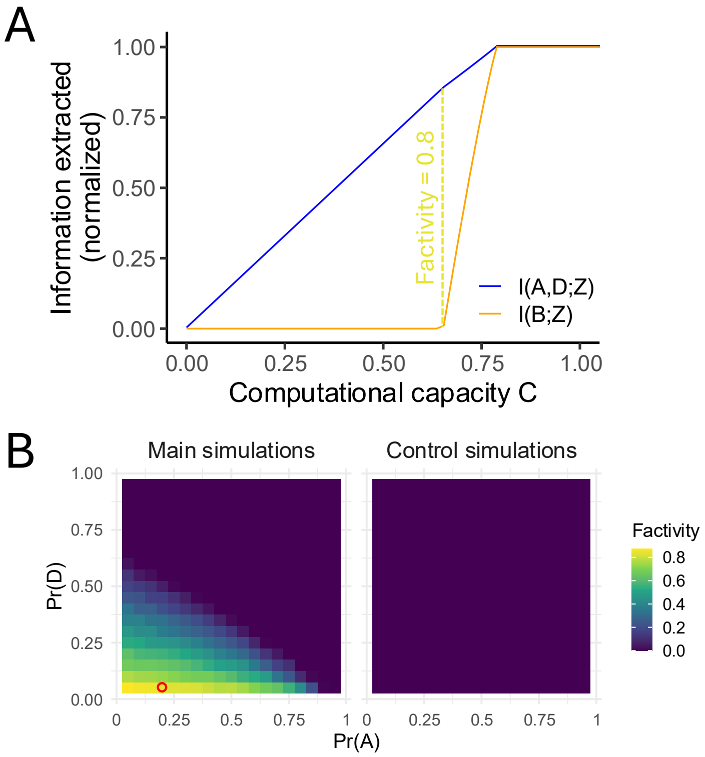


Figure 3: **A:** Amount of information that resource-rational observers extract from knowledge-relevant variables A and D (blue), and belief-relevant variable B (orange), as a function of the observer's computational capacity C, shown here for $N = 3$, $Pr(A) = .2$, $Pr(D) = .05$. Each point on the x-axis corresponds to a different resource-rational observer. Information extracted is normalized such that 1 represents the amount of information extracted by the observer with the largest computational capacity. **B:** Prevalence of factive policies across parameter space, shown for $N = 3$. Red dot corresponds to parameters used in panel A. Factive policies are prevalent in ecologies in which the likelihood Pr(D) of an unwitnessed transfer is low, and the likelihood Pr(A) that the actor witnesses the original hiding is low-to-intermediate. Factivity is computed as the maximum value of $I(A, D; Z) - I(B; Z)$ across values of C, normalized as in A. Intuitively, the brightness of a tile indicates how much higher than the orange line the blue line can get in a plot such as in panel A—see dashed line. In control simulations, the observer does not have a pre-existing representation of S.

actor might go toward any box, and predicts that a Knowledgeable actor will go to box

S (the box that actually contains the item).

In contrast, in social ecologies with high values of $Pr(D)$ and $Pr(A)$, or for observers with high computational capacity, the resource-rational policy is closer to a meta-representational policy. The lower panel on Figure 4 is an idealized depiction of a meta-representational policy. The observer represents the content of the actor's beliefs, like the belief that the item is in box 1. The representation Z extracts all the available information about B ($I(B;Z)$ is high), and the observer does not use its own representation of the state of the world S.

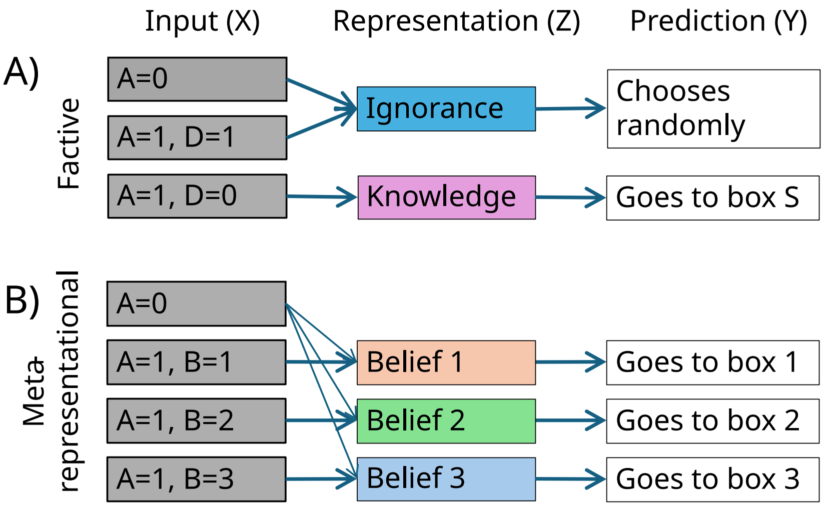


Figure 4: Schematic description of the $\vec{X} \rightarrow Z \rightarrow \hat{Y}$ mapping in a factive observer (**A**) and a meta-representational observer (**B**), shown for $N = 3$. The meta-representational observer in the state $A = 0$ maintains a uniform probability distribution over the three possible belief states. Note that actual policies are typically more stochastic than these simplified mappings, see Figure S3.

Figure 3B shows the prevalence of factive mindreading across social ecologies. Factive mindreading is especially prevalent for low values of $Pr(D)$ and low-to-intermediate values of $Pr(A)$. The likelihood of a false belief is equal to $Pr(A)Pr(D)$, and factive policies make sub-optimal predictions when the actor has a false belief, since in this case the item location S is not sufficient to predict what the actor will do. Factive mindreading is also more prevalent with increasing N (see SI), because the information-theoretic cost of extracting B increases with the number of possible beliefs the actor could have.

3.1 Experiments

Here we take a closer look at the performance of resource-rational observers by performing 'in-silico' experiments in our mindreading tasks. We also compare these results to existing empirical findings in similar tasks in human and non-human primates. We present results for three observers, a representative each of an 'automatic' policy ($C = 0$), a low-resource- ($C = .5$), and a high-resource observer ($C = 1$). The automatic policy cannot extract any information from the input data and is therefore 'blind' to differences between tasks. The low-resource observer is of special interest because it is a factive mindreader, as can be seen in Figure 3A. To generate the predictions of an observer, we compute $\Pr(\hat{Y}|\vec{X}, S)$ by marginalizing over Z:

$$\Pr(\hat{Y}|\vec{X}, S) = \sum_{Z} q(\hat{Y}|Z, S)q(Z|\vec{X}) \tag{4}$$

We use a social ecology with $N = 3$ boxes, $Pr(A) = .2$ and $Pr(D) = .05$, and report experiments for other social ecologies in the Supplementary Information.

3.2 Predicting behavior

In our main series of tasks, the observer has to predict which box the actor will reach toward. In **Experiment 1**, the observer must predict the behavior of an actor who knows the location of the item ($A = 1$, $D = 0$, upper-left on Figure 5). We find that all observers correctly predict that the actor will reach for the box containing the item, although this inference is stronger in observers with more cognitive resources. This result mirrors experiments in adults, children, and non-human primates; individuals from these populations can attribute knowledge, but human adults do so more reliably [18, 20, 29, 40, 70, 71].

In **Experiment 2**, the observer predicts the behavior of an actor who has a false belief ($A = 1$, $D = 1$, and $b \neq s$, upper-right on Figure 5). Only the high-resource observer correctly predicts that the actor will reach for the item where it last saw it. The low-resource observer is mostly agnostic, maintaining an almost uniform distribution over boxes, with only a slight bias toward the actual location of the item.[2] This pattern

[2] This bias emerges for the following reason. An observer who encodes no information about $\vec{X}$ should have a bias toward the reward's actual location, because when averaging across tasks the actor goes toward the reward's location more often than toward other locations. The low-resource observer encodes some information about $\vec{X}$, but this encoding is unreliable; so even in a false belief task a slight bias toward the reward location subsists.

again reflects experimental results: human adults can pass false belief tasks while non-human primates often fail them ([18, 19, 40, 72], but see [21–24]). Moreover, non-human primates fail false-belief tasks in the same way as the low-resource observer: they find each outcome equally surprising, including seeing the actor go toward a box where the item was never located [38]. Young human children also struggle with false belief tasks, although they fail in a slightly different way than the low-resource observer, because they predict that the actor will look for the item at its actual location ([16], see also [26]).

In **Experiment 3**, the actor is ignorant ($A = 0$, lower-right in Figure 5). The high-resource and low-resource observers correctly predict that the actor might go toward any box. Young human children and non-human primates similarly make different predictions depending on whether the agent is knowledgeable or ignorant [73–75], although children's predictions about ignorant agents are sometimes biased [76, 77]. See also Supplementary Information for more discussion on the implications of our results for theories of ignorance representation.

Experiment 4 has the structure of a 'Gettier case' in epistemology [43]. Outside of the actor's awareness, the item is removed from its original box but then put back in exactly the same box; as a result the actor has an *accidentally true belief* ($A = 1$, $D = 1$, $s = b$, upper-right on Figure 5). While the high-resource observer succeeds at the task, the low-resource observer expects that the actor might look at any location. The pattern of results for the low-resource observer is similar to that of non-human primates [20, 71], who also fail to represent an actor's belief if that belief is true only by luck. Similar patterns have been observed in human children ([20, 78, 79], but see [80]).

A striking finding from Experiment 4 is that the low-resource observer performs worse than the automatic policy, despite having higher computational capacity. The low-resource observer defaults to a near-uniform distribution whenever the actor lacks perceptual access to an event; this is usually a good strategy but it backfires in a Gettier case, where a simple bias toward the item's actual location actually does a good job of predicting the actor's behavior.

3.3 High-resource observers flexibly switch between knowledge and belief representation.

High-resource observers in our simulations successfully pass false belief and Gettier tasks. This finding might indicate that high-resource observers implement a fully meta-representational strategy: they encode the content B of the other agent's belief whenever

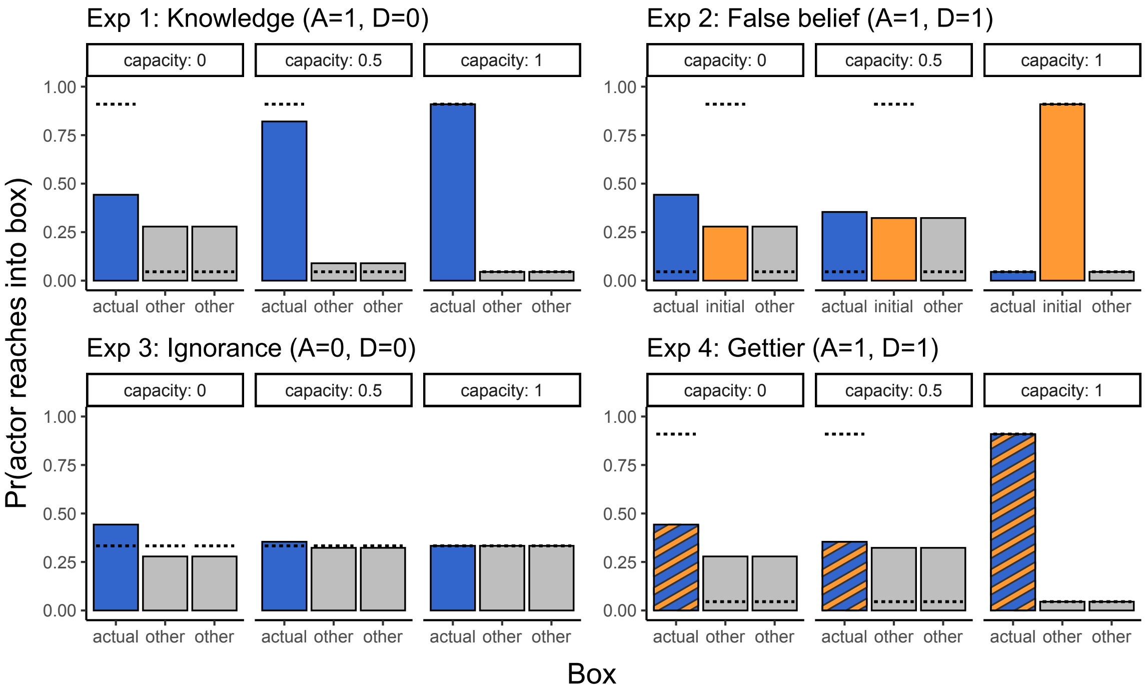


Figure 5: Predictions made in our experiments by resource-rational observers with different computational capacity. ‘Actual’: actual location s of the item; ‘Initial’: initial location b of the item. Dashed lines represent the ideal non resource-limited policy. Parameters used were $N = 3$, $Pr(A) = .2$, $Pr(D) = .05$.

that agent has perceptual access ($A = 1$), see Figure 4 lower panel. Alternatively, high-resource observers might use a flexible strategy: they only encode the content of the actor’s belief when the actor was deceived, i.e. when $A = 1$, $D = 1$, and track the actor’s knowledge otherwise. This strategy allows an observer to perfectly predict behavior, while potentially saving cognitive resources.

To assess which strategy more closely describes high-resource observers in our simulations, we computed the amount of information that an observer extracts about variables A, D and B, relative to the maximum possible information that can be extracted about that variable (its Shannon entropy). For high values of $Pr(D)$, high-resource observers approximate a fully-metarepresentational strategy, extracting a high portion of the available information about B. In contrast, for low values of $Pr(D)$, high-resource observers approximate a fully-flexible strategy: they represent knowledge by default, and only encode the content of an actor’s belief when this actor has a false or accidentally true belief (Figure S4).

3.4 Learning about the world

In the Supplementary Information, we report additional experiments investigating the performance of factive mindreaders in a simple social learning task (inspired by [40, 72]). We show that the representations optimized for our *first* task (predicting behavior from the state of the world), can also be co-opted for predicting the state of the world from observation of another individual's behavior.

3.5 Control simulations

We claim that factive mindreading can be adaptive because observers are already representing the state of the world S, and so can use this information at no extra cost for predicting others' behavior. In the Supplementary Information we report a complementary set of simulations where we abandon this assumption, and find that factive mindreading does not emerge when observers must pay the additional cost of representing S for mindreading-specific purposes—showing that this assumption is indeed essential to our results.

4 Discussion

Social prediction is a key adaptive challenge for many organisms. This challenge has been a key driver of the evolution of cognitive systems for 'mindreading', the ability to represent the mental states of other individuals. There has been considerable debate about the development of mindreading in humans, and the extent of its sophistication in other species [1–3, 16]. Explaining how mindreading works across ages and species requires reconciling two prima facie contradictory phenomena. On the one hand, mindreading seems highly imperfect: in some populations (like non-human primates and young human children) mindreaders make systematic patterns of mistakes that suggest the absence of a capacity to represent beliefs [1, 15, 16, 20, 71]. On the other hand, mindreading seems to function well: young human children and non-human primates are quite adept at predicting the behavior of others outside of contrived situations involving false beliefs [20, 35, 40, 70, 81]. Moreover, the best existing computational accounts of mindreading in human adults assume that mindreaders make approximately rational inferences [7, 10, 14].

Here we present a simple formal model that reconciles these two sets of observations, explaining the successes and limitations of mindreading from first principles. We suggest that cognitive systems for mindreading are subject to a trade-off between accuracy and tractability: they need to generate good predictions about the behavior of other individuals, while not exceeding the computational capacity of the mindreader [30]. From this perspective, the optimal cognitive policy depends on the computational resources that the mindreader can dedicate to the task of social prediction: mindreaders with fewer resources need to trade accuracy for computational efficiency, resulting in systematic mistakes.

In a simple social prediction task, we derived the optimal cognitive policy for mindreading across different social ecologies. Our key finding is that for mindreaders with low computational capacity, it might not be optimal to explicitly represent the beliefs of the other agent. In social ecologies where agents rarely acquire false beliefs, low-capacity mindreaders are better off simply tracking the facts that the other agent knows. This *factive* strategy occasionally generates the wrong predictions, but this inaccuracy is compensated by the computational savings that come from representing knowledge instead of belief. As other researchers already emphasized, the use of a factive strategy can explain many findings about mindreading in young human children and non-human primates [1, 35, 40].

Crucially, factive mindreading explains not only why individuals in these populations struggle in false belief tasks, but also their overall accuracy in more natural tasks that do not involve deception [20, 40, 70, 81]. Our approach can also explain the good performance of human adults, under the assumption that they can assign a lot of computational power for mindreading. In sum, the pattern of successes and mistakes in mindreading across ages and species might be explained in a unified way: organisms execute policies that are optimally designed for good predictive performance under cognitive constraints, and these constraints vary across ages and species. This proposal is consistent with the fact that human adults tend to have higher information-processing capacity than human children and other primates [82].

More speculatively, our framework might also explain why success in false belief tasks has been reported more often in great apes than in monkeys [21, 22, 24]. Great apes have larger information-processing capacity than monkeys [83] and might be able to deploy meta-representational strategies more easily. In general, the resource-rational perspective presented here predicts that signatures of meta-representational mindreading will be found more often in species with higher information-processing capacity, even among closely related species. Systematic tests of this prediction are a fruitful avenue for future research.

4.1 The logic of factive mindreading

Factive mindreading saves computational resources by exploiting the substantial overlap that exists between a mindreader's world model and that of other agents in the same environment. In virtue of living in a shared world, individuals tend to have similar world models. This overlap allows a mindreader to predict another agent's behavior by keeping track of which facts in the mindreader's own world model are also in the other agent's world model. Our simulations confirm that this overlap between individuals' world models is key to the emergence of factive mindreading. First, factive mindreading did not emerge in social ecologies where individuals often have false beliefs (which diverge from the mindreader's world model). Second, factive mindreading did not emerge when the mindreader cannot use its own world model for the purpose of social prediction.

Our perspective gives a principled explanation for why false belief tasks are difficult. Namely, resource-rational policies for mindreading are optimized for good performance in situations that occur frequently, and cases of deception are relatively infrequent. Therefore factive mindreaders use ecologically valid cues (like perceptual access) to track what an agent knows, but do not explicitly represent beliefs. This perspective makes a surprising prediction: mindreaders might find it hard to predict the behavior of an agent with a *true belief* if that true belief does not meet the input conditions to be represented as knowledge [1, 35, 40]. In our simulations, this situation arises in experiments that involve accidentally true beliefs, mirroring the structure of 'Gettier' cases in epistemology [43]. In these experiments, the actor sees an item placed in a box and then goes away. The item is then taken out of the box, but is subsequently put back in the same box. A factive mindreader observing this situation will initially represent the actor as knowing the item's location. When the item is removed from its location outside of the actor's awareness, the mindreader discards this knowledge attribution and represents the actor as ignorant. Once the knowledge attribution is discarded, there is no way to re-create it when the item is put back in its original location. Therefore, the factive mindreader fails to reliably predict that the actor will go toward the item's actual location.

Crucially, experiments with this exact structure have been conducted in non-human primates, revealing that chimpanzees and rhesus macaques fail to reliably predict the behavior of individuals with an accidentally true belief, even though they succeed in closely matched situations where the individual has actual knowledge [20, 71]. In sum, non-human primates in these experiments perform like the factive mindreaders in our simulations. We find other points of convergence between non-human primates and our factive mindreaders: for example they find each outcome equally surprising in a false-belief task, including seeing the agent go toward a box where the item was never located [38].

4.2 Competence and performance

We explain the difficulty of belief attribution tasks from an abstract normative perspective. This ultimate-level explanation [84] is compatible with different explanations at the proximate level [3, 80, 85]. On one hand, belief attribution might be difficult because of a *competence* problem: a participant might fail a task because they do not have the capacity to represent beliefs. On the other hand, *performance* issues, like difficulty understanding a question, cognitive load, or lack of ecological validity, might mask the participant's competence. Supporting the performance interpretation, human infants and non-human primates can pass false belief tasks under some conditions [3, 17, 21, 22, 86], although the interpretation of these results is debated [3, 21, 87–90].

From our perspective, the interesting phenomenon is that belief attribution is in general more difficult than knowledge attribution: performance factors that might interfere with belief attribution do not seem to interfere with knowledge attribution to nearly the same extent [39, 40]. This pattern is consistent with our claim that knowledge attribution tasks can be passed while deploying fewer cognitive resources. A mindreader that is in principle capable of belief attribution might sometimes default to a simpler factive strategy depending on the amount of cognitive resources they currently have at their disposal [91]. This might happen when task demands (e.g. the need to interpret a complex verbal question) limit the resources that can be attributed to mindreading proper.

4.3 Hybrid strategies for mindreading

Our results are also relevant to understanding mindreading in human adults. Consider that high-resource observers in our model often use a hybrid strategy that blends metarepresentation and factive representation. They represent knowledge by default, and only encode the content of the actor's belief when a knowledge representation would not allow them to accurately predict behavior (cases of false or accidentally true belief). Human adults might use a similar strategy. Supporting this hypothesis, people can engage in mental state inference in contexts like conversation that require quick and spontaneous mindreading [37], even though they find it difficult to compute beliefs in these same contexts [92]. Convergent evidence comes from tasks that require people to make explicit knowledge or belief judgments [93, 94]. When people are asked what an agent knows, they respond either as fast or *faster* than when they are asked what the agent believes [93, 94]. This is the reverse of what one would predict if human adults used a purely meta-representational strategy (under this hypothesis, people would judge whether an agent knows something by first computing the agent's belief, and then assessing whether this belief matches reality). Similarly, neural activity in knowledge attribution tasks does not exhibit the signatures of inhibitory processing found in belief attribution tasks [93, 95].

4.4 Limitations and directions for future research

Because we set out to explain qualitative empirical patterns in mindreading from normative principles, we kept our model as simple as possible. So, by design we made very few assumptions about details of cognitive architecture, and used the abstract framework of information theory to operationalize cognitive costs [62]. Our model leaves out many factors that might influence performance in mindreading tasks, like altercentric biases in infancy or difficulty understanding questions in childhood [80, 96]. As a predictable consequence, there are empirical regularities that our model does not capture. For example young children tend to fail false belief tasks differently than our model: they predict that the agent will look for the item in its actual location [16]. Factive observers in our model have a slight bias toward the actual location of the item, but this bias is probably too small to fully account for the magnitude of this effect in young children. In our simulations, factive mindreading is prevalent in ecologies where ignorance is relatively common, and ignorant agents often do not reach for the actual location of the item; as a result a very strong actuality bias is not adaptive in these ecologies.

Future research should extend our approach to a broader range of tasks. Our analysis focuses on a classic experimental paradigm where false beliefs are caused by unwitnessed changes in location. Other situations can create challenges for mindreading: false beliefs can for instance be caused by mis-perception. Mindreading can also involve inferring the goals and desires of an agent [4, 9]. A promising direction for future research is to apply our framework to goal inference tasks, and explore the connections between our approach and existing resource-rational models of goal inference [97–99]. While we focus on social prediction, mindreading can also be deployed with the goal of influencing the behavior and mental states of other agents [100, 101].

Finally, while we focus on the distinction between knowledge and belief representation, the general principle we identify here is much broader. On our account, mindreaders save computational resources by representing some parts of their own world model as being shared by another agent. In the setting we use here these parts of the world model are facts about the world, but in principle they can be other things, such as concepts. For instance Alice might assume that Bob's concept of APPLE is the same as her own, instead of creating a meta-representation of Bob's concept of APPLE. We suspect that much of social cognition relies on such strategies, and that meta-representation is the exception rather than the norm.

References

[1] Martin A, Santos LR. What cognitive representations support primate theory of mind? Trends in cognitive sciences. 2016;20(5):375-82.

[2] Premack D, Woodruff G. Does the chimpanzee have a theory of mind? Behavioral and brain sciences. 1978;1(4):515-26.

[3] Baillargeon R, Scott RM, Bian L. Psychological reasoning in infancy. Annual review of psychology. 2016;67(1):159-86.

[4] Gergely G, Csibra G. Teleological reasoning in infancy: The naıve theory of rational action. Trends in cognitive sciences. 2003;7(7):287-92.

[5] Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in ”theory of mind”. NeuroImage. 2003;19(4):1835-42.

[6] Richardson H, Lisandrelli G, Riobueno-Naylor A, Saxe R. Development of the social brain from age three to twelve years. Nature communications. 2018;9(1):1027.

[7] Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour. 2017;1(4):0064.

[8] Quillien T, German TC. A simple definition of ‘intentionally’. Cognition. 2021;214:104806.

[9] Lucas CG, Griffiths TL, Xu F, Fawcett C, Gopnik A, Kushnir T, et al. The child as econometrician: A rational model of preference understanding in children. PloS one. 2014;9(3):e92160.

[10] Jara-Ettinger J. Theory of mind as inverse reinforcement learning. Current Opinion in Behavioral Sciences. 2019;29:105-10.

[11] Marr D. Vision: A computational investigation into the human representation and processing of visual information. MIT press; 1982.

[12] Anderson JR. The Adaptive Character of Thought. Psychology Press; 1990.

[13] Cosmides L, Tooby J. Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. Cognition. 1994;50(1-3):41-77.

[14] Jern A, Lucas CG, Kemp C. People learn other people's preferences through inverse decision-making. Cognition. 2017;168:46-64.

[15] Rakoczy H. Foundations of theory of mind and its development in early childhood. Nature Reviews Psychology. 2022;1(4):223-35.

[16] Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition. 1983;13(1):103-28.

[17] Onishi KH, Baillargeon R. Do 15-month-old infants understand false beliefs? Science. 2005;308(5719):255-8.

[18] Marticorena DC, Ruiz AM, Mukerji C, Goddu A, Santos LR. Monkeys represent others' knowledge but not their beliefs. Developmental science. 2011;14(6):1406-16.

[19] Martin A, Santos LR. The origins of belief representation: Monkeys fail to automatically represent others' beliefs. Cognition. 2014;130(3):300-8.

[20] Kaminski J, Call J, Tomasello M. Chimpanzees know what others know, but not what they believe. Cognition. 2008;109(2):224-34.

[21] Krupenye C, Kano F, Hirata S, Call J, Tomasello M. Great apes anticipate that other individuals will act according to false beliefs. Science. 2016;354(6308):110-4.

[22] Kano F, Krupenye C, Hirata S, Tomonaga M, Call J. Great apes use self-experience to anticipate an agent's action in a false-belief test. Proceedings of the National Academy of Sciences. 2019;116(42):20904-9.

[23] Lurz RW, Krachun C, Mareno MC, Hopkins WD. Do chimpanzees predict others' behavior by simulating their beliefs? Animal Behavior and Cognition. 2022;9(2):153-75.

[24] Padberg M, Hanus D, Haun D. Great apes show altercentric influences when confronted with conflicting beliefs. Animal Behaviour. 2025;227:123304.

[25] O'Laughlin C, Thagard P. Autism and coherence: A computational model. Mind & Language. 2000;15(4):375-92.

[26] Berthiaume VG, Shultz TR, Onishi KH. A constructivist connectionist model of transitions on false-belief tasks. Cognition. 2013;126(3):441-58.

[27] Goodman ND, Baker CL, Bonawitz EB, Mansinghka VK, Gopnik A, Wellman H, et al. Intuitive theories of mind: A rational approach to false belief. In: Proceedings of the twenty-eighth annual conference of the cognitive science society. vol. 6. Cognitive Science Society Vancouver; 2006. .

[28] Wang L, Hemmer P, Leslie AM. A Bayesian framework for the development of belief-desire reasoning: Estimating inhibitory power. Psychonomic Bulletin & Review. 2019;26(1):205-21.

[29] Berke MD, Horschler DJ, Royka A, Santos LR, Jara-Ettinger J. What Primates Know About Other Minds and When They Use It: A Computational Approach to Comparative Theory of Mind. bioRxiv. 2025. Available from: https://www.biorxiv.org/content/early/2025/09/10/2023.08.02.551487.

[30] Lieder F, Griffiths TL. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and brain sciences. 2020;43:e1.

[31] Gershman SJ, Horvitz EJ, Tenenbaum JB. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. Science. 2015;349(6245):273-8.

[32] Lewis RL, Howes A, Singh S. Computational rationality: Linking mechanism and behavior through bounded utility maximization. Topics in cognitive science. 2014;6(2):279-311.

[33] Leslie AM. Pretense and representation: The origins of ”theory of mind.”. Psychological review. 1987;94(4):412.

[34] Sperber D. Metarepresentations: A multidisciplinary perspective. Oxford University Press; 2000.

[35] Nagel J. Factive and nonfactive mental state attribution. Mind & Language. 2017;32(5):525-44.

[36] Phillips J, Norby A. Factive theory of mind. Mind & Language. 2021;36(1):3-26.

[37] Westra E, Nagel J. Mindreading in conversation. Cognition. 2021;210:104618.

[38] Royka A, Horschler DJ, Bargmann W, Santos LR. Exploring the evolutionary roots of theory of mind: Primate errors on false belief tasks reveal representational limits. Cognition. 2026;270:106400.

[39] Dungan J, Saxe R. Matched false-belief performance during verbal and nonverbal interference. Cognitive science. 2012;36(6):1148-56.

[40] Phillips J, Buckwalter W, Cushman F, Friedman O, Martin A, Turri J, et al. Knowledge before belief. Behavioral and Brain Sciences. 2021;44:e140.

[41] Pöppel J, Kopp S. Satisficing mentalizing: Bayesian models of theory of mind reasoning in scenarios with different uncertainties. arXiv preprint arXiv:190910419. 2019.

[42] Williamson T. Knowledge and its Limits. Oxford University Press; 2002.

[43] Gettier E. Is justified true belief knowledge? Analysis. 1963.

[44] Icard T. Bayes, bounds, and rational analysis. Philosophy of Science. 2018;85(1):79-101.

[45] Sims CR. Rate–distortion theory and human perception. Cognition. 2016;152:181-98.

[46] Wei XX, Stocker AA. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. Nature neuroscience. 2015;18(10):1509-17.

[47] Sims CR, Jacobs RA, Knill DC. An ideal observer analysis of visual working memory. Psychological review. 2012;119(4):807.

[48] Gershman SJ. The rational analysis of memory. In: Oxford handbook of human memory. Oxford University Press Oxford, UK; 2021. .

[49] Futrell R. Information-theoretic principles in incremental language production. Proceedings of the National Academy of Sciences. 2023;120(39):e2220593120.

[50] Zaslavsky N, Hu J, Levy RP. A rate-distortion view of human pragmatic reasoning. arXiv preprint arXiv:200506641. 2020.

[51] Taylor-Davies M, Lucas CG. Balancing utility and cognitive cost in social representation. arXiv preprint arXiv:231004852. 2023.

[52] Taylor-Davies M, Quillien T. An information bottleneck view of social stereotype use. In: Proceedings of the cognitive science society; 2025. .

[53] Sims CA. Implications of rational inattention. Journal of monetary Economics. 2003;50(3):665-90.

[54] Polanía R, Woodford M, Ruff CC. Efficient coding of subjective value. Nature neuroscience. 2019;22(1):134-42.

[55] Binz M, Schulz E. Modeling human exploration through resource-rational reinforcement learning. Advances in neural information processing systems. 2022;35:31755-68.

[56] Lai L, Gershman SJ. Human decision making balances reward maximization and policy compression. PLOS Computational Biology. 2024 04;20:1-32. Available from: https://doi.org/10.1371/journal.pcbi.1012057.

[57] Ortega PA, Braun DA. Thermodynamics as a theory of decision-making with information-processing costs. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013;469(2153):20120683. Available from: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2012.0683.

[58] Arumugam D, Ho MK, Goodman ND, Van Roy B. Bayesian Reinforcement Learning With Limited Cognitive Load. Open Mind. 2024 04;8:395-438. Available from: https://doi.org/10.1162/opmi_a_00132.

[59] Cheyette SJ, Wu S, Piantadosi ST. Limited information-processing capacity in vision explains number psychophysics. Psychological Review. 2024.

[60] Icard T, Goodman ND. A Resource-Rational Approach to the Causal Frame Problem. In: Proceedings of the cognitive science society; 2015. .

[61] Kinney DB, Lombrozo T. Building Compressed Causal Models of the World. Cognitive Psychology. 2023.

[62] Tishby N, Pereira FC, Bialek W. The information bottleneck method. arXiv preprint physics/0004057. 1999.

[63] Berger T. Rate-distortion theory. Wiley Encyclopedia of Telecommunications. 2003.

[64] Blahut R. Computation of channel capacity and rate-distortion functions. IEEE transactions on Information Theory. 1972;18(4):460-73.

[65] Arimoto S. An algorithm for computing the capacity of arbitrary discrete memoryless channels. IEEE Transactions on Information Theory. 1972;18(1):14-20.

[66] Gondek D, Hofmann T. Conditional information bottleneck clustering. In: 3rd ieee international conference on data mining, workshop on clustering large data sets; 2003. p. 36-42.

[67] Zaslavsky N, Kemp C, Regier T, Tishby N. Efficient compression in color naming and its evolution. Proceedings of the National Academy of Sciences. 2018;115(31):7937-42.

[68] Todd PM, Gigerenzer G. Ecological rationality: Intelligence in the world. OUP USA; 2012.

[69] Simon HA. A behavioral model of rational choice. The quarterly journal of economics. 1955:99-118.

[70] Pillow BH. Early understanding of perception as a source of knowledge. Journal of experimental child psychology. 1989;47(1):116-29.

[71] Horschler DJ, Santos LR, MacLean EL. Do non-human primates really represent others' ignorance? A test of the awareness relations hypothesis. Cognition. 2019;190:72-80.

[72] Krachun C, Carpenter M, Call J, Tomasello M. A competitive nonverbal false belief task for children and apes. Developmental science. 2009;12(4):521-35.

[73] Luo Y, Johnson SC. Recognizing the role of perception in action at 6 months. Developmental science. 2009;12(1):142-9.

[74] Hare B, Call J, Agnetta B, Tomasello M. Chimpanzees know what conspecifics do and do not see. Animal Behaviour. 2000;59(4):771-85.

[75] Townrow L, Krupenye C. Bonobos point more for ignorant than knowledgeable social partners. Proceedings of the National Academy of Sciences. 2025;122(6).

[76] Friedman O, Petrashek AR. Children do not follow the rule "ignorance means getting it wrong". Journal of Experimental Child Psychology. 2009;102(1):114-21.

[77] Chen Y, Su Y, Wang Y. Young children use the "ignorance= getting it wrong" rule when predicting behavior. Cognitive Development. 2015;35:79-91.

[78] Fabricius WV, Boyer TW, Weimer AA, Carroll K. True or false: Do 5-year-olds understand belief? Developmental Psychology. 2010;46(6):1402.

[79] Fabricius WV, Gonzales CR, Pesch A, Weimer AA, Pugliese J, Carroll K, et al. Perceptual access reasoning (PAR) in developing a representational theory of mind. Monographs of the Society for Research in Child Development. 2021;86(3):7-154.

[80] Oktay-Gür N, Rakoczy H. Children's difficulty with true belief tasks: Competence deficit or performance problem? Cognition. 2017;166:28-41.

[81] Hare B, Call J, Tomasello M. Do chimpanzees know what conspecifics know? Animal behaviour. 2001;61(1):139-51.

[82] Cantlon JF, Piantadosi ST. Uniquely human intelligence arose from expanded information capacity. Nature Reviews Psychology. 2024;3(4):275-93.

[83] Roth G, Dicke U. Evolution of the brain and intelligence in primates. Progress in brain research. 2012;195:413-30.

[84] Mayr E. Cause and effect in biology: kinds of causes, predictability, and teleology are viewed by a practicing biologist. Science. 1961;134(3489):1501-6.

[85] Perner J, Leekam SR, Wimmer H. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. British journal of developmental psychology. 1987;5(2):125-37.

[86] Buttelmann D, Buttelmann F, Carpenter M, Call J, Tomasello M. Great apes distinguish true from false beliefs in an interactive helping task. PloS one. 2017;12(4):e0173793.

[87] Butterfill SA, Apperly IA. How to construct a minimal theory of mind. Mind & Language. 2013;28(5):606-37.

[88] Heyes C. Submentalizing: I am not really reading your mind. Perspectives on Psychological Science. 2014;9(2):131-43.

[89] Burge T. Do infants and nonhuman animals attribute mental states? Psychological Review. 2018;125(3):409.

[90] Horschler DJ, MacLean EL, Santos LR. Do non-human primates really represent others' beliefs? Trends in Cognitive Sciences. 2020;24(8):594-605.

[91] Lieder F, Griffiths TL. Strategy selection as rational metareasoning. Psychological review. 2017;124(6):762.

[92] Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. Cognition. 2003;89(1):25-41.

[93] Bricker AM. The neural and cognitive mechanisms of knowledge attribution: An EEG study. Cognition. 2020;203:104412.

[94] Phillips J, Knobe J, Strickland B, Armary P, Cushman F. Evidence for evaluations of knowledge prior to belief. In: Proceedings of the cognitive science society; 2018. .

[95] Gonzalez B, Armary P, Dungan J, Strickland B, Knobe J, Cushman F, et al. Knowledge without belief. 2025. Available from: https://osf.io/preprints/psyarxiv/ht65f_v2.

[96] Manea V, Kampis D, Grosse Wiesmann C, Revencu B, Southgate V. An initial but receding altercentric bias in preverbal infants' memory. Proceedings of the Royal Society B. 2023;290(2000):20230738.

[97] Blokpoel M, Kwisthout J, van der Weide TP, Wareham T, van Rooij I. A computational-level explanation of the speed of goal inference. Journal of Mathematical Psychology. 2013;57(3-4):117-33.

[98] Chandra K, Chen T, Li TM, Ragan-Kelley J, Tenenbaum J. Inferring the future by imagining the past. Advances in Neural Information Processing Systems. 2023;36:21196-216.

[99] Zhi-Xuan T, Kang G, Mansinghka V, Tenenbaum JB. Infinite Ends from Finite Samples: Open-Ended Goal Inference as Top-Down Bayesian Filtering of Bottom-Up Proposals. Proceedings of the Annual Meeting of the Cognitive Science Society. 2024 Jul;46(46).

[100] Ho MK, Saxe R, Cushman F. Planning with theory of mind. Trends in Cognitive Sciences. 2022;26(11):959-71.

[101] Sell A, Sznycer D, Al-Shawaf L, Lim J, Krauss A, Feldman A, et al. The grammar of anger: Mapping the computational architecture of a recalibrational emotion. Cognition. 2017;168:110-28.