How Is Mindreading Possible: A Contextual Analysis of Theories of Mind

Xuning Gao

Abstract: The question of how we successfully interpret and predict other people's minds without having access to them—commonly referred to as "mindreading"—has fueled the ongoing debate between theory-theory and simulation theory. Theory-theory argues that mindreading relies on folk psychological frameworks, while simulation theory suggests it is based on imaginative projection. Yet, the debate is intractable and this impasse has paved the way for pluralism—the idea that multiple theories can be valid, each offering accurate depictions of people's mental states in different contexts. This paper advances a pluralist perspective, proposing that the applicability of each theory depends on two key factors: the familiarity of mindreaders—individuals who interpret others' minds without direct access—with the person being predicted and with the situation in which mindreading occurs. By categorizing mindreading scenarios into four types based on these two dimensions, I demonstrate how theory-theory and simulation theory can be contextually integrated. Three of the four scenarios allow successful mindreading, where individuals rely on either theory-theory, simulation theory, or a combination of both, depending on the strengths and limitations of each approach. In the fourth scenario, mindreading is not possible due to insufficient information. This categorical analysis shows that neither mindreading theory is universally applicable, but both play crucial roles in different contexts. By emphasizing the specific contextual applicability of each theory, this paper offers a more nuanced understanding of mindreading under a pluralism account, bridges the gap between the two competing theories, and provides a framework that recognizes the strengths of both approaches in predicting mental states.

Introduction

How do we successfully interpret and predict other people's thoughts and emotions when the only mind we truly access is our own? Although this task of mindreading seems logically unachievable, the constant accuracy of predictions about others' minds plays a crucial role in facilitating our daily interactions. Various theories have been developed to explain our success in mindreading, with simulation theory and theory-theory being the most well-regarded theories. Theory-theory argues that we rely on folk psychological theories to infer others' mental states. Folk psychological theories here refer to theories that people generate using concepts like 'desire" and "belief," which interpret, predict, and explain the behaviors or feelings of others. Simulation theory, in contrast, claims that we understand others by imaginatively projecting ourselves into their situations, simulating how others would feel or act based on our own potential reaction.

The proponents of the two theories remain fundamentally at odds with each other, leading to the rise of a new pluralist view on mindreading. Pluralists argue that multiple theories of mindreading can coexist, each suited to a different context in diverse human interactions. Advancing this pluralistic perspective, this paper examines the applicability of theory-theory and simulation theory, proposing that their use depends on two dimensions: the mindreader's familiarity with the individual being interpreted and the context of the situation.

In this paper, I first introduce the debate between these two dominant theories to provide the context for my view, outline the framework in which we are situated, and raise the question we seek to answer. Then, I present the pluralist perspective in this debate and provide evidence supporting the possibility of cooperation between the two theories. Next, I explore different categories and explain which theories should be applied in specific thought experiments. Finally, I address potential objections and offer responses.

Theory-Theory and Simulation Theory at a Glance

The debate between theory-theory and simulation theory centers on the question: "What sort of process subserves our capacity to predict, explain, and interpret other people?"[1] Theory-theorists and simulation theorists offer distinct perspectives on this issue. One of the most widely discussed arguments of theory-theory involves Alison Gopnik's examination of the false-belief task, which demonstrates how children's ability to predict others' beliefs evolves between ages three and five is based on their attainment of folk psychological theories.[2] For simulationists, philosophers like Goldman and Gordon have received considerable attention, proposing that people predict others' minds and behaviors by imaginatively "putting themselves in the other's shoes."[3] In this section, I introduce the basic ideas of mainstream theory-theorists and simulation theorists with the help of the false-belief experiment.

Theory-theory (mainly by Gopnik, Wellman, Stich, and Nichols) in a nutshell is the idea that people utilize folk psychological theories to predict people's mental states. In their 1994 article, Gopnik and Wellman compare children's generation of folk psychological theories with scientists' development of scientific theories (i.e. the atomic theory), saying that "scientific theory change and conceptual change in childhood are both the product of human minds trying to understand the world around them" even though children's theories are much less formally and rigorously tested.[4] In other words, people attain folk psychological theories by generalizing their past experiences into straightforward formulas such as "If x then y."[5] Gopnik and Wellman then extend a theory-theorist perspective to explain the false-belief task.[6] Before further illustrating the false-belief experiment from the theory-theorists' perspective, I will first outline the relevant experiments related to desires, beliefs, and predictions in their original form.

To start with a desire-based task by Woolley and Wellman—a simpler version of the false-belief task featuring only desire—children aged two, three, and five are told a story about a person and asked to predict what the person would do next, examining their understanding of the relationship between desires and human actions.[7] For example, the children are told that "Tom wants to drink water" and are asked to predict Tom's subsequent actions. Across all ages, most children understand the desire-action relationship and predict that Tom will go drink water. The task becomes more challenging when beliefs are introduced. In a study by Karen Bartsch and Henry Wellman, children aged three and four were presented with a scenario involving a person's simple action and asked to

---

1 Shaun Nichols and Stephen P. Stich, "Folk Psychology: Simulation or Tacit Theory?" *Philosophical Issues* 3 (1993): 240.

2 Gopnik, Alison, and Henry M. Wellman. "The Theory Theory." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, edited by Lawrence A. Hirschfeld and Susan A. Gelman (Cambridge University Press, 1994).

3 Robert M. Gordon, "Folk Psychology as Simulation," *Mind & Language* 1, no. 2 (1986): 162.

4 Gopnik and Wellman, "The Theory Theory," 258.

5 Note that one of the reasons why folk psychological theories are classified as "theories" is that they are generalized and may apply not only to an individual, but to a group of people.

6 Gopnik begins by comparing scientific theory change in adults with conceptual change in children. She argues that, despite differences in cognitive ability and methodological rigor, these types of change share "deep similarities between the underlying cognitive mechanisms involved in the epistemological endeavors of childhood and of science" (Gopnik and Wellman, "The Theory Theory," 259). According to Gopnik, key features of scientific theories—such as falsifiability, the inclusion of abstract entities, causal explanations, systematic coherence, and more—should also apply to children's understanding of the mind. In this view, children's predictive abilities arise from their possession of folk psychology theories. Children *with* differing theories are likely to interpret evidence distinctively, leading them to different conclusions about others' mental states.

7 Henry M. Wellman and Jacqueline D. Woolley, "From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology," *Cognition* 35, no. 3 (1990): 245–75.

interpret that action.[8] When asked, for example, "Why is Amy searching for her dog under the sofa instead of outside the house?" both three- and four-year-olds generally respond with explanations hybrid with belief terms and desire terms, such as "Amy *wants* her dog and *thinks* it is under the sofa," with four-year-olds demonstrating better accuracy.[9] Wellman and Woolley's 1990 study further explores children's recognition of belief and desire by performing similar experiments featuring either belief or desire performances on two-year-olds. The results suggest the concept of belief is not fully accessible to two-year-olds, and that "many two-year-olds pass desire tasks but fail comparable belief tasks."[10] Between the ages of two and four, children's understanding of the concept of belief appears to grow in direct proportion to their age.

The experiment that compares three-year-olds' and five-year-olds' understanding of the concept of belief, which is when the false belief task comes into play, is much more intriguing: "Children may see a candy box that has pencils instead of candy inside. They are asked what someone else [who just walks into the room] will think is inside the box. Three-year-olds typically say "pencils." They behave as if there is a simple and reliable causal link between the real state of affairs in the world, and our mental states about it."[11] While three-year-olds can talk about beliefs, they struggle to accurately predict others' beliefs as shown in the experiment. By contrast, five-year-olds successfully answer "candy" instead of "pencil," demonstrating a more sophisticated understanding of beliefs as different representations of the world instead of direct perceptions. The false-belief task further illustrates the direct correlation between maturation and the understanding of others' minds.

Theory-theorists explain this correlation by linking children's development of understanding to shifts in their acquisition of folk psychological theories at different ages. They argue that two-year-olds, who lack a concept of belief, operate on a desire-perception theory: "If an agent desires x, and sees that x exists, he will do things to get x."[12] By applying this folk psychological theory, two-year-olds can infer that Tom will drink from a water fountain if he is thirsty and perceives one.

By the age of three, children's responses shift to incorporate beliefs, marking a theoretical evolution. At this stage, children begin to understand fictional mental states such as beliefs, dreams, and images, and they form a new theory within the desire-perception framework. This new theory, still heavily influenced by the simple desire-perception view, remains nonrepresentational—just as one's perception has a direct causal link with reality, three-year-olds also think there exist direct causal connections between beliefs and real-world objects.[13] For instance, in the false-belief task where children are shown a candy box that contains a pencil, three-year-olds mostly conclude that someone

8 Karen Bartsch and Henry Wellman, "Young Children's Attribution of Action to Beliefs and Desires," *Child Development* 60, no. 4 (1989): 946–64.
9 Bartsch and Wellman, "Young Children's Attribution."
10 Wellman and Woolley, "From Simple Desires to Ordinary Beliefs," 267.
11 Gopnik and Wellman, "The Theory Theory," 266.
12 Gopnik and Wellman, "The Theory Theory," 265.
13 Gopnik and Wellman, "The Theory Theory," 265.

seeing a candy box would think it contains a pencil due to their possession of the theory that there is a direct causal link between tangible reality (what is actually inside the candy box) and belief (thought of what is inside the candy box).

By age five, children understand beliefs as representations, explaining their success in false-belief tasks. They realize that people believe what looks like the truth to them rather than being directly aware of the truth of the world. Such a realization marks a shift in children's minds toward representational beliefs similar to those of adults. This shift also transforms their non-representational theory of belief—which assumes a direct causal connection between belief and reality—into a representational theory. In the candy-pencil false-belief task, five-year-olds tend to respond with "candy" instead of "pencil," indicating their grasp of representational mental states. A representational theory for five-year-olds might take the form: "There exists an object P, which appears as p1 from an outside perspective. Person A observes from the outside and thus perceives P as p1." Providing a coherent explanation for the false-belief task, theory-theorists suggest that children's predictive responses evolve due to the development of their folk psychology theories, and people make predictions through applying folk psychology theories.

Nonetheless, theory-theory is "not the only game in town."[14] Simulation theorists like Goldman and Gordon argue that prediction does not require such folk psychological theories—people do not need to generalize their experience into theories to predict or explain others' behavior. Gordon claims the decisions that people make are "often products of practical reasoning,"[15] or in Goldman's words "decision-making mechanism,"[16] without clarifying the composition of such a mechanism. Since simulation theorists presuppose that practical reasoning is the underlying basis for all individuals' decision-making, people generate similar actions when having similar beliefs and desires. Thus, the predictors imagine being in others' positions and use their practical reasoning to predict or explain the predicted person's behaviors.

A simple way of simulating a person is "to answer the question, "What would I do in that person's situation?""[17] When we are trying to predict what we will do in an imaginary scenario, we put ourselves in a pretend world with pretend intentions, which allows our practical reasoning (or decision-making mechanism) to work offline (i.e., working without making actual decisions in real life) and predict what we will do if such an imaginary event happens. For instance, if I want to know what I would do if my cat went missing, I put myself into a modified world where I have lost my cat. My practical reasoning will then tell me that, in such a possible world, I would want to call my friend to come over and help me look for the cat in the neighborhood, which is a prediction of my own mental state if I had actually lost my cat. In self-prediction, we modify certain parameters while keeping others constant. In the cat example, we adjust our desires—the

14 Nichols and Stich, "Folk Psychology," 239.
15 Gordon, "Folk Psychology as Simulation," 160.
16 Alvin I. Goldman, "Conceptualizing Simulation Theory," in *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (Oxford University Press, 2006), 27.
17 Gordon, "Folk Psychology as Simulation," 161.

intention to find my lost cat—while retaining our beliefs, like that "my friends can help me." Similarly, when predicting others' actions, we engage in a process of "quarantining [our] own idiosyncratic desires and beliefs"[18] to fully adopt their perspective, which involves putting ourselves in their shoes and incorporating their specific desires and beliefs rather than our own.[19] If I want to predict what Marielle will think or do after she receives a sapphire necklace as a birthday present, I would imagine that I am Marielle and have her beliefs and desires. I know that Marielle loves the color blue and jewelry, but she is not aware that I am giving her a present on her birthday. Pretending to have such desires and beliefs, I imagine that I would feel very surprised and happy when seeing the birthday gift in this pretend world, which are the emotions I attribute to Marielle. Simulation theorists claim people do not predict through a rigorous, systematic format in which we strictly follow explicit rules that generalize the behaviors of a large group of people, rather, their predictions are individually focused and shaped by the predictors' past experience with the predicted. To return to the false-belief experiment, typical simulation theorists would explain the false-belief task by saying that the change from failure at the age of three to success at the age of five is due to the completeness of the mental ability. Children at the age of three simply do not have the mental capacity to simulate the other's perspective fully. While at the age of five, their capacity develops and allows them to simulate a pretend world from the other's perspective, which is why they answer "candy" instead of "pencil."[20]

To end this section, a helpful analogy can be made to compare the two theories. Stich and Nichols compare an engineer predicting how a "plane will behave at a certain speed"[21] to a person mindreading how an individual will behave or feel in a situation. There are two ways of figuring out how a plane will perform after taking off. The first way is to use physical and mathematical functions to calculate the precise data points, and the other is to imagine the plane taking off based on the engineer's experience. The first approach aligns with theory-theory, that people utilize theories that work for all to make predictions, while the second aligns with simulation theory, which asserts that people predict by envisioning themselves in similar situations as the predicted individuals. These two theories seem to be fundamentally at odds with each other. Now the question is, can we reconcile these two seemingly opposing theories of mind? If so, how?

18 Goldman, "Conceptualizing Simulation Theory," 30.

19 I understand there are differences between Gordon's and Goldman's versions of simulation theory. However, this paper does not delve into these distinctions. as the focus is on utilizing the core principles of simulation theory and theory-theory to construct a pluralistic framework. Due to space constraints of this paper, Goldman and Gordon are treated here as proponents of the mainstream simulation theory.

20 In this paper, I focus on the arguments presented by mainstream theory-theorists and simulation theorists. While I acknowledge the existence of alternative perspectives, such as Jane Heal's co-cognition, which could offer grounds to reconsider my argument, I do not address these views due to the limited scope of this paper.

21 Nichols and Stich, "Folk Psychology," 229.

Distinguishing Ways of Predictions

Many philosophers adhere to a dichotomy, viewing theory-theory and simulation theory as incompatible frameworks such that only one of them is true. However, this approach has proven unproductive, as neither theory can definitively refute the other. Instead, the debate often results in each theory talking past the other, reinforcing their incompatibility rather than advancing the conversation. The failure to resolve the debate comes from an underlying assumption that one theory must dominate all mindreading contexts, while in fact, each theory has its own strengths and weaknesses. Pluralism offers a solution by recognizing that the applicability of each theory depends on context, allowing us to move beyond the dichotomy and embrace both approaches. Theory-theory is useful for drawing quick conclusions in familiar situations with generalized theories that apply to all, while simulation theory helps with getting a more specific, nuanced understanding of an individual's mental states. The complementary strengths of these theories demonstrate their collaborative potential, showing that mindreading is not a one-size-fits-all process, but rather a dynamic skill that adapts to various contexts.

Pluralists such as Jane Suilin Lavelle and Shannon Spaulding claim there are many ways to answer the question "How do we predict people's minds" depending on the context. These different ways are all valid interpretations of human behaviors because they are applied in multiple contexts for different purposes. Following pluralism, I propose that theory-theory and simulation theory can complement each other in context-specific ways. This section specifies the problem we are addressing, provides evidence for the compatibility of the theories, and analyzes how they can collaborate.

In addressing the ongoing debate between theory-theorists and simulation theorists, it is crucial to consider the scope in which mindreading is applied. As Lavelle points out in her book *Mindreading and Social Cognition*, traditional theorists of mind such as Stich and Nichols take mindreading to be ubiquitous, in a sense that we both "engage in mindreading for mundane chores . . . [and] in loftier endeavors like trying to glean Descartes's reasons for thinking that many ideas are innate."[22] Lavelle refers to this traditional view as "the ubiquity principle," which holds that mindreading is a constant, pervasive part of human interaction. Nonetheless, following the pluralistic perspective, I argue that it is superfluous and impractical for people to attribute mental states to everyone in every single daily interaction. Many interactions occur without conscious awareness of others' mental states, as these processes are so habitual that we seldom pause to analyze or think about them. When we are checking out at a convenience store, we do not bother to attribute mental states to the cashier. When the cashier says "19.99," we don't need to analyze their beliefs and desires, imagine ourselves in their position, or apply a theory to interpret their words and predict their intentions. Instead, we focus solely on the number we hear. Our cultural traditions, customs, and constant visits to supermarkets assured us that, in this context, the number the cashier states is simply the price we need to pay for our items. Since such scenarios happen constantly, we develop a mental shortcut that bypasses the need to consider the speaker's beliefs and desires to understand their intent. Instead, we respond automatically to the words we hear without

22 Shaun Nichols and Stephen P. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Clarendon Press, 2003), 2.

referring to the speaker's mental state. Therefore, our reading of mind does not happen in mundane scenarios, but rather only in the scenarios that catch us by surprise or force us to think about other people's minds. We do not mindread the cashier when they say, "How are you?" or "Your total is $19.99," but when they start to act in a way that is absurd within the context, for example, when they burst into tears or start praying, people will be confused and utilize their ability of mindreading. In brief, mindreading does not happen on a daily basis. It is not a mundane or ubiquitous activity—people only mindread when something unexpected takes place. Some might argue that mindreading now sounds more like explaining or interpreting behaviors rather than predicting their mental states in hypothetical scenarios. To clarify, the term "mindreading" in this paper does not merely refer to forecasting others' mental states and behaviors, it also encompasses the explanation of off-script behaviors by referencing the agents' mental states. Having such an interpretive role, mindreading allows us to provide reasoning behind people's behavior (i.e., the cashier dancing or crying), and use the reasoning as evidence of people's beliefs and desires for predictions in the future.

My argument here aligns with the *script theory* raised by Nelson and Gruendel, they introduced the concept of a *script* in their 1979 paper as a typical sequence of events within a specific context or routine such as "going to bed, attending a birthday party, eating at a restaurant, taking a bath, or having a telephone conversation."[23] Though their experiment investigates the scripted content of children's conversation, their findings can be applied more broadly, that a large part of people's daily interactions are scripted. Adults also endorse commonly acknowledged routines across multiple contexts that, according to Lavelle, shape our expectations and "limit the amount of information one needs to process in a particular situation."[24] Script theorists claim that the script provides a framework and guides what we should focus on in a given context and what we can ignore. Since a large amount of people's daily interactions with each other follow the script, they are able to find the corresponding meaning of each other's actions for each script, which saves them from extensive mindreading and attributing mental states to each other. When following a script, predictions become mechanical without referencing the mental states of others. Thus, unlike the traditional view that supports the ubiquity principle of mindreading, real-world interactions do not always require attribution of mental states to people we are interacting with. The debate between theory-theory and simulation theory arises only when events deviate from the expected script or are entirely devoid of a script. It is only when people go completely off-script that they are compelled to see them as intentional beings with feelings and mental states rather than as scripted characters who interact with us mechanically.

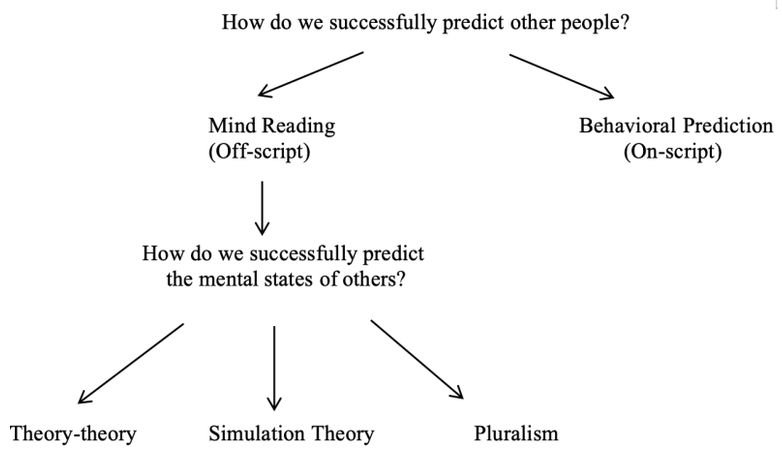
With this in mind, we can outline the problem at hand more clearly. There are two ways to answer the question, "How do we successfully predict other people?" The first is to predict according to the script and current actions and perform a behavioral prediction that only concerns their behavior. For example, we predict that the next move

23 Katherine Nelson and Janice M. Gruendel, "At Morning It's Lunchtime: A Scriptal View of Children's Dialogues," *Discourse Processes* 2, no. 2 (1979): 78.

24 Jane Suilin Lavelle, *Mindreading and Social Cognition* (Cambridge University Press, 2022), 36.

of the cashier will be to give me my receipt, because I asked for it. The second way is to perform a mental prediction, or mindreading. When people we are interacting with get off the script, we have no script to rely on and need mindreading to explain or predict their move (see *Diagram No. 1*). We think of other people's minds and mindread only when an off-script behavior takes place or when we are forced to think about other people's mental states, and the circumstances where people feel the necessity of ascribing mental states varies by person. The following discussions of the term "mindreading," therefore, implies the appearance of off-script behaviors.

The question for the new mindreading then is: "How do we successfully predict the mental state of others?" Several different answers can be considered as the answer to this question. There are the mainstream accounts—theory-theory and simulation theory, and there is the pluralist account—that we have different theories for different contexts and circumstances we are in. The distinction between mindreading and behavioral prediction clarified above is central to how we interpret the positions of theory-theorists and simulation theorists, as well as for generating thought experiments later on in this paper. In the following sections, I will challenge the traditional view and expand on pluralism to demonstrate how theory-theory and simulation theory can work together in a complementary way for different scenarios in mindreading.



(Diagram No. 1)

Many theorists of mind, even the traditional ones, have made the argument that theory-theory and simulation theory do not entirely contradict each other. Simulation theorist Goldman specifies in the footnote of his 2009 paper that his previous book *Simulating Minds,* "does not claim that all mindreading is executed by simulation. It leaves room for theory-based mind reading, and hence [his] overall approach is a simulation-theory hybrid."[25] Moreover, another simulation theorist Jane Heal offers a less ambitious version of simulation theory in her book *Mind, Reason and Imagination*. She points out that "[simulationists] do not yet have enough grip on how that project might

25 Alvin I. Goldman, "Mirroring, Simulating and Mindreading," *Mind & Language* 24, no. 2 (2009): 24.

actually work out in detail to have any confidence that we are working in terms of the right architecture."[26] This modest simulationist view acknowledges the limitations and complexity of the mindreading project. Even though they are significant defenders of simulation theory, neither Goldman nor Heal claims it is the only method people use to make predictions. Simulationists are not the only ones who take a step back on the battlefield, though the most prominent theory-theorists like Stich and Nichols are still holding the line. Peter Carruthers, another defender of theory-theory, sketched a limited view of simulation similar to Heal's version of simulation theory that a person can simulate the inferential role of a concept in others' minds relying on their grasp of the same concept. Carruthers claims that this limited simulation theory is something that "a theory-theorist should have no principled objection to."[27] His proposal allows the collaboration between the theory- and simulation theory that keeps the theoretical framework from theory-theory, and adds in advantages simulation has, such as "fine-grained predictions and explanations of thoughts, feelings, and actions of other people."[28]

Even though many traditional theorists of mind accepted the collaboration between the two theories, they still want to argue that their own theory is the dominant one instead of accepting the pluralist account that these two and more theories are equally explanatory for people's prediction behavior at different times for different purposes. To put it in Lavelle's words, they are still disagreeing upon the "how" question of mindreading, ignoring the other elements of the mindreading debate such as "what," "why," and "when." From a pluralist standpoint, the goal isn't solely to understand *how* people mindread. Instead, pluralists recognize the variations of theories among scenarios and cultures, emphasizing the value of exploring how mindreading varies in different contexts. Since the context determines which theory is applied, questions about "what," "why," and "when" we mindread are deeply connected to "how" we mindread. Pluralists view the two theories less as contradictory frameworks for explaining how people predict and more as different tools among a box of tools that individuals use for prediction in various circumstances. Although pluralism suggests that there is no single approach to mindreading and that individuals use different methods depending on the situation, pluralistic arguments focus more on various kinds of mindreadings while rarely addressing the specifics of when or where people apply either simulation theory or theory-theory—a gap that the following analysis aims to fill.

Collaborating Theories of Mind

Building on pluralism, I propose that simulation theory and theory-theory can function together not merely as one being a supplement for another, but are equally useful as distinct tools suited to different contexts due to their differing requirements for background knowledge for the target of mindreading. One major difference between the implications of the two theories is their acquisition of information. Theory-theory is often considered "knowledge-rich" because its application relies on using folk psychological theories to predict others' actions, which requires predictors to have the knowledge of theories. In contrast, simulation theory is viewed as "knowledge-poor," as it depends on imagining oneself in another's position, which requires minimal background knowledge about folk psychology. Just as the theories differ in being either knowledge rich or knowledge poor, they are more suited to certain types of scenarios corresponding to their features. Given their distinct characteristics, I argue that people apply different theories based on two contextual elements: (1) how familiar we are with the situation we're trying to predict, and (2) how well we know the person we are predicting. To give precise definitions, a person is familiar to me when I know the uniqueness of their beliefs and desires, especially the parts that are similar to or differ from generalized or stereotypical knowledge of a social group that they belong to. The word "situation" here implies specific contexts the predicted person is in, such as nationality, disease, and any factors external to their minds that affect our prediction. Nonetheless, it is important to keep in mind that since these situations all concern mindreading, they are all either off-script scenarios or circumstances in which people are forced to refer to other people's minds. For example, a familiar situation might be something that we have encountered several times, such as when a person receives his check and immediately shouts.[29] A severe example of an unfamiliar situation might be when a person discovers that her friend has secretly been campaigning for the president of their country. In the following paragraphs, I will evaluate four scenarios in detail, providing specific examples to support my argument that three of them qualify as mindreading scenarios, each involving the application of a different theory or theories. It is worth noticing that the examples discussed in this paper all focus on the mindreading of an individual's emotions and feelings for simplicity and consistency, while the argument can extend to all other mental states such as desires, beliefs, and decision-making.

(1) *Familiar Person with Unfamiliar Situation (FPUS)*

By mixing and matching these two dimensions, the four distinct scenarios in mindreading are: predicting a familiar person with an unfamiliar situation, an unfamiliar person with a familiar situation, a familiar person with a familiar situation, and an unfamiliar person with an unfamiliar situation (see *Graph No.2*). I argue that in each scenario, people tend to apply different theories of mindreading. To start with the first scenario, if a person has never encountered, seen, or heard of the situation they are in, it is by definition impossible for them to have a generalized folk-psychology theory and predict the mental states of others with it. Simulation theory is more useful here since it does not rely on pre-existing folk psychological theories. Instead, we simulate the mental

26 Jane Heal, *Mind, Reason and Imagination* (Cambridge University Press, 2003), 35.

27 Peter Carruthers, "Simulation and Self-Knowledge: A Defence of Theory-Theory," in *Theories of Theories of Mind*, edited by Peter Carruthers and Peter K. Smith (Cambridge University Press, 1996), 25.

28 Carruthers, "Simulation and Self-Knowledge," 25.

29 Note that a familiar situation happens when the attribution of mental states is necessary, yet the same situation has been repeatedly encountered in various ways. The event is not as mechanical as scripted behaviors but also not a brand new, extremely shocking circumstance as in unfamiliar situations.

state of a familiar person, using our past experiences of their reactions to predict future response. Moreover, as we are familiar with the predicted person, we are better equipped to attribute beliefs and desires to them and predict their mental states more accurately. This familiarity with the person allows us to refine our predictions by drawing on the historical tendency of their typical reactions or decision-making.

For example, if Amy is a five-year-old kid who wants to predict how her 30-year-old cousin Betty would feel when Betty's mom passes away, she tends to simulate Betty's mind instead of applying theories to her.[30] Suppose the "death of someone she knows" is a situation that Amy has never encountered in her life, and nor has anyone talked to her about the very notion of death. It is then impossible for her to develop a folk psychological theory in her mind about what death is and how it affects people's mental states. She lacks the developed theories that adults often have, such as those linking death to sadness, sorrow, and suffering. Betty tells Amy that after someone dies, you will never be able to meet them again. With no theory in this unfamiliar situation, Amy's only way to mindread Betty's feelings when her mother dies is by putting herself in Betty's shoes and imagining how she would feel if she could never see her own mother again. Amy knows that both she and Betty love their mothers deeply, and she can imagine the sadness and loneliness she would feel if her own mother passed away. By performing such a simulation with her knowledge of the situation Betty is in, Amy predicts by projecting her own emotions onto Betty that she might have a similar experience instead of utilizing a folk psychological theory such as, "If a person lost their mother, they tend to feel sad and cry." As a result, predicting a familiar person's reaction to an unfamiliar situation can be done without relying on a formal theory. Simulation theory brings us more successful predictions compared to theory-theory in the case of FPUS.

(2) *Unfamiliar Person with Familiar Situation (UPFS)*

When predicting an unfamiliar person in a familiar situation, theory-theory is most effective, as we apply general folk psychological theories based on prior experience with similar situations despite limited knowledge of the individual. Since we have encountered similar situations before, we have generalized folk psychological theories and are knowledge-rich in the context we are predicting. Yet, our limited knowledge of a stranger's mental states makes it difficult to accurately simulate their feelings or even build an imaginary world from their perspective. By drawing on general theories we've developed about different groups with specific traits, we can apply these insights to individuals within those groups and accurately mindread without needing to fully adopt their perspective. It is particularly useful when predicting the actions of strangers, whom we know little about and whose social group we can only infer from their appearances or behaviors.

For instance, if my friend asked me the question when we walked past an old European lady on the street, "What would the lady think after she sees a black cat walking past her?" I can barely relate to the old lady since we are strangers and I do not know

30 Death of a parent is an off-script event to Amy since it is not something that she can understand by mechanically interacting with her cousin. People relate to the relatives of the dead and naturally refer to their mental states.

her beliefs and desires at all.[31] It would be difficult for me to simulate and construct an imagined world from her perspective. Fortunately, I have a folk psychological theory that "If a European sees a black cat crossing their path, they may feel uneasy, viewing it as a sign of bad luck." I can apply this folk psychological theory to the old lady since she belongs to the social group mentioned in the theory, and predict her mental states without knowing her beliefs and desires. From this example, we can conclude that using theory-theory's approach to mind reading can be an effective and practical method when predicting the behavior of unfamiliar people in familiar situations. We use generalized theory to mindread when simulation is not available.

Critics might object that this scenario (UPFS) should not be considered as mindreading because theory-theory is less specific than simulation theory. Mindreading in this case is predicting the behavior of an entire social group rather than mindreading a particular individual, which diminishes both its particularity and accuracy. However, this concern is not warranted. While accuracy is certainly a key strength of any mindreading theory, a theory with relatively low accuracy does not mean it is rarely used in practice. Even simulation theory cannot guarantee one hundred percent accuracy in mindreading. In the UPFS scenario, theory-theory (predicting based on the typical mindset of a social group) is exactly what people rely on, and is the best option available. Without theory-theory, we would be left to guess the preferences and beliefs of an unknown individual, which introduces a vast array of possibilities and leads to no clear conclusions. Simulation theory cannot be effectively put into use in this scenario since we are not familiar with the predicted person. Therefore, while theory-theory may have comparatively lower accuracy than simulation theory in other scenarios, it is still the dominant (and perhaps the only) approach that enables our successful mindreading in this case.[32]

(3) *Familiar Person with Familiar Situation (FPFS)*

In familiar situations with familiar people, both theory-theory and simulation theory are at play. Familiar situations allow us to have a generalized theory while our familiarity with the person gives us prior knowledge about their unique beliefs and desires. Thus, we use general theories to quickly draw a conclusion while using simulation to make predictions more precisely about the individual we are predicting through our previous experience of their specific reactions and feelings, which explains why our predictions of familiar people and familiar cases tend to be more accurate.

For example, Amy from the first scenario is in her thirties and receives the message that her uncle, Betty's father, is very ill and about to pass away. Shaken by the shocking news, Amy tries to predict how Betty might feel about her father's passing. Since Amy is now a grown-up and is a lot more familiar with the situation of people she knows passing

31 This is not so much an off-script behavior, but it is a scenario in which I am forced to refer to other people's mental state because I am asked to do so. Therefore, it should be considered as a mindreading scenario.

32 The theories discussed and categorized in this paper do not guarantee success in mindreading; rather, they offer frameworks for how people engage in mindreading, without addressing the accuracy of the process.

away, she will quickly come up with a conclusion that Betty will be very sad according to the folk-psychology theory she generalized—"People generally feel sad when their parents pass away." However, Amy's prediction does not stop with the generalized theory as it did in the previous UPFS scenario. Instead, since Amy knows how much Betty loves her father and has constantly seen their interactions, Amy can further understand why Betty is sad and knows the objects and activities that would possibly trigger her sadness (i.e., crafts that Betty and her dad made together when she was young). When mindreading a familiar person, the prediction of feelings is much more nuanced and concise than mindreading an unfamiliar one, even though both types of mindreading can be considered successful. Since you know the person well, you're able to understand their perspective more deeply. In this scenario, we use theory-theory to reach a general conclusion, and simulation theory to predict one's feelings with much more detail. The two theories work together to bring us accurate predictions of the minds of people we know.

(4) *Unfamiliar Person with Unfamiliar Situation (UPUS)*

Lastly, it is almost impossible to predict an unfamiliar person in an unfamiliar situation because we completely lost the ground for any theory to be useful. Since it is an unfamiliar situation, people do not have a pre-existing theory readily available to apply. Moreover, people can barely simulate the individual they are predicting since they know nothing about their background mental states. There is no stable basis for the predictor to begin their mindreading. The result of mindreading in this scenario would likely be, "I don't know what that would be like." An example for UPUS might be reading in the newspaper that a man from a savage tribe is about to be sent to Saturn, and you are asked to mindread his feelings when he arrives there for the first time.[33] The man from the savage tribe is an individual who has completely different customs and habits from you, and you have never heard of someone being on Saturn. Thus, there is no way to tell whether he will be happy or sad since you have no reference point for where to begin. As there is no ground for the two theories to build upon, the only answer for UPUS is "I don't know." It is arguably possible to develop folk psychological theories about the savage after learning more about them, however, when we have a developed theory, they should already be considered a familiar situation to us. In comparison to the UPUS case, if I were to predict my friend's feelings when they arrived on Saturn (FPUS), I could reach a serious prediction since I know her desires and beliefs well. Through my prior knowledge, she loves space and especially Saturn, I can accurately mindread what she will feel when she arrives on Saturn using simulation theory. However, under the UPUS scenario, there is too little information that theory and simulation cannot reach serious predictive conclusions that allow us to reference and interact with the predicted individual accordingly.

33 Not all off-script cases can be successfully mindread. The example here is clearly an abnormal case in which we tend to refer to a person's mind. However, there is too little information for us to successfully mindread.

Person (P)/ Situation (S)	Familiar P	Unfamiliar P
Familiar S	ST and T-T	T-T
Unfamiliar S	ST	Unable to predict

(Table No.2: Four Scenarios for Contextual Mindreading)

Views Against Collaborative Theories

My pluralistic argument entails certain limitations on the simulation theory, particularly that it is not the dominant theory people use when dealing with the UPFS scenario. Simulation theorists like Goldman might argue that simulation is not limited by the familiarity of people, and we can still use simulation theory to predict people's mind when we do not know them. Prediction does not differ according to familiarity, instead, simulationists claim that we apply one single method for all circumstances—we imagine a world from the perspective of the person we are predicting using their desires and beliefs.[34] Going back to the example raised for UPFS, simulationists might claim that we can simply picture ourselves in the old European lady's shoes using general human knowledge and mindread by imagining how we will react when seeing the black cat to ascribe the same decision to the lady.

However, the problem with the case Goldman is making here is that we cannot attain the old lady's beliefs and desires without using a theory of mind. Since we do not know this lady beforehand, the only way for us to have reasonable assumptions about their beliefs and desires is to induce through the situation they are in. In order to get from the information in a situation to an individual's potential mental state while largely maintaining accuracy, we must apply a theory instead of using imagination. Imagining what the lady will do when sees a black cat without a starting point has so many possible results and is getting us nowhere. Instead, knowing that this old lady is European, we are able to apply the theory, "If a European sees a black cat crossing their path, they consider it as a sign of bad luck and want to avoid it," to get the conclusion that the lady might decide to step back and avoid the cat. The only way to draw trust-worthy conclusions about someone's mental state in an unfamiliar situation is to induce through our experiences with similar situations or the stereotypes we hold. Therefore, in order to have enough information about an individual's beliefs and desires to simulate their perspective, we need to apply theory-theory. Goldman might further argue that theory-theory here is just the way we gain information for performing simulation, while simulation is still the main theory we use.[35] Yet, applying folk psychological theories is still the crucial step that we must take under the UPFS scenario to make simulation possible. Moreover, theory-theory also seems to be a more direct route to get to the results of prediction compared to first applying theory-theory and then performing simulation given that we are already using

34 Goldman, "Conceptualizing Simulation Theory," 27–30.

35 Goldman, "Conceptualizing Simulation Theory," 44.

theories in this scenario. The indispensability and convenience of using theory-theory are why I identify it as the dominant theory for UPFS.

Another possible objection is a general one concerning the pluralistic nature of my argument. Since my view is grounded in a pluralistic response to the problem of prediction, it is important to address significant objections to pluralism that may also apply to my argument. The most immediate and urgent attack on the general pluralist account might be the commonality problem. The commonality problem arises when we define something as "A is B or C." For B and C to both belong to A, they must share a common feature. If such a feature exists, it replaces the disjunction as the definition of A. Yet, without a shared feature, A cannot coherently include both B and C.

The problem is more pressing when it comes to my proposal on mindreading. My definition of mindreading can be expressed as "Mindreading is 'ST for FPUS,' or 'ST and T-T for FPFS,' or 'T-T for UPFS.'" Having disjunction as a definition, I should encounter the same problem of commonality other pluralistic statements had. Critics might argue that these three scenarios are too distinct to belong under a single concept of mindreading. If there is no unifying feature, mindreading risks incoherence. Conversely, if I identify a commonality, mindreading will simply reduce to that feature, undermining my pluralist framework.

However, this critique misunderstands the nature of my pluralism and does not threaten the pluralism that I propose in this paper. By comparing the two statements, we can identify several crucial distinctions between my mindreading statement and the statement of defining A as B or C. An example of the former is the grue problem introduced by Nelson Goodman.[36] Grue is defined as either "being green before time t" or "blue after time t," an artificial grouping of unrelated properties. The statement answers the question, "What is grue?" However, definitions for green or blue are completely different. Grue lacks coherence because it forces unrelated concepts (green and blue) into a single category at the same conceptual level, rather than a higher-level category such as "color" that logically encompasses them. The connective, "or," in this definition creates a disjunction that lacks intrinsic logic and is ultimately arbitrary.

By contrast, my statement treats mindreading as a higher-level category unified by its purpose: understanding and interpreting others' mental states. It does not force unrelated scenarios into an artificial unity. ST, T-T, and their combinations are distinct methods, but they are unified because they share a common goal of mindreading. Nonetheless, this commonality in purpose does not dissolve the distinctions between the methods. The disjunctions in my statement reflect the diversity of methods, which entails the contextual adaptability of mindreading instead of a lack of coherence. Thus, the commonality problem does not undermine my pluralism, as the shared purpose of mindreading provides the necessary unity without collapsing distinctions between methods. Since "to mindread" operates at a higher conceptual level than the individual processes (ST, T-T, and their combinations), the critique that applies to the grue problem does not apply to my pluralist account of mindreading.

36 Nelson Goodman, "The New Riddle of Induction," in *The Pragmatism Reader: From Peirce Through the Present*, edited by Robert B. Talisse and Scott F. Aikin (Princeton University Press, 2011).

To summarize this section, simulationists might object that their theory can still function when we are not familiar with the person we are predicting. However, I point out that one's beliefs and desires can only be obtained through the familiar situation and general theories for unfamiliar individuals. This shows that theory-theory is crucial and irreplaceable under the UPFS scenario and should be the dominant theory people use. Furthermore, to answer the commonality problem, mindreading is a higher-order category that encompasses multiple valid processes (ST, T-T, and combinations thereof), each serving a shared purpose: understanding others' mental states, which defines mindreading. Yet, finding the definition of mindreading is not what we are looking for in this paper. Instead, I focus on when and how we engage in mindreading. My answer deals with the general inquiries about mindreading but does not seek to define it in a similar manner as the grue statement. This distinction ensures that my pluralistic proposal avoids the incoherence of the commonality problem while maintaining the flexibility needed to account for the complexity of mindreading.

Returning to the False-Belief Task

It is time to revisit the first section and connect the four scenarios to the well-discussed false-belief experiment. I argue against theory-theory that the cognitive differences observed between children at ages two, three, and five stem from biological development instead of the obtainment of folk psychology theories. Around the age of five, children's biological functions are mature enough to enable them to recognize differences in people's viewpoints. The ability to differentiate perspectives is critical for the mindreading behavior across the three mindreading scenarios. The recognition of this fact does not fall within either dimension (familiarity with people or situations) but instead serves as a precondition for mindreading to occur. Theory-theorists suggest that this change of understanding results from children at the age of five having acquired more folk psychological theories. According to this view, repeated experience familiarizes children with mindreading tasks, accelerating their development of awareness of different perspectives. Theory-theorists are taking an empirical bet that simply increasing a child's exposure to mindreading situations would accelerate the ability to obtain theories that differentiate perspectives before the biological readiness around the age of five. I challenge this claim and argue that the difference is rooted in biological maturation with the simulation theorists, but with more complex reasons.

Around the age of two, children's cognitive processes can be classified as a special case within the fourth non-mindreading category in *Table No. 2* in Section III—unfamiliar people and unfamiliar situations. At this age, children are not aware of the beliefs of others or representational folk psychological theories. Yet, this case is not that easy to get out of like any other examples in the fourth category, at least not by simply informing the children what theory they should use or changing the person they are predicting. Their unawareness of any mental state of others is not due to their unfamiliarity with them, but because of their underdevelopment in the biological brain. Changing the individual they are mindreading to their guardians would not help with the results. Similarly, children's unfamiliarity with situations is also not because they have not yet experienced them before—they are asked to mindread in experiments several times—but due to their lack

of the biological capacity to apply the theory and acknowledge the uniqueness of each person's mind. Children's brains are still developing when they are three-years-old and reach full mental capacity regarding desires and beliefs at the age of five.

In all, this paper explores the enduring question of how people predict each other's mental states and behaviors, proposing a pluralist framework that integrates theory-theory and simulation theory. By categorizing mindreading scenarios based on familiarity with individuals and situations, I argue that these theories are not mutually exclusive but contextually collaborative. Simulation theory dominates in predicting familiar individuals in unfamiliar situations and using personal knowledge of individuals to simulate while theory-theory is more efficient when predicting unfamiliar individuals in familiar contexts, relying on generalized folk psychological theories. In situations in which both the individual and the context are familiar, the theories work together, enabling more detailed and accurate predictions. This paper advances the pluralist argument by providing concrete scenarios and categorizations, addressing the potential triviality in the application of pluralism to real-life cases, and laying the groundwork for a more nuanced discussion on which theory should be applied in specific contexts to enable successful mindreading. The success of human prediction lies in our ability to flexibly employ different theories depending on the context, a skill essential for navigating diverse social situations. These insights not only advance philosophical debates but also hold practical implications for fields such as psychology. Future research could build on this framework by exploring additional mindreading theories and their application across diverse cultural and situational contexts, deepening our understanding of this fundamental human capacity.

Works Cited

Bartsch, Karen, and Henry Wellman. "Young Children's Attribution of Action to Beliefs and Desires." *Child Development* 60, no. 4 (1989): 946–64.

Carruthers, Peter. "Simulation and Self-Knowledge: A Defence of Theory-Theory." In *Theories of Theories of Mind*, edited by Peter Carruthers and Peter K. Smith. Cambridge University Press, 1996.

Goldman, Alvin I. "Conceptualizing Simulation Theory." In *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.

Goldman, Alvin I. "Mirroring, Simulating and Mindreading." *Mind & Language* 24, no. 2 (2009): 235–52.

Goodman, Nelson. "The New Riddle of Induction." In *The Pragmatism Reader: From Peirce Through the Present,* edited by Robert B. Talisse and Scott F. Aikin. Princeton University Press, 2011.

Gopnik, Alison, and Henry M. Wellman. "The Theory Theory." In *Mapping the Mind: Domain Specificity in Cognition and Culture,* edited by Lawrence A. Hirschfeld and Susan A. Gelman. Cambridge University Press, 1994.

Gordon, Robert M. "Folk Psychology as Simulation." *Mind & Language* 1, no. 2 (1986): 158–71.

Heal, Jane. *Mind, Reason and Imagination*. Cambridge University Press, 2003.

Lavelle, Jane Suilin. *Mindreading and Social Cognition*. Cambridge University Press, 2022.

Nelson, Katherine, and Janice M. Gruendel. "At Morning It's Lunchtime: A Scriptal View of Children's Dialogues." *Discourse Processes* 2, no. 2 (1979): 73–94.

Nichols, Shaun, and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Clarendon Press, 2003.

Nichols, Shaun, and Stephen P. Stich. "Folk Psychology: Simulation or Tacit Theory?" *Philosophical Issues* 3 (1993): 225–70.

Wellman, Henry M., and Jacqueline D. Woolley. "From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology." *Cognition* 35, no. 3 (1990): 245–75.