AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2026, Vol. 133, No. 4, 919–956
https://doi.org/10.1037/rev0000623

Reverse Engineering the Centered Self

L. A. Paul[1, 2, 3], Tracey Mills[4], Tomer D. Ullman[5], Julian De Freitas[6], Cédric Colas[4, 7], and Joshua B. Tenenbaum[4]

[1] Department of Philosophy, Yale University
[2] Department of Psychology, Yale University
[3] Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich
[4] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[5] Department of Psychology, Harvard University
[6] Department of Marketing, Harvard Business School, Harvard University
[7] Flowers AI and CogSci Laboratory, Inria, Talence, France

In certain problem-solving contexts, people organize their domain through treating themselves as the perceptual and cognitive center of their world. They identify and solve a particular problem from their perspective as a particular agent, with a particular location, at a particular time, in a particular environment. When they do this, they engage in a distinctive kind of *agent-centered* problem-solving. For many contexts involving intelligent agents, partially observable Markov decision processes (POMDPs), a framework for modeling decision-making in uncertain environments unfolding over time, have effectively become a "standard model." Yet, as these models are ordinarily interpreted, they do not explicitly represent agent-centered problem-solving. Accordingly, to model this type of problem-solving, we begin by extending the standard POMDP framework to define "ePOMDPs." This formalism models how an agent, once it centers itself on a particular self-and-world representation, plans and acts rationally from its own perspective. To capture the way that such agents choose which problem to solve, we build on our ePOMDPs to develop a "meta-ePOMDP" agent within a hierarchical Bayesian framework. We implement meta-ePOMDP agents for two different suites of "centering game" tasks that highlight different aspects of our theory. We find that our models explain signatures of agent-centered problem-solving not captured by alternative models, in particular, the difficulty of navigating spaces of possible problem representations. We close by suggesting that our model could provide the beginnings of a computational framework for how a person could have a self.

Keywords: self, agent, planning, centering, partially observable Markov decision processes

Computational accounts of human cognition often rely on a framework of approximate or bounded rational agency as a foundation for understanding goal-directed behavior. One influential framework models a rational agent as one that aims to choose actions that approximately maximize its expected utility in a given environment, subject to the agent's approximately Bayesian beliefs about the structure and state of the environment as well as constraints on the agent's own computational resources (e.g., Gershman et al., 2015; Griffiths et al., 2010, 2024; Lieder & Griffiths, 2020; Russell & Norvig, 2016, 2020; Savage, 1954; Simon, 1955; Von Neumann &

Elena L. Grigorenko served as action editor.
L. A. Paul https://orcid.org/0000-0002-2280-7695
Tracey Mills https://orcid.org/0000-0001-6278-6721
Tomer D. Ullman https://orcid.org/0000-0003-1722-2382
Julian De Freitas https://orcid.org/0000-0003-4912-1391
Cédric Colas https://orcid.org/0000-0003-0212-427X
Joshua B. Tenenbaum https://orcid.org/0000-0002-1925-2035

L. A. Paul and Tracey Mills contributed equally to this article. Data, study materials, and analysis code for Studies 1, 2, and 3 are available on the Open Science Framework at https://osf.io/48gav/. The preregistration for the parameter fitting analysis for Study 2 is available on the Open Science Framework at https://osf.io/5tcze/. The human data analyzed in Study 1 were published in De Freitas et al. (2023). A preliminary version of the modeling work described in Study 1 was presented at a poster session at the 6th International Workshop on Intrinsically Motivated Open-ended Learning (IMOL 2023). Figure 1 ("You are Here Now") was created by Melisa Machuret ("Whiteboard-Girl").

The authors have no conflicts of interest to disclose. This work was supported by the Center for Minds, Brains and Machines under National Science Foundation Science and Technology Center Grant CCF-1231216, Tomer D. Ullman is supported by the Jacobs Foundation.

The authors are grateful to Adam Bear, John Bengson, José Luis Bermudez, Paul Bloom, Joshua Cohen, Philip Corlett, Fiery Cushman, Stan Dehaene, Nina Emery, Branden Fitelson, Sam Gershman, Dan Greco, Max Kleiman-Weiner, Marta Kryven, Michael Littman, Dilip Ninan, Brian Scholl, Ted Shear, Jack Spencer, Barry Taylor, R. Jay Wallace, Ahmet K. Uğuralp, Zeliha Oğuz-Uğuralp, Ilker Yildirim, members of the Columbia University First Person Discussion Group, members of the Yale Perception and Cognition Lab, and members of the Cognition and Neural Computation Lab, and several referees for helpful comments on earlier versions of this article.

L. A. Paul played a lead role in writing–original draft. Tracey Mills played a supporting role in writing–original draft. Tomer D. Ullman played a supporting role in writing–original draft. Julian De Freitas played a supporting role in writing–original draft. Cédric Colas played a supporting role in writing–original draft. Joshua B. Tenenbaum played an equal role in writing–original draft.

Correspondence concerning this article should be addressed to L. A. Paul, Department of Philosophy, Yale University, 451 College Street, New Haven, CT 06511, United States. Email: la.paul@yale.edu

Morgenstern, 1994). This expected utility framing is designed to quantitatively capture the rational core of intentional agency: An agent believes that its actions will lead to the fulfillment of its goals (cf. Dennett, 1987).

Models characterizing human agents in this way have proven useful and successful in many fields, including operations research, robotics, computational neuroscience, and computational cognitive science as well as the broader landscape of the computational social sciences. The approach also extends to proposals for how people model the minds of others (see, e.g., Baker et al., 2009, 2017; Butterfield et al., 2009; Jara-Ettinger et al., 2016; Jern & Kemp, 2015).

Yet, as these models are ordinarily interpreted, they leave something out. They do not include a characterization of a distinctive way that people can formulate and solve problems, one which has been identified in the philosophical literature on "self-locating" attitudes and "centering" (Perry, 1979; Stalnaker, 2008). This literature inspires our work here on what we call *agent-centered problem-solving*, or for short, "centered problem-solving."

Centered problem-solving is inherently perspectival. As biological creatures with brains and bodies that evolved in the physical world, human beings naturally organize their actions through mental models that treat themselves as agents who are the perceptual and cognitive centers of their world. That is, their mental models presuppose that they are embedded in a particular environment, oriented as a particular agent, with perceptual access and causal action affordances keyed to some particular temporal and spatial location. Given a problem or a task to solve, people formulate plans to achieve their goals and execute these plans from this agent-centered perspective (see, e.g., Avraamides & Kelly, 2008; Burgess, 2006; for philosophical discussion, see, e.g., Bermúdez, 2017, 2018; List, 2023; Paul, 2016; Paul & Quiggin, 2018; Perry, 1979; Recanati, 2012a).

From this centered perspective, people can engage an appropriate mental model to guide their actions, in a vast range of possibilities, using a wide variety of cues and signals to do so, including novel kinds of environments and data sources with little or no precedent in their evolutionary or personal histories. People can also recognize when they have "mislocated" themselves in some sense and thus need to *recenter* themselves: That is, they need to change the agent-centered problem they are solving or the way they are representing themselves as agents in the problem to be solved. For example, people can relocate themselves on a map when they realize they are not where they thought they were, or change their plans when they realize it is later than they thought it was. Such examples parallel early psychological studies of the sensorimotor foundations of agency, in which people infer "I caused that" by integrating motor intentions, action outcomes, and sensory feedback (Haggard et al., 2002; Moore & Fletcher, 2012), and suggest that in a very basic sense, knowing where—and when—you act is a metacognitive prerequisite for goal-directed action. Here, we extend this perspective to a broader framework of centered problem-solving in physical and digital worlds.

In this article, we develop and test a computational framework for modeling this kind of flexible and adaptive agent-centered problem-solving, with the aim of reverse engineering the capacities that people use to solve problems from a centered perspective and to recognize when they need to adaptively recenter, changing what they are doing in order to achieve their goals.

The plan of the article is as follows: We begin with concrete examples that illustrate flexible centering and adaptive recentering and their role in distinctively human forms of problem-solving (we defer to Appendix A for a more technical introduction to the philosophical framework that underlies the concept of centering, and that motivates some of these examples). Importantly, we explore the way that centering and recentering can be understood as a solution to the meta-problem of deciding which problem to solve, a problem that requires specifying a representation of the world but also of ourselves within the world.

After presenting our examples, we introduce our computational approach based on a hierarchical, approximately Bayesian, framework for inference and decision-making. This framework extends the tradition of rational agent models to model centered problem-solving. We argue that the capacity to center and recenter is a key computational building block that human intelligence relies on to solve perspective-dependent problems. Our computational proposal unfolds in two formal steps, each capturing a distinct facet of centered problem-solving:

1. *ePOMDPs:* We begin by extending the standard POMDP (partially observable Markov decision process) agent formalism for rational decision-making and planning under uncertainty to what we call extended POMDPs (ePOMDPs), which explicitly separate the agent's state from its environment and represent the problem solver's world model from the agent's perspective, locating the agent within the environment, with a certain position, orientation and perspective on its environment. This formalism models *agent-centered problem-solving*: It models how an agent, once centered on a particular self-and-world representation, plans and acts rationally from its own perspective.

2. *meta-ePOMDPs:* We then lift this account into a hierarchical decision framework, the meta-ePOMDP, where the very choice of which ePOMDP to solve becomes an explicit rational inference and decision problem. A meta-ePOMDP decision-maker *centers* itself by choosing the ePOMDP that currently best captures the problem it needs to solve from a set of possible ePOMDPs. When it makes its choice of ePOMDP, it takes into account its goals and beliefs about the world it is in. It can also *recenter*, by adopting a new ePOMDP as the agent-centered problem it is solving now, when evidence or circumstances suggest that a different problem formulation is required, or when a change in its higher level goals require it to consider different, alternatively centered ePOMDPs. This formalism models how an agent decides which problem to solve.

To test our framework, we implement meta-ePOMDP agents for two suites of "centering games" that probe complementary aspects of our theory—inferring which agent one is, and selecting which problem to solve—in different video game domains, including the popular puzzle video game "Baba Is You" (Hempuli, 2019). In each domain, we compare the behavior (solving speeds and accuracies) of meta-ePOMDP and alternative models against human participants. Across both domains, we find that meta-ePOMDP models best capture the efficiency of human agent-centered planning and people's characteristic patterns of centering and recentering. Together,

this hierarchical modeling framework qualitatively and quantitatively predicts relative task difficulty, demonstrating that our two-step framework mirrors key features of how people engage in centered problem-solving.

We close the article by suggesting that the ability to center (and recenter) could provide the beginnings of a computational foundation for "having a self": hence, our title. This is, admittedly, a philosophical suggestion: Our thought is that the distinctive way that people center themselves by locating and representing themselves as agents, and then planning and acting from their perspective, is conceptually related to what are called "self-locating" attitudes in the philosophy of language and mind (e.g., Bermúdez, 2017, 2018; Bratman, 2018; Ismael, 2007; Lewis, 1979; List, 2023; Paul, 2016; Paul & Quiggin, 2018; Peacocke, 1999; Perry, 1979; Recanati, 2012a; and for further discussion, see our Appendix A). These discussions relate philosophical concepts of identity and self to the use of indexicals in language (especially the first-person pronoun "I") and belief, and to the epistemology of action, decision-making, bodily navigation, and awareness. Juxtaposed with this philosophical context, our computational framework can be understood as a proposal for reverse engineering the computational ground of the self.

However, we hasten to add that our empirical interpretation of centering is necessarily philosophically minimal and is not intended to capture the deeper and theoretically rich content explored in the philosophical debates about the de se, self-location, the first person, or self-awareness understood more broadly. With this in mind, our concluding discussion engages with psychological characterizations of inferential and computational features of how people have selves that relate to our minimal conception of centering, and our ultimate hope is that our computational model could provide a candidate general framework for developing ways to integrate appropriate self-type inferences with extant philosophical discussions.

Centering: A Philosophical Introduction

Centering as Self-Location: You Are Here Now

What does it mean to say that people can flexibly center themselves perceptually and cognitively in order to solve certain problems? A person centers themselves by representing themselves as located within their world, with a certain perspective on their world that triangulates between their first-person representation of who, where, and when they are, and a third-person representation of themselves as at their spatial and temporal position. This process parallels classic studies of first-person self-localization, where multisensory integration (e.g., visual–tactile synchrony) determines bodily self-location (Lenggenhager et al., 2007; Petkova & Ehrsson, 2008; Schettler et al., 2020) but effectively involves a hybrid of more familiar first-person and third-person perspectives.

There is a subtle feature of the human computational feat of centering, involving a kind of self-awareness, that deserves particular attention. When you, as a thinking, experiencing person, represent and orient yourself perceptually in your environment, you do it by taking yourself to be the agent that is at the center of this representation. You do not just coordinate representations of where and when you are; you also coordinate first-person and third-person representations of *who* you are.

We rely on such coordination at an intuitive level, one that is perhaps so basic that it is difficult to recognize. An example can help to detect it. Suppose that you have never encountered the notion of a physical (or virtual) map (Figure 1). One day, you are visiting a new city, and you find yourself lost. To solve this problem, you ask someone in front of you for directions, and they point to a dot on a nearby picture of a street grid labeled with "you are here now." A potential response on your part might then be: "no, I'm over here, in front of you. That is a dot on a piece of paper on a billboard. Rude."

The initial barrier to cross here is understanding that *you* are the referent of the dot on the paper, such that this dot represents your spatial location from the map's global or "bird's-eye" perspective. You center yourself cognitively by recognizing that you, as the agent, are represented by the dot on the map in the map's overhead view, and take the dot to "be" you. You also center yourself perceptually, by coordinating visual lines of sight and other sensory information. This dual act—perceptually locating yourself in the world and cognitively identifying the dot as "being" you—is the essence of the perspectival self-location at the heart of agent-centered problem-solving. Such centering naturally draws on body schema representations—mental representations of the body's parts and interconnections (Berlucchi & Aglioti, 1997; Botvinick & Cohen, 1998; Head & Holmes, 1911)—integrating multisensory inputs to establish both where and who one is from a first-person perspective. In daily life, as we will suggest in our discussion below, you center on your body; in our example, you center on the dot; and we will also see that the same logic applies when centering on

Figure 1
Locating Oneself on a Map: You Are Here Now

Note. See the online article for the color version of this figure.

characters in a story or a game, or a video game avatar. From your center, you can solve problems.

Importantly, centering involves the coordination of one's egocentric, first-person information, such as visual lines of sight and other sensory representations, with allocentric, observational information about one's spatial and temporal position. Once you have centered yourself on the dot, you can then solve problems from this perspective, using the map to perform tasks that you could not otherwise perform. After all, a map is useless until you locate yourself on it.

For example, you might see a building in front of you that looks like a school. You could have no interest in the school for its own sake, but you could use this information to determine that you need to turn left to get to a cafe, that is, to know where you are, by recognizing its relationship to your location as it is represented by the dot on the map, which also marks the location of the cafe and the school. Then you can triangulate your first-person line of sight on the building with your external (third-person) representation of yourself at your map location in order to figure out how to get to your goal.

The map example exemplifies agent-centered problem-solving: First, a person centers themselves in a representation of their environment at a time, in a place, and as an agent, and once centered, they solve a particular problem from their centered perspective (as that particular agent, from that particular location, at that particular time, in that particular environment).

Our account of centered problem-solving in human action and planning extends a long tradition of philosophical thought experiments that illustrate the linguistic and cognitive importance of centering for perspectival thought and action, and for conceptions of self-involving content (for contemporary discussions, see Dennett, 1991, 1992; Ismael, 2007; Nagel, 1989; Paul, 2014, 2017a; Perry, 1979; Pollock, 2006). We note however, that given our empirical approach, we interpret first- and third-person perspectives in a minimal sense that does not address many of the richer philosophical dimensions of such perspectives discussed in philosophical contexts. For a brief introduction to the philosophical tradition and an account of centering in terms of philosophical models of indexical attitudes and self-involving content, see Appendix A.

Centering as Problem Identification: Which Problem Should I Solve?

We have shown how centering can be understood as a form of self-location, such as when locating yourself on a map of a city: You center yourself when you locate yourself temporally, spatially, and "agentially," as an agent with specific goals and affordances with respect to an otherwise fixed representation of the problem at hand. However, centering can also involve a more abstract type of task, namely, determining which problem it is that you want to solve. Illustrating this with our map example, we can put the point this way: as you center yourself on a map, you must also solve a series of high-level problems. For example, which map should you use? What scale of detail should you focus on? Should you take a bike or the subway, or a combination of the two, instead of simply walking—and if so what map is an appropriate guide for this kind of planning?

To make these questions more concrete, consider navigating Paris. To successfully navigate Paris, at a minimum, you need a map that appropriately represents the structure of Paris. If you tried to use a map that mixed up the different neighborhoods, did not include the Eiffel Tower, or swapped the Louvre for Centre Pompidou, you would quickly find yourself lost. In this sense, an appropriate map must be structure preserving: Spatial relationships between sets of points, regions, and paths on the map should (at least approximately) preserve the corresponding spatial relationships between analogous points, regions, and paths in the real-world environment—that is, Paris—with sufficient fidelity that you can use it to find your way.

But being an appropriate map—being the right map for an agent's purposes—is not only about being accurate in this sense. The map must have the right scope: It must span enough of Paris to include both the agent's location and the locations of the goals the agent needs to navigate to, and the goal locations must be clearly marked on the map and findable in the world. The map should have the right scale: It should have enough detail about navigable paths through the city for the agent to plan efficient and effective sequences of steps and be able to identify in the real world when and where to turn as they follow these paths, but not so much low-level detail as to leave the planner lost in complexity, pondering every cobblestone and curb. If the agent's goal is to get to a cafe across town within 30 min, the map ought to include metro lines and stops, because walking is unlikely to be sufficient (it could even be solely a metro map if the goal is just to transfer from one train station to another). In short, out of the infinitely many structure-preserving maps of Paris, only a small subset may be appropriate for a resource-bounded agent to use to achieve their goals.

When making a choice of which map to use, an agent is making a choice about which version of the navigation problem they are going to solve. The selection of the map is thus a centering operation: By choosing a particular map, you choose a particular representation of the city. And crucially, any navigating agent must make this choice. Even if they do not have the luxury of an external map to consult, they need to construct some version of a map in their mind to guide their path planning and online navigation. Your choice of how to represent the city functions is part of the choice of how you want to be guided through the city, defining how you want to construct and execute your plans.

In this sense, your choice of map is a choice of world model, a choice of model for the part of the world you want to navigate. The same holds for other kinds of world models. We can think of a world model as any behaviorally efficacious representation of the world and its dynamics, generated (in animals) by the brain through perceptual contact with the environment. A map is one form of world model, but there are many others that are also structure preserving such that the agent can use them for flexible belief formation and action planning (Bi et al., 2024; Tolman, 1948; Yildirim & Paul, 2024). Just as you can use a spatial map of Paris to spatially navigate Paris, you can use a domain-specific world model as a type of "conceptual map" to navigate a domain. For example, when you solve a problem involving everyday physics, such as throwing a ball to hit a target, you can use an intuitive physics model, or a world model that represents the physical dynamics of your space (Battaglia et al., 2013; Ullman et al., 2018).

Note that, when using a world model as your conceptual guide to (efficiently) throw a ball and hit a target, you need a world model that correctly (or at least well enough) represents the physics of the world. Importantly, the model includes rules for how the world unfolds. For example, our physics world model includes representations of the laws

of physics (or causal mechanisms compatible with them) that determine the trajectory of a ball with mass *m* when thrown in such and such a way. If you tried to use a world model that misrepresents the physics of the actual world, for example, it miscalibrates the strength of gravity (e.g., using the earth's gravity while throwing a ball on the moon or in outer space), inverts the force of gravity, or fails to include the mass of the ball, you would not reliably hit your target. And just as with a mental map of a space, a mental model of the world's physics can vary greatly in scope, detail, and degree of approximation. People are adept at constructing simplified models of the world with the right content and grain, whether for the purposes of efficiently planning actions (Ho et al., 2022), making physical predictions (Chen et al., 2026), or solving physical problems (Allen et al., 2020). Adopting or constructing a particular physical world model that is appropriate for the agent's goals is a crucial element of the choice about what problem the agent is going to solve.

More generally, formulating an agent-centered problem, in addition to formulating a sufficiently precise description of the task, involves selecting a world model as a conceptual guide to define, frame, and achieve your specified objectives for your local environment. Selecting among alternative world models thus involves a metacognitive decision about which representation of the rules of the world to use—akin to judgments of task difficulty and strategy selection in problem-solving (Metcalfe & Shimamura, 1994). We can now see how agent-centered problem-solving involves more than locating oneself at a spatial or temporal location (and more than locating which agent one is). People also center themselves in a more abstract sense, by locating themselves within a space of conceptual possibilities representing different problems they could be solving or thinking about. Putting all these pieces together, we can characterize successful agent-centered problem-solving as, at least in part, locating yourself at the right problem representation, that is, locating yourself in the right place in the space of conceptual possibilities that represents the different problems you could be solving or thinking about.

In this way, we interpret the meta-problem of "which problem do I need to solve?" as the problem of locating yourself on a map of conceptual possibilities. Locating yourself at a conceptual possibility amounts to choosing a particular problem representation, which, in addition to its dependence on how abstractly a problem is defined or described, includes choosing a world model that represents the world as being a particular way. For short: By locating yourself in this way, you are conceptually defining and representing yourself as solving *this* problem.

Recentering

As you wander through Paris, you become disoriented. When you emerge from the Métro, you recalibrate yourself with your phone's GPS, perhaps by turning your body or your phone or walking a few steps in a certain direction. This is an ordinary way to relocate yourself in physical space by recentering yourself on your map. When "recentering" in this way, a person discovers that they are misaligned, and in order to properly realign, performs a type of cognitive adjustment to relocate (or find) themselves by representing themselves at a new location, often with a new perspective. Metacognitive monitoring of these cues—deciding which perspective to trust when they conflict—reflects higher order reasoning about one's own perceptual estimates (Fleming & Dolan, 2012). This type of minor recentering in space or time does not ordinarily amount to a change of world model or a change in who one takes themselves to be.

However, a more dramatic change of plans, due to a change in your understanding of the situation at hand, could require a change in world model. Just as with spatial disorientation and reorientation, the way you think your world is and the way your world actually is can come apart. If you are confused or make a mistake about what your world is actually like (or your world suddenly changes), you will choose the wrong world model, and any thinking you do based on that mental model is effectively trying to solve the wrong problem. To act efficiently, you must be able to recognize when this happens and correct for it, that is, you must recenter.

For example: Perhaps, as you attempt to find your way to the Eiffel Tower, union leaders call a sudden strike, and the trains are halted. This means that using the Métro to get to the Eiffel Tower is no longer a good strategy; you would be solving the wrong problem by trying to do so. You recognize this, and spying a rack of green mechanical bicycles nearby, you rent one, changing your route altogether. Here, you change the way you get to the Eiffel Tower by changing which problem you are solving, choosing a new world model (for this little part of the world), and conceptually recentering yourself using this new model. This conceptual shift resembles the "representational change" in insight problem-solving where overcoming functional fixedness requires reframing the problem space (Chrysikou, 2006; Ward et al., 1997).

However, you can also be conceptually disoriented in even deeper and more dramatic ways (Ongchoco et al., 2024; Paul, 2014). For example, you could choose a radically mistaken world model, entailing that you use the wrong conceptual framework altogether. Imagine you are playing an immersive virtual reality game that inserts you randomly into a city. The city around you looks a little like Paris, with names of streets you recognize from Paris, and so you engage your world model of Paris. Yet, as you play, you keep finding yourself confused about where you are, because you do not see the landmarks or streets you expected. And then at some point, you see a building with the name of "Bank of Montreal," and you realize your mistake.

Even more disorienting, you could discover that you have also been mistaken about your own agency, and thus, you need to replace more than just your world model: You need to change the way you understand yourself as the agent who is solving the problem in your world model. This involves an important change in one's beliefs, a type of self-locating adjustment as part of changing the way you represent your world (Paul, 2014; Perry, 1979). In this situation, you need to revise the "agent-centered" element of the problem you are solving, with consequent revisions in your understanding of how you think of yourself as thinking and acting in the world, that is, with consequent changes in which problem you take yourself to be solving. The conceptual and psychological importance of needing to correctly center on who you are in order to solve the right problem, and the consequent changes of how one understands oneself as a problem-solving agent in one's world, can be illustrated by an additional example, this one drawn from a famous children's story.

Woozle Tracks

One morning, just after the snow stops, Piglet happens upon Pooh as he is walking very slowly, staring intently at the ground. Pooh

tells Piglet that he is tracking a Woozle, and they proceed to track together. The tension escalates as a second set of tracks, and then a third, joins the first. When a fourth set of tracks appears, Piglet abruptly leaves the area. Pooh, left alone to reflect, is interrupted from his thoughts by Christopher Robin, who has been watching the whole scene from up above, perched in the old oak tree.

"Silly old Bear," says Christopher Robin. "What were you doing? First you went round the spinney [a small tree] twice by yourself, and then Piglet ran after you and you went round again together, and then you were just going round a fourth time." On hearing this, Pooh stops, sits down, and thinks. Then, very slowly, he fits one of his paws into the tracks, and it dawns on him: He was making the tracks himself! (Milne & Shepard, 1926).

When Pooh fits his paw into the track, he combines his own perceptual inputs and his own perspective on himself with observational information of the scene given by another perspective (that of Christopher Robin). This additional information changes the beliefs he has about who was responsible for making the tracks in the snow, changing his beliefs about who exists in his world and what they are doing, and importantly, what he is doing.

More generally, when he found himself solving the wrong problem, he responded by discovering and solving a new kind of problem. He did so by reviewing the way he was thinking of the agents in the world and selecting a new representation of himself that better fit his evidence (the evidence provided by Christopher Robin plus his matching footprint) as one of those agents. Pooh's problem was that he misrepresented himself and his agency in his world and that his original world model included an agent that did not exist (a Woozle). Recentering himself on a new agent (himself), he replaces his world model, conceptually realigning himself and solving his problem.

Our goal in this article is to develop a theory and computational modeling framework that is rich enough to include even this type of case, so as to capture a sense in which agent-centered problem-solving aligns with how people understand themselves as thinkers, solving problems from their own perspective.

Philosophical readers will recognize that our Woozle example mirrors Perry's (1979) classic example of spilling the sugar in the supermarket (we note that Milne's Winnie the Pooh story predates the Perry example by more than 50 years). As such, our account of agent centered problem-solving in human action and planning extends a long tradition of philosophical thought experiments illustrating the linguistic and cognitive importance of centering for belief and agency (for contemporary discussions, see Dennett, 1991, 1992; Ismael, 2007; Nagel, 1989; Paul, 2014, 2017a, 2017b; Perry, 1979; Pollock, 2006). For readers who wish to explore the philosophical debates in more depth, we provide further technical discussion and additional philosophical context in Appendix A. One might also wonder how our proposal is connected to extant work on creative problem-solving. For a discussion of the relationship between our proposal and those literatures, see Appendix B.

A Computational Formalization of Centering and Recentering

We will now offer a computational formalization of our conceptual framework for agent-centered problem-solving. We first formalize (with "ePOMDPs") the way that grounded, embodied agents distinguish between themselves and the world to solve these types of problems. Building on our ePOMDP approach (using "meta-ePOMDPs"), we then formalize our account of centering as self-location as the meta-problem of choosing which agent (at a time and location) to center on, and centering as problem identification as the meta-problem of which problem to solve (which can include self-location as a proper part). We also include formalizations of recentering.

Our formalization builds on well-established machinery for modeling rational agents, starting with what has effectively become a "standard model" of intelligent agency in artificial intelligence (AI) and computational cognitive science. On this approach, rationality is a standard of correctness for actions, and a rational agent is one that chooses actions to maximize its expected performance measure based on its current knowledge and percepts of the environment. The many computations and capacities that comprise intelligence in such an agent can be captured by what Peter Norvig has called the "fundamental equation of AI" (Figure 2).

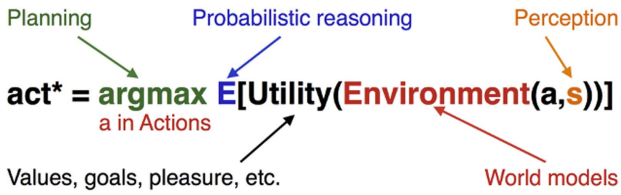
Formally, an agent is represented as making *plans* to take actions (a) that maximize (*argmax*) its utility in expectation (E), where the expected utility is a function of the agent's beliefs formed by *probabilistic reasoning* over its current state (s), which itself is inferred from a model of the *environment* and its dynamics, updated based on the agent's perceptual observations and its own actions.

Note that this equation can apply to subjectively rational agents with limitations on memory, compute time, knowledge, and so on (see, e.g., Bratman, 1987, 2018; Bratman et al., 1988; Gershman et al., 2015; Lieder & Griffiths, 2020; Paul & Quiggin, 2018). However, the structure by itself admits arbitrary utilities, beliefs, environments, and actions. Actual implementations require additional models of the world in order to be effective. Below, we build on the standard model to present a computational model of the way people center and recenter themselves in order to solve problems.

We propose to use POMDPs to implement our model of agent-centered problem-solving. We choose POMDPs as they are well-suited as a model for intelligent action under uncertainty

Figure 2

Schematic of the Fundamental Equation of Artificial Intelligence



Note. Effectively searching over a large space of actions requires *planning*. Knowing which actions are available in a given state requires a *world model*, as does knowing how the environment will evolve given a particular state and action pair. Calculating the expectation requires *probabilistic reasoning* about likely results and priors over states and observations. Utilities can be more than simple mappings from states to an ordering and can involve *goals and values* which are themselves drawn from expectations about correct behavior. The state is not directly given, but rather is filtered through observation, and the link between the true state, the observation, and the resulting belief requires a model of *perception*. Figure design based on a concept from Max Kleiman–Weiner, with inspiration from Peter Norvig (Kleiman-Weiner et al., 2017). See the online article for the color version of this figure.

(Kaelbling et al., 1998) in the sense of achieving a goal in a situation that is well-specified by the system's engineer.[1] We emphasize that our approach does not hinge on using this particular framework; it is just one way to concretely implement our abstract conceptual claim.

As an example of a simple POMDP, consider a scenario in which an agent navigates a small grid world to find sources of reward, given its beliefs and desires (see Figure 3). Given a specific setup (possible states and actions as well as utility, observation, and transition functions), the solution to this POMDP will be a series of actions that moves the agent from its initial state to maximize its utility. For example, the POMDP solution could dictate a series of moves to get to L, assuming the reward for L outweighs the cost of motion given the agent's location.

In this example, the state is unitary and does not distinguish the agent from the world (this will be relevant to the extensions of the model we propose below). While the state description includes the agent location, it is no different from including the location of a wall or a potential goal. In this sense, then, it fails to incorporate the perspectival nature of the type of problem-solving we intend to capture. The type of problem-solving we are exploring can be characterized by the extensions we detail next.

Modeling Agents Separately From the Rest of the World: ePOMDPs

To focus on the particular space of interest, we extend the standard POMDP machinery (we propose "extended POMDPs," or for short, "ePOMDPs") to represent the way that agents organize their domain from their agent-centered perspective. More specifically, ePOMDPs are a class of POMDPs that include:

1. *Separation between world and agent states:* The most basic distinction missing in a standard POMDP is a distinction between the representation of the agent and the representation of the rest of the world. In a POMDP without any other specifications, there is a single unitary notion of "state" that includes both the state of the world (e.g., which objects are where) and the state of the agent in the world (e.g., "where am I?"). Our extension of a POMDP distinguishes the states of the world and the states of the agent, replacing the unitary space of abstract states with a cross product of two spaces: one for the state of the world (W) and one for the state of the agent (M). For now, it does not matter that the agent correctly distinguishes W and M, only that this distinction exists in the representation of the overall state. We note that such a distinction is not only natural but already useful in many existing POMDP formulations. For example, consider how Rabinowitz et al. (2018) separated W in the sense of a transition function that dictates movement in a grid world with walls, from M in the sense of agents with different utilities and observations (see also Haber et al., 2018; Kim et al., 2020, for examples of explicitly separating a self-model and a world model in a learning agent). This distinction allows one to put different agents into the same world, or the same agent into different worlds, with varying expected behaviors. The notion of "same" here for "same world" and "same agent" rely on distinguishing agent states and world states. Again, our intention is to call attention to the fact that the standard formal setup of a planning framework such as a POMDP does not explicitly distinguish agent and world states nor require it, and yet that this distinction is necessary, natural, and consequential. Our conceptual demarcation is needed to capture the way human agents distinguish between themselves and the rest of the world, centering and recentering themselves in their environments in order to act effectively (Bermúdez, 2017; Ismael, 2007; Paul & Quiggin, 2018).

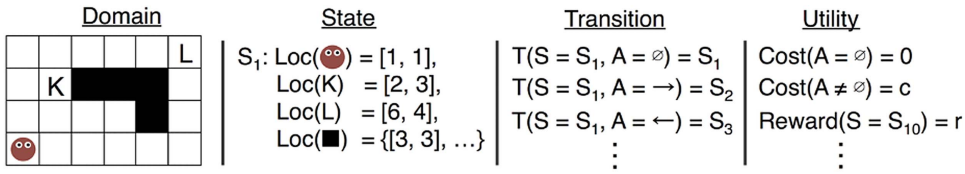
2. *Entity classes:* Beyond the split between the agent state and the world state, an ePOMDP differentiates structure in the world state W through classes of entities that constrain and shape utilities, transitions, observations, and actions. Classes of useful entities can be thought of as theories, and particularly useful ones include objects (possible sources of reward, and targets of actions), actors, and barriers (constraints on actions and observations; Lake et al., 2017). To give a specific example, consider the grid world described in Figure 3. Building a POMDP of the grid world, an engineer could in principle arbitrarily assign numbers to undifferentiated states and use those to learn a full $S \times S \times A$ transition matrix (which describes the probability of transitioning from state s in S to a different state in S given an action in A). But this is not done in practice. Instead, the engineer may create an entity class such as a "wall," and a corresponding piece of code stating agents cannot move through walls. New environments can then populate the transition matrix without much hand coding, by porting over the "wall" class.

There are a number of other extensions that are useful both in principle and in practice for defining a full ePOMDP in the sense that it extends the "fundamental equation of AI" of Figure 2. These include extended observation functions O (replacing the arbitrary mapping from observations to states with sensory perception); a distinction between first- and third-person representations; extended transition functions (replacing the arbitrary mapping from state and actions to states with intuitive physics); extended action sets A (replacing a large, fixed set of actions with actions generated on the fly); and extended utilities U (replacing an arbitrary mapping from state to rank/number with a more temporally extended decision process based on world models and theories of harm, intention, and morality). Such extensions would bring the example shown in Figure 3 closer in line with the model sketch in Figure 4. However, we omit the details of these other extensions, as they are not strictly necessary for the main theoretical points, thought experiments, and empirical experiments we consider here.

Though these extensions to the standard POMDP machinery allow us to model an agent that solves a problem from a centered perspective, they do not yet allow us to capture the notion of switching from one conceptual possibility to another. To do this, we need to attend to one level above this planning framework. We need to move to the level of an agent that sets up and solves an ePOMDP for itself while representing itself as performing this process, the

[1] A formal overview and interactive introduction to POMDPs as models of agents is available in Chapter 3 of Evans, Stuhlmüller, Salvatier, and Filan (2017; https://agentmodels.org/chapters/3-agents-as-programs.html).

Figure 3
A Simple Partially Observable Markov Decision Processes



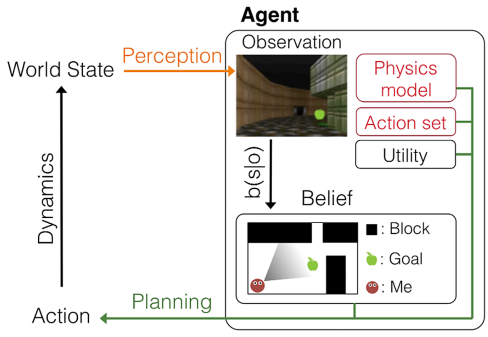
Note. An example of a planning domain, in which the state encodes the location of the agent and items, and the position of the walls. The transition function implements grid-like motion, and the utility is composed of costs of actions and rewards from reaching particular states. We have omitted the observation function for brevity. The "solution" to the partially observable Markov decision processes with this setup would dictate the motion of a rational agent, such as moving toward L. Here, the transition, utility, and observation functions have all been prespecified, and implement a grid-like world with particular rewards. The general case partially observable Markov decision processes can involve arbitrary action sets, states, transitions, utilities, and observation functions. See the online article for the color version of this figure.

level of the meta-ePOMDP. We can think of an agent solving this higher level problem as generating different ePOMDPs for different tasks and selecting the right one.

Modeling Centering With Meta-ePOMDPs

If we consider an ePOMDP as capturing a more humanlike way of accurately representing a specific problem, such as driving a car, playing chess, or picking up the kids from a movie, then even when solving the problem as formulated by that ePOMDP, the person solving the ePOMDP is more than just the solver of that particular problem. In order to account for the way that people reason and choose between actions, we must consider that the person could in principle generate either the driving ePOMDP or the chess-playing ePOMDP or the child-pickup ePOMDP, or a myriad of others, and

Figure 4
A Simple Extending the Standard Partially Observable Markov Decision Processes



Note. A simplified depiction of a model in which an agent observes its environment through first-person perception (of observations o) and updates its beliefs (b(s|o)) about its state (s), which includes a model of the entities in play, their types, and their effects on the environment. The agent can use its intuitive physics model (a world model) and beliefs to choose an action from the action set it believes is available to it, and plan to take actions that change the world state to increase its utility. Note that the agent's belief about the state includes an identification of itself. See the online article for the color version of this figure.

then select between them. Such an agent is not described by a single ePOMDP, but rather by an ePOMDP generator, together with a process for evaluating and selecting an ePOMDP.

If we envision all ePOMDPs that the agent might represent as inhabiting a space, then centering as problem identification involves the task of selecting the right ePOMDP from this space to account for a privileged set of independently specified perceptual data. The agent solves a meta-ePOMDP (for short, a "mePOMDP") by selecting a possibility from a possibility space and in effect, locating itself at this possibility. The representational capacities of the agent will determine the space of possible ePOMDPs it might consider. While adult humans have extremely general, compositional representational capacities that allow us to generate infinitely large and diverse spaces of possible problem formulations, other agents might be more constrained, and the ePOMDPs an agent considers in practice will also depend on the efficiency with which it navigates this space. The ability to center oneself by flexibly adopting useful problem representations thus falls on a spectrum.

Recentering, or changing the problem that is being solved, then involves switching from one ePOMDP and replacing it with another one, presumably in response to a breakdown in the match between the current ePOMDP and the world or the agent's goals. Here, we take the ePOMDP to be replaced, not merely refined. Rational agents may often update their models of a given situation in small ways, realigning themselves in space and time, like you did when you emerged from the Metro and used your GPS to update your location on your map. We take these kinds of small adjustments to be better described as ePOMDP "refinements" rather than as instances of recentering. As we will use "recentering" here, we take it to involve a replacement of the problem one is solving through substantial realignment and effective ePOMDP replacement, like switching plans to rent a bicycle instead of a taking the subway, or even more substantially, when discovering you are doing something you did not know you were doing, like Pooh did in the Woozle example.

How do we formalize the replacement of one ePOMDP by another on this account? From a functional perspective (see, e.g., Marr & Poggio, 1976; Oaksford & Chater, 2001; Tenenbaum et al., 2011), it is relatively straightforward to represent a switch from ePOMDP 1 (e1) to ePOMDP 2 (e2) as a rational decision. "Simply" do the following: define the space of all possible ePOMDPs,

calculate the expected values of e1 and e2 at time t, and select the ePOMDP with the higher value (see Figure 5).[2] The expected value of an ePOMDP can be decomposed (broadly) into the following considerations: (a) whether the ePOMDP is in line with relevant preferences, such that solving it would increase the utility of the agent, and (b) whether the ePOMDP correctly describes the situation, such that plans that the agent makes to solve the ePOMDP will be effective in the real world. Each ePOMDP may receive low or high marks for either consideration: Conditioned on the agent's independently held preferences, staying home to play chess may receive lower marks than picking up the kids. Of course, the agent may prefer being Julius Caesar at the battle of Alesia to both of the previous ePOMDPs, but that option does not adequately describe who, what, when, and where the agent is.

To help to illustrate our proposal regarding recentering, in Figure 6, we represent Pooh as an agent solving a particular ePOMDP e1 (in which the Woozle is creating the tracks) and then, after a point T, solving a different ePOMDP e2 (in which Pooh is creating the tracks). Note that each ePOMDP represents a space for learning, belief revision, and action, and the transition represents a radical departure, a switch of ePOMDP, in which possibilities involving belief, action, and agent spaces are revised. The engineering question that this framework aims to tackle is how to build an agent that can construct a space of ePOMDPs and choose to transition from e1 to e2.

The Woozle example illustrates the conceptual importance of recentering in three ways. First, the change in ePOMDPs entails a change of world model, that is, a change in the problem being solved, through a change in interpretation of the data (the individual responsible for the tracks), stemming from a switch of agent assignments (from "a Woozle" to "me"). The change of interpretation is substantial, stemming from Pooh's discovery that he is the causal source of the tracks, that is, Pooh also centers on a new representation of himself as agent. The correction here is not merely that he was mistaken about which agent was making the tracks: that more minimal sort of correction would have occurred if the switch had been from a Woozle to, say, a Wizzle. What is more substantial is that Pooh recenters in a way that, by moving to a new location in conceptual space, he switches to a model of the world where *he* is the source of the tracks, and thus discovers that there is no Woozle. Through recentering himself, he replaces his model of a Woozle world with one where he is the primary causal agent, and, importantly, one where he is the source of his own deception (for in-lab examples of misattributions of self action to other and vice versa, see, e.g., Wegner et al., 2003; Wegner & Wheatley, 1999).

Second, the Woozle example involves substantial self-correction, involving the reinterpretation of several variables. As part of his recentering, Pooh updates the variables in a way that backward corrects the previously held values. In this example of recentering, it is not merely that the world was X, and now it is Y. The agent also reasons that, from some point in time, the world was always Y, and grasps how many values need to be reset and updated accordingly, which is algorithmically challenging.

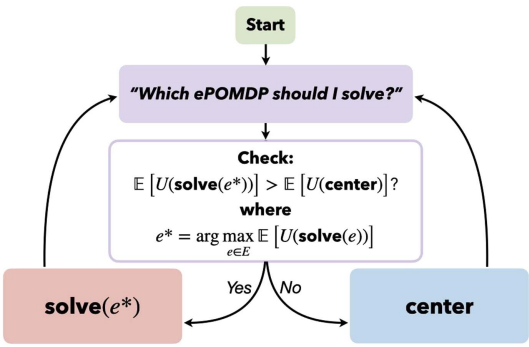
Finally, in this process of updating and backward-correction, the agent (Pooh) does not transition mindlessly from ePOMDP e1 to ePOMDP e2. Pooh chooses to switch e1 with e2, and represents that this replacement occurs. This is recentering oneself in a possibility space, with the resultant backward-correction involving recognizing he was wrong and updating his beliefs. Switching from e1 to e2 reflects Pooh's realization that he was mistaken about what he was doing and allows him to correct his mistake (if not necessarily to understand fully why he was mistaken, or what he was mistaken about). This type of "self-understanding" is, obviously, psychologically important.

Pooh discovers these things about himself and his world in virtue of the fact that he is acting at the meta-ePOMDP level to choose between representations of his world, and within a representation, is able to locate himself-as-agent, in particular, as the agent that caused the tracks. This second way of regarding himself is structurally similar to how, when playing a computer game, human agents regard potential avatars as representations of themselves-as-agents. (We will return to this idea in our discussion below.) In each case, choosing a new representation amounts to recentering oneself in a possibility space, switching from one ePOMDP to another.

A general schematic of an agent that solves a mePOMDP by centering and recentering is shown in Figure 5. Our next step is to introduce a concrete computational model that instantiates this general framework to illustrate and test the way that people act at the meta-ePOMDP level to center and recenter themselves.

Figure 5

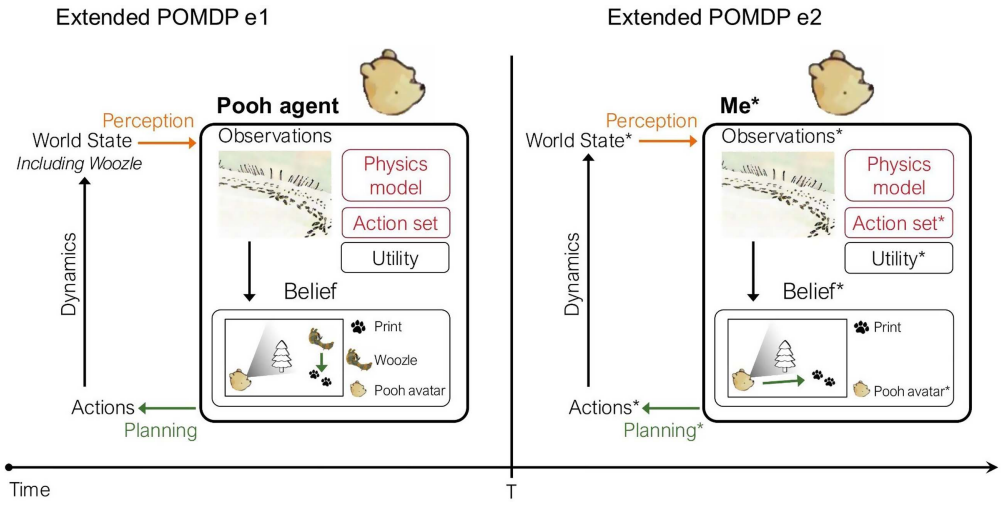
Schematic of a General Meta-ePOMDP Agent



Note. The agent starts by considering which ePOMDP to solve. To answer this question, it computes which ePOMDP, e^*, would be most valuable to solve out of the space, E, of ePOMDPs it represents. It then computes whether beginning (or continuing) to solve this ePOMDP has a higher expected value than beginning (or continuing) to center. If so, it makes progress toward solving e^*, while continuing to monitor its value relative to other ePOMDPs. If not, it makes progress toward centering, for example, by considering alternative ePOMDPs or gathering more information about its environment. Bold text denotes agent actions and decisions, while italic text denotes branch labels and mathematical variables and quantities. ePOMDP = extended partially observable Markov decision process. See the online article for the color version of this figure.

[2] We recognize that while we termed the representation above as "simple" in a conceptual sense, it is no simple task algorithmically to represent and reason over the full space of all ePOMDPs an agent could undertake to solve, even if centering and recentering can be captured in a representationally rich probabilistic programming language (see, e.g., Cusumano-Towner et al., 2019; Evans et al., 2018; Goodman et al., 2008, 2016), particularly those designed to model agents and their recursive representations of themselves and other agents (e.g., Chandra et al., 2025).

Figure 6
Schematic of ePOMDP Agents in Woozle World



Note. A simplified presentation of the Woozle world. Pooh is first solving a particular ePOMDP (e1). After time T (fitting his paw into the print) Pooh transitions to a different ePOMDP (e2), a world in which there is no Woozle, and in which various spaces of possibilities and functions have changed (indicated by a *). It is not just that Pooh's specific belief over the world state has changed, but that his space of possibilities for the world states and their relation to actions has changed: He chooses a different location in conceptual space. ePOMDP = extended partially observable Markov decision process. See the online article for the color version of this figure.

Testing the Framework With Concrete Meta-ePOMDP Models

We aim to further demonstrate and empirically validate the value of the meta-ePOMDP framework by using games that allow us to behaviorally observe key aspects of self-location and problem identification in agent-centered problem-solving. The first set of games, the avatar games, tests players' ability to perform centering and recentering as self-location tasks, by adopting and flexibly updating an agential perspective within an otherwise fixed problem. Our second set of games, based on the popular puzzle video game "Baba Is You" (Hempuli, 2019), tests players' ability to center and recenter themselves in conceptual space, by considering and selecting not only from possible perspectives within a problem, but from possible problems themselves. That is, to win the game, they must identify the right problem, which can involve changing the rules of the game as well as changing who they are in the game. For each set of games, we implement a model that plays the games in accordance with our theory by solving a mePOMDP, and compare its performance to human players as well as alternative models that do not engage in explicit centering.

Modeling Self-Location in Avatar Games

As we indicated above, in the Centering as Self-Location section, just as you must center yourself on the dot on the map to navigate Paris, you must center yourself on an agent to play (navigate) a video game. Here, we model the importance of centering oneself on a representation of oneself in an environment (on a dot or an avatar) in order to have agency within that environment. We interpret avatar identification in video games as a version of centering oneself by selecting an ePOMDP with a particular agent representation. In most games, this centering process is largely taken for granted, occurring quickly and internally to the player; for instance, during the prerace countdown in the four-player version of the game Mario Kart, players might look for their selected character or rev their engines to identify which corner of the split screen their avatar occupies. Here, we study this process directly by designing games that emphasize and reveal behavioral correlates of centering.

In the avatar games, an agent is placed in a simple grid-world environment where multiple characters move about, and the agent's keypress actions control a single character (the avatar). However, which character is the avatar is initially unknown and can also change during the game. Building on earlier theoretical work (Paul et al., 2023), De Freitas et al. (2023) used these games to analyze the role of self-representation in human and AI gameplay by comparing the performance of human and deep neural network-based reinforced learning (RL) agents on a range of game types which pose different challenges for avatar identification. The RL agents took orders of magnitude longer than human players to learn to successfully complete the games and to adapt to perturbations (such as adding an additional character) and failed to employ a robust and generalizable strategy for efficiently moving their avatar to the goal. In contrast, human players behaved in a way consistent with a two-stage hierarchical strategy: They first attempted to identify which character they could control, and only then navigated that character along an efficient path to a goal object, accounting for possible switches to which character they controlled in the middle of the game.

We build on, but go beyond, the work of De Freitas et al. (2023) in three key ways. First, we provide the theoretical foundation needed to explain the patterns in human behavior observed in that work: We

interpret the two-stage hierarchical strategy described by De Freitas et al. as preliminary evidence that humans center themselves on an avatar and recenter themselves as needed in order to play the game.

Second, we implement our theory in a computational model. While De Freitas et al. (2023) provided a set of simple models that qualitatively captured the observed two-stage strategies, those models were hard-coded, game-specific policies that were designed to implement the observed strategy in a manner tailored to the particular mechanics of each game, without an attempt to explain why people acted in this way or to derive game-specific policies from a more general rational-inference architecture. In contrast, the goal of our modeling here is to test whether we can explain human behavior in terms of a single, unified strategy that applies across all the avatar games and that explicitly captures the cognitive processes of agent-centered problem-solving, centering, and recentering that we posit as the basis for how people play these games. Our model takes the form of a (boundedly, see Simon, 1955) rational hierarchical inference and planning agent that operationalizes both agent-centered problem-solving and centering, with the former computation nested inside the latter, as an instance of the general mePOMDP architecture shown in Figure 5. That is, it solves a meta-inference or centering problem about which agent-centered model is best to use to play any given avatar game, given the game state and observational data the player confronts, and then uses that ePOMDP model to plan efficient actions it expects to win the game. It also monitors its own performance, so that it can recognize if and when its representation of itself as an avatar-based agent has become misaligned—as some of the avatar games are designed to do—and hence needs to recenter. More specifically, we present a meta-ePOMDP solving agent, or "mePOMDP agent," that plays each game by first (a) taking actions aiming to generate data for inferring (in an approximately Bayesian fashion) which character is its avatar by maximizing expected information gain (EIG). Thus centered on the appropriate ePOMDP, it then (b) goes on to solve this ePOMDP with shortest path planning that exploits the ePOMDP's action affordances. While doing so it monitors sense data to detect if and when it needs to recenter, by inferring whether that ePOMDP is still appropriately predicting the expected effects of its action and progress toward its goal, and if not, it infers the new ePOMDP most appropriate to replace it.

Third, and finally, we test human players and our mePOMDP agent on a much larger set of avatar games to empirically validate our theory. We test the mePOMDP agent in two forms: One is an ideal learner, and the other is a resource-limited version that incorporates cognitively plausible constraints on attention and memory. We also compare with a simple heuristic alternative that past work finds is much closer to human levels of efficiency than deep-RL algorithms (De Freitas et al., 2023), but that does not attempt to recenter or explicitly identify its avatar; rather it (optimistically) always tries to move the closest character to the inferred goal.

Below we describe the avatar games we used, give a formal account of the space of possible ePOMDPs and the way the mePOMDP agent works (in both optimal and resource-constrained versions) and present our empirical findings. We begin by recapping the four original avatar games used by De Freitas et al. (2023) and then move on to our larger set of games. In each level of each original game variant, there are four qualitatively identical characters and one goal (see Figure 7). One character is the avatar, which can be controlled by pressing the arrow keys. A level of the game ends when the avatar reaches the goal. Time moves one step forward at each keypress, and the goal is placed in the center. Human players were simply told to "use the arrow keys to play the game," without any further instructions, meaning they were given no instructions about the meaning of the squares, goals, and so on.

The four game variants were designed such that players need to exploit different types of cues in order to successfully identify the avatar. In the logic game (Figure 7A), only the avatar-square moves in response to keypresses. In the contingency game (Figure 7B), each nonavatar character moves in one direction (left–right or up–down) randomly sampled at the beginning of the level, one step per timestep. The switching mappings game (Figure 7C) is a variant of the contingency game in which the key–action mapping of the avatar itself is randomly sampled at the beginning of each level and must be inferred. The switching embodiments game (Figure 7D) is another variant of the contingency game, in which the avatar periodically and without warning switches to a new character entirely.

Meta-ePOMDP Implementation of the Avatar Games

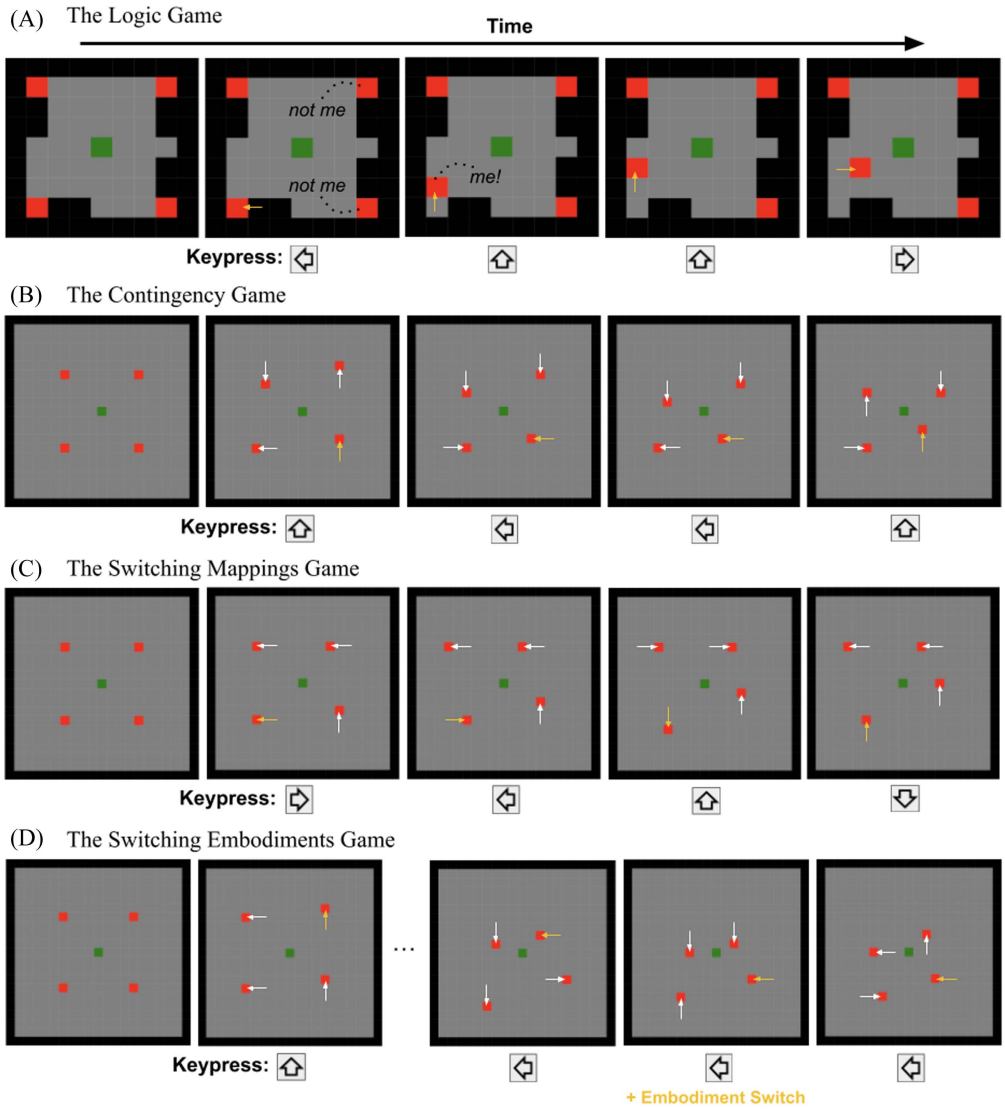
We now describe a model that implements the essential pieces of a meta-ePOMDP framework needed to solve the avatar games, including the four introduced in De Freitas et al. (2023) as well as several additional variants, and test its ability to explain signatures of human gameplay. Our mePOMDP agent constructs a space of possible ePOMDPs, centers on a particular ePOMDP in this space, attempts to solve this ePOMDP, and recenters when necessary by inferring probabilistically when it is likely that its current ePOMDP is out of date. In effect, our agent solves our computational version of the self-location problem of "Who am I (in this world)?"

Constructing a Space of ePOMDPs

The mePOMDP agent is designed to be the minimal agent that can solve the meta-ePOMDP formulation of the avatar games. Thus, its perceptual and representational capacities are meant to be fairly general within a limited game world, albeit far from general purpose in any real-world sense. It simply perceives the location of each object in the grid world as indicated by the cell color (e.g., walls, characters, goals) and can only represent worlds which fit the following assumptions:

1. Every object in the world is fully observable.
2. Black cells are walls, gray cells are open, red cells are characters, green cells are goals.
3. The avatar is one of the characters.
4. Characters cannot move through walls or each other.
5. The level ends when the avatar reaches the goal state.
6. At each timestep, the agent can take one of four actions, each of which specifies a movement in a unique direction.
7. At each timestep, the avatar will attempt the movement specified by the agent's action with probability $1 - \alpha$, and choose a direction of movement (or to stay put) uniformly at random otherwise.

Figure 7
Depiction of the Original Avatar Games



Note. Snapshots of the four avatar games as described in De Freitas et al. (2023). Arrows indicate the direction of the previously attempted movement by each character: white for nonavatar characters and yellow for the avatar. (A) The logic game. There are four characters (red blocks), one of which is the avatar. The avatar attempts to move in the direction indicated by arrow keypresses. Other characters do not move. The level ends when the avatar reaches the goal. Scenes are annotated with logical conclusions drawn based on keypresses and resulting character movements. (B) The contingency game. Here, nonavatar characters move randomly in one direction (left–right or up–down) sampled at the beginning of the level. (C) The switching mappings game. Identical to the contingency game, except that the mapping between arrow keypresses and avatar movements is sampled randomly at the beginning of the level. (D) The switching embodiments game. Identical to the contingency game, except that the avatar switches to a new character every seven steps (shown between the third and fourth frames). See the online article for the color version of this figure.

8. At each timestep, nonavatar characters will choose a direction of movement (or to stay put) uniformly at random, and attempt that movement with probability $1 - \alpha$.

9. At each timestep, there is some probability P_{switch} that the avatar switches to another character.

Here, α is a noise parameter set to 0.01. These constraints on the kinds of worlds the mePOMDP agent can represent limit the space of ePOMDPs it can consider. For example, the agent cannot consider an ePOMDP in which the avatar can break through walls. Though this implementation does not capture the full richness of human representational spaces, our goal is to demonstrate how a

minimal implementation of the meta-ePOMDP framework allows an agent to flexibly navigate the space of possible worlds it *can* consider, viz., ones involving differences in self-location.

Because the mePOMDP agent does not initially know which character is the avatar or which keypresses map to which actions, it considers a space of possible ePOMDPs (E), one for each possible agent realization. When the action mapping is known, this results in four possible agent realizations: one for each possible avatar. When the action mapping is not known, there are 96 possible agent realizations: one for each possible combination of avatar and action mapping. For the mePOMDP agent solving our simple meta-ePOMDP, constructing the space of possible ePOMDPs amounts to considering possible agent realizations in this way (i.e., instantiations of particular avatar identities and their action-mappings). Thus, we refer to ePOMDPs and agent realizations interchangeably from here onward.

Determining Which ePOMDP to Solve

We have proposed that the value of solving a particular ePOMDP should depend on (a) whether the ePOMDP is in line with relevant preferences, and (b) whether the ePOMDP correctly describes the situation. Here, all possible ePOMDPs are equally in line with the agent's preferences, because they all specify the same task: Navigate the avatar to the goal. Relative values thus depend only on the accuracy with which the ePOMDP represents the world. Solving the mePOMDP therefore reduces to inferring which ePOMDP describes the true agent realization (the avatar identity and action-mapping). This is performed with Bayesian inference: The mePOMDP agent starts with a uniform prior over possible ePOMDPs E ($e \sim \text{Uniform}(E)$). At each timestep t, the mePOMDP agent takes an action a, observes its consequence o_t, and updates the posterior probability distribution over possible ePOMDPs accordingly:

$$P_t(E|o_{1:t}, a) \propto P_{t-1}(E|o_{1:t-1}, a) \cdot P(o_t|E, o_{1:t-1}, a), \quad (1)$$

where $P_{t-1}(E|o_{1:t-1}, a)$ is the prior over ePOMDPs at timestep t, $P(o_t|E, o_{1:t-1}, a)$ is the likelihood of observing o_t after action a, and $P_t(E|o_{1:t}, a)$ is the posterior over ePOMDPs.[3] The action is selected to maximize an approximation to one-step information gain (see Appendix C).

The mePOMDP agent repeatedly acts, observes, and updates its distribution over possible ePOMDPs until one is confidently estimated to be more valuable than the others. At this point, the mePOMDP agent selects the most valuable ePOMDP to center on and solves it using relatively standard planning methods. For simplicity, we take this decision for selecting an ePOMDP to be based on a relative probability threshold, defined as the point where the posterior probability of one ePOMDP is at least some constant multiplicative factor higher than any other. Although other ways of selecting an ePOMDP might also be appropriate, this particular implementation choice is inspired by literature suggesting that human confidence in a given theory (in this case, a particular agent realization) is determined by its probability relative to other leading theories (Li & Ma, 2020). We call this factor the *confidence ratio* (or "confidence" for short), and in the remainder of this article, set it to a fixed value of 1.5, but we note that more sophisticated mePOMDP agents could incorporate an adaptive threshold (e.g., based on an agent's higher order preferences or beliefs).

Solving an ePOMDP

In our original set of games, solving a particular ePOMDP was quite simple: One must navigate the avatar to the goal. At each timestep, the mePOMDP agent computes the shortest path from the current location of the avatar to the goal using the A* algorithm, a standard pathfinding algorithm for finding the shortest path between two locations, and takes the first step in that path. Later, we will describe an extension of this planning module which generalizes to certain nondeterministic environments, and test our mePOMDP agent in that setting as well.

Recentering: Choosing a New ePOMDP to Solve

While solving a particular ePOMDP, the mePOMDP agent continues to monitor the value of each possible ePOMDP by updating the probability distribution over agent realizations at each timestep. It assumes there is some small probability, P_{switch}, that the character it controls might switch at any timestep. P_{switch} is set to the true switch probability of 1/7 for the switching embodiments game type and 0.01 for the other game types.

$$\begin{aligned} P_t(E|o_{1:t}) \propto\ & P_{\text{switch}} \cdot (P_{t-1}(E|o_{1:t-1}, \text{switch}) \cdot P(o_t|E, o_{1:t-1})) \\ & + (1 - P_{\text{switch}}) \cdot (P_{t-1}(E|o_{1:t-1}, \text{no switch}) \\ & \cdot P(o_t|E, o_{1:t-1})). \end{aligned} \quad (2)$$

If at any point the currently selected ePOMDP is no longer clearly the most valuable (as determined by the *confidence ratio*), the mePOMDP agent stops solving the selected ePOMDP and begins to recenter. Specifically, it stops trying to navigate the character it originally centered on toward the goal and instead begins to take informative actions that disambiguate which of the other agent realizations it should center on instead. A schematic of the full centering and recentering process is depicted in Figure 8.

Model Variants

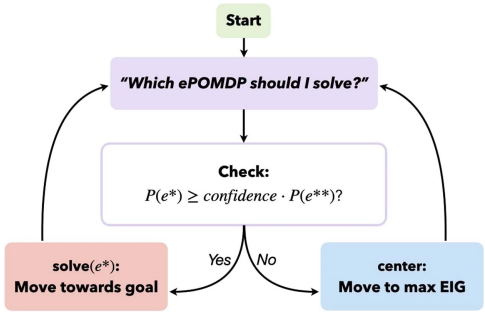
Before comparing the mePOMDP agent to human performance on the avatar games, we additionally propose two alternative models: a *resource-limited* variant of our mePOMDP agent that accounts for certain cognitive limitations in attention and memory likely faced by human players, and a *proximity heuristic* agent that implements a computationally simpler strategy that is weakly sensitive to some of the same cues as our agent but does not rely on the meta-ePOMDP framework.

Resource-Limited Model Variant

While we have introduced our mePOMDP agent to illustrate how people might flexibly and efficiently solve the avatar games through centering, this agent has certain cognitive capacities that make it an unrealistic model of human players. In particular, the mePOMDP agent attends to every one of the movements of each of the four characters and, when computing the posterior over agent realizations at each timestep, takes all of this information into account. Additionally, it has

[3] Note that the full posterior updates also accounts for the probability of switching to a new character (P_{switch}), see full equations (Equation 2) in "Recentering: Choosing a New ePOMDP to Solve."

Figure 8
Meta-ePOMDP Agent Schematic for the Avatar Games



Note. The mePOMDP agent starts by attempting to decide which ePOMDP would be most valuable to solve, and whether it should begin solving that ePOMDP or continue centering. In this particular implementation, the agent does so by checking if any ePOMDP is clearly more valuable than the others, as described in *centering* and *recentering* (here, e^* represents the most probable ePOMDP, and e^{**} represents the second most probable ePOMDP). The agent takes a solving action in e^* if so, and takes a centering action if not. When selecting a solving action in e^*, the mePOMDP agent navigates the avatar to the goal (as specified by e^*),. When selecting a centering action, the mePOMDP agent selects an action that maximizes its expected information gain over the relative value of ePOMDPs. Note that this is a specific implementation of the more general architecture shown in Figure 5. Bold text denotes agent actions and decisions, while italic text denotes branch labels and mathematical variables and quantities. EIG = expected information gain; ePOMDP = extended POMDP; mePOMDP = meta-ePOMDP; POMDP = partially observable Markov decision process. See the online article for the color version of this figure.

perfect memory of its beliefs about the relative probability of each possible ePOMDP, even in the *switching mappings* game (with 96 possible ePOMDPs). In this sense, this mePOMDP agent is "optimal" but probably too powerful relative to humans. We therefore implement a resource-limited mePOMDP agent that follows the same rational-inference framework but subject to cognitive constraints on both attention and memory which plausibly emulate those of human players.

This resource-limited variant simulates limits on human attention by attending to the movements of just one character at a time, where the attended character is randomly selected at each timestep; other character movements do not affect the belief update. The observations of the resource-limited mePOMDP agent at each timestep o_t are thus a subset of those of the full meta-ePOMDP, thereby limiting the quality of both the posterior updates (Equation 2) and the action selection (Equation 3).

In addition, in the *switching mappings* game in which players must learn new action mappings on each level that conflict with the typical (and for human players, highly overlearned) mappings of keyboard arrows onto movement directions, the resource-limited mePOMDP agent does not perfectly remember the computed relative probability of each ePOMDP. Instead, it partially forgets the distribution over action mappings for each possible character identity of the avatar, in favor of the typical action mapping (i.e., left arrow keypress causes the avatar to move left) that we assume will always be competing in human players' memories with the new mapping—and potentially interfering with learning it.

Specifically, at each timestep, the resource-limited mePOMDP agent updates its posterior over agent realizations based on its most recent observation and then, conditioned on each possible character identity c_i, takes the weighted average between the updated posterior over action mappings and a biased prior distribution P_0(mapping), which heuristically captures the anchored memory of the typical action mapping by making it 10 times more probable than any other. The relative weights of the two distributions are determined by a forgetting factor *ff*, which we initially set to 0.2, and fit to human data in later analyses. The resource-limited mePOMDP agent's posterior over action mappings for the given character identity is then replaced with this biased posterior:

$$P_t(\text{mapping}|o_{1:t}, c_i) = \begin{array}{l}(1-ff)\cdot P_{t-1}(\text{mapping}|o_{1:t-1}, c_i) \\ \cdot\, P(o_t|\text{mapping}, c_i)/P(o_t|c_i, o_{1:t-1}) \\ +\, ff\cdot P_0(\text{mapping})\end{array} \quad (3)$$

Given these limits on the model's attention and memory, we expect the resource-limited mePOMDP agent to solve each of the game types less efficiently than the optimal mePOMDP agent but at a rate more similar to human players.

Proximity Heuristic Agent

The proximity heuristic agent offers a way of playing the avatar games that does not involve centering and recentering as specified by our framework. Just like the mePOMDP agents, and unlike deep-RL agents, the proximity heuristic agent explicitly represents the world state, including the locations of walls, the goal, and each character. However, it does not attempt to infer which avatar it controls, nor even assume that it controls its avatar. Rather, it follows a heuristic policy of choosing at each timestep the character that is closest to the goal and attempting to move this character toward the goal using the same strategy as the mePOMDP agent.[4] To avoid certain failure cases, such as never attempting to move the avatar when it starts further from the goal than another character in the logic game, the agent takes a random action with probability $1-\epsilon$ at each timestep, with $\epsilon = 0.1$. See Appendix C for other specific details related to each game type.

Study 1: Modeling the Original Avatar Games

Method

We report the performance of our mePOMDP agent, its resource-limited variant, and the proximity heuristic agent for each of the four original avatar games, and compare it to that of the human players from De Freitas et al. (2023).[5]

We used two behavioral metrics: (a) the average number of steps taken to solve the game (solving time) and (b) the average number of steps taken to center on a particular avatar (centering time). While the RL agents tested in De Freitas et al. (2023) did not center and failed to capture the solving time of humans, we expected the optimal mePOMDP agent and its resource-limited variant to better account for these behavioral metrics both across and within games. First, we expected these two agents to capture the relative difficulty

[4] De Freitas et al. (2023) found that this heuristic outperformed the hand-coded baseline agents in the switching embodiments game, making it a simple and potentially promising alternative strategy for solving some types of avatar games.

[5] Data, study materials, and analysis code (for Studies 1 and 2) are available at https://osf.io/48gav/.

across games: Agents should take a longer time to center and solve games that pose greater challenges to humans. Second, we expected these agents to better capture the absolute values of our two behavioral metrics within any game: Their distributions of centering and solving time should be similar to that of human players. In all cases, while we expected both the optimal and resource-limited mePOMDP agents to capture aspects of human behavior, we expected the resource-limited variant would better capture human behavior than the optimal variant.

Participants and Models. De Freitas et al. (2023) recruited distinct sets of 18–20 human participants to play each game type. All participants were provided the minimal instruction to "use the arrow keys to play the game" before completing 100 randomly selected levels. Participants therefore had to learn the dynamics of each game as they played. For more participant details, see De Freitas et al.

Because we are interested in formalizing the ability to solve the avatar games *after* participants have learned the game dynamics, we only analyzed participant data after the average participant performance reached an asymptote in each game type, that is, from level 35 to 100 (cf. De Freitas et al., 2023). We ran each model with 20 random seeds for each game type to match the number of participants and computed solving times by averaging across levels within participants (for humans) or seeds (for models).

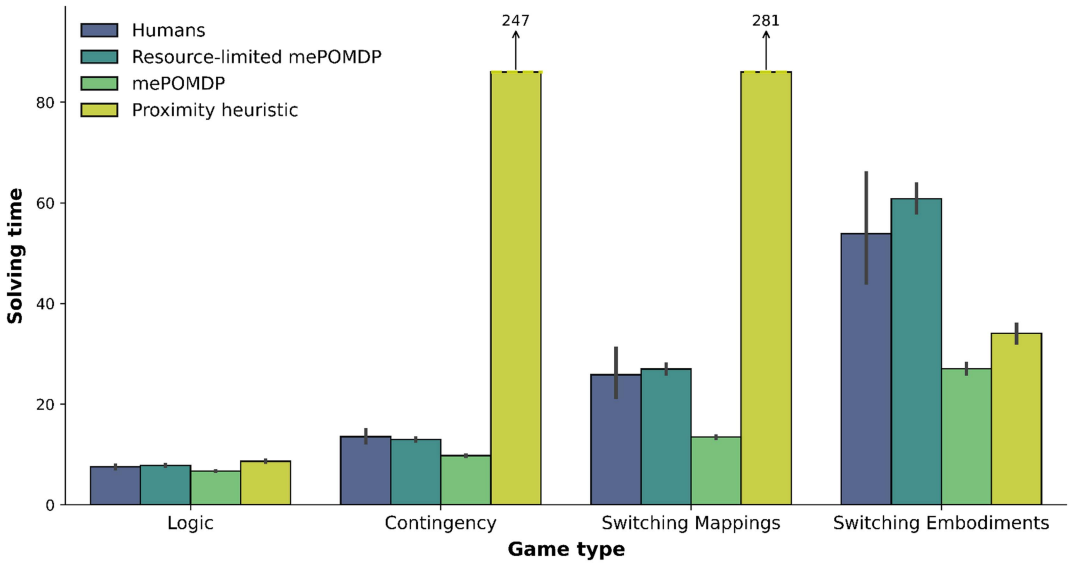
Results and Discussion

Solving Performance. Each of the models displayed noticeable differences in solving times across game types (Figure 9). Looking first at human performance, we see that the easiest game was the *logic* game, followed by *contingency*, *switching mappings*, and *switching embodiments*. The optimal mePOMDP agent and resource-limited variant showed the same ordering, whereas the proximity heuristic agent did not: Though the proximity heuristic agent also solved the *logic* game the fastest, its next fastest was the *switching embodiments* game, followed by *contingency* and *switching mappings*.

Both the optimal and resource-limited mePOMDP agents displayed similar behavior to humans in terms of the relative difficulty of the different game types. However, the optimal mePOMDP agent solved each game significantly faster than humans, whereas the resource-limited variant better matched human game playing times (Figure 9) and did not differ significantly from human solving times in any of the four game types (see Table C1 for *t* tests between model and human solving times in each game type). Qualitatively, while both variants captured the increase in solving time for the *switching embodiments* game, the resource-limited variant showed a sharper increase in line with human performance. This increase likely reflects the computational difficulty of repeatedly recentering, which requires first noticing that the currently selected ePOMDP should no longer be solved, then abandoning it in favor of a new one.

While the proximity heuristic agent did not strongly match the human pattern of performance across game types, one exception is in the *logic game*, in which all models and humans performed similarly well. The heuristic strategy works well in this game type; since only the avatar moves, and each character starts equidistant from the goal, once the proximity heuristic agent happens to successfully move the avatar it will continue doing so. However, unlike

Figure 9
Solving Times for the Original Avatar Games (Study 1)



Note. The average number of steps each model took before completing a level. We plot the steps taken to complete each level, both for human participants and randomly initialized seeds of the models. Error bars show 95% confidence intervals. Error bars are omitted for the proximity heuristic in the contingency and switching mappings game types. mePOMDP = meta-extended partially observable Markov decision process. See the online article for the color version of this figure.

the other models, the proximity heuristic's solving time increased dramatically in the more complex *contingency* and *switching mappings* game types. Interestingly, and analogously to De Freitas et al.'s (2023), the proximity heuristic solved the *switching embodiments* game much faster than humans and the resource-limited mePOMDP agent. The fact that human players could have employed this computationally simpler proximity heuristic, yet behaved more similarly to the resource-limited mePOMDP agent which centers on the inferred agent realization, demonstrates the default human tendency to engage in centering—even when it is not always strategically advantageous to do so.

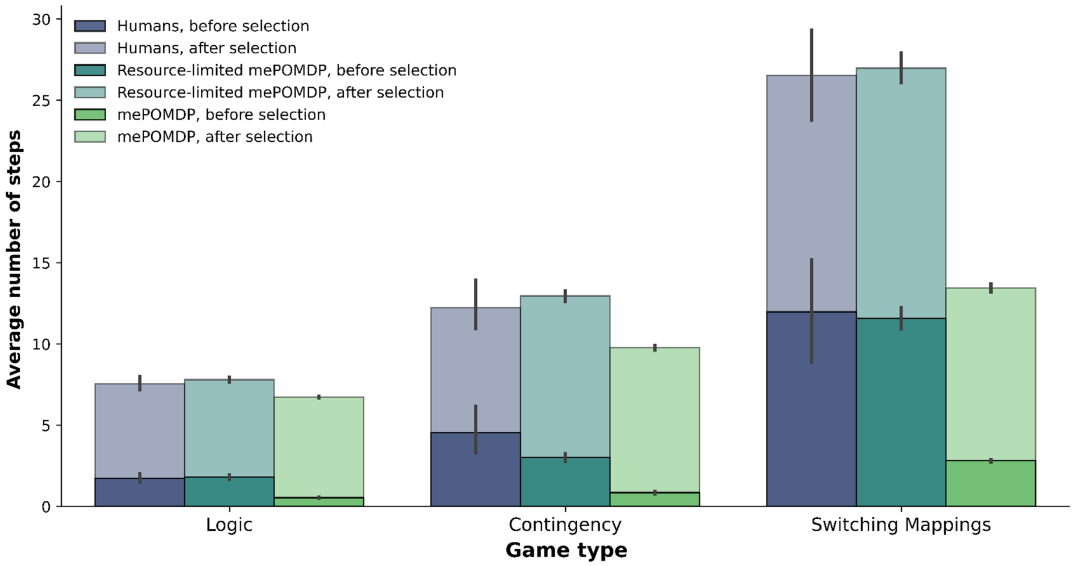
Centering Performance. To further analyze the dynamics of human game play, De Freitas et al. (2023) examined the point at which participants identified their avatar in different game types. For the *logic* game, in which only the avatar moves, they took this to be the point at which the avatar first moved. For the more complex *contingency* and *switching mappings* game types, De Freitas et al. ran a follow-up study in which they asked participants to explicitly report when they identified their avatar. Two distinct sets of 20 participants were recruited and again tasked with playing 100 levels of their assigned game type. This time, however, participants were additionally instructed, "In each level, as soon as you find which square you are controlling, please click on that square. You can only click one time per level." This explicit instruction to identify the avatar did not significantly affect the average number of steps taken to solve the game between the original and new participants, when comparing levels 35 (the earliest point after which the original participants' performance plateaued across all game types) to 100, *contingency*: $M_{\text{original}} = 13.55$, $M_{\text{new}} = 12.23$, $t(38.0) = -1.27$, $p = .211$, $d = 0.40$; *switching mappings*: $M_{\text{original}} = 25.84$, $M_{\text{new}} = 26.54$, $t(38.0) = 0.24$, $p = .813$, $d = 0.08$.

We compared the number of steps after which humans and the mePOMDP agents selected an avatar (centered) in these games. For humans, selecting an avatar corresponded to successfully moving the avatar (*logic* game) or clicking the avatar square (*contingency* and *switching mappings* games). For the mePOMDP agents, selecting an avatar corresponded to choosing an ePOMDP to solve (transitioning from centering to solving an ePOMDP as described in *centering*). As seen in the dark bars of Figure 10, the optimal mePOMDP agent consistently selected an avatar more quickly than humans on average: *logic*: $M_{\text{model}} = 0.54$, $M_{\text{human}} = 1.74$, $t(19.7) = 8.11$, $p < .001$, $d = 2.56$; *contingency:* $M_{\text{model}} = 0.85$, $M_{\text{human}} = 4.55$, $t(19.1) = 5.00$, $p < .001$, $d = 1.58$; *switching mappings*: $M_{\text{model}} = 2.82$, $M_{\text{human}} = 11.98$, $t(18.0) = 5.30$, $p < .001$, $d = 1.74$. In contrast, the resource-limited variant better matched human behavior, and the difference in centering time between them was not significant: *logic*: $M_{\text{model}} = 1.82$, $M_{\text{human}} = 1.74$, $t(38.0) = -0.47$, $p = .644$, $d = 0.15$; *contingency*: $M_{\text{model}} = 3.02$, $M_{\text{human}} = 4.55$, $t(20.0) = 2.04$, $p = .055$, $d = 0.64$; *switching mappings*: $M_{\text{model}} = 11.59$, $M_{\text{human}} = 11.98$, $t(19.5) = 0.22$, $p = .826$, $d = 0.07$.

These results offer preliminary evidence that human behavior in the avatar games might be well described as solving a meta-ePOMDP in a resource-limited approximation. Although both the optimal and resource-limited mePOMDP agents capture human patterns of solving performance across game types, the model with resource limitations on attention and memory additionally captures solving and centering performance within games.

Figure 10

Centering Times for the Original Avatar Games (Study 1)



Note. The average number of steps before selecting an avatar and before completing a level. Humans selected an avatar by successfully moving the character (*logic* game) or by clicking on the character square (*contingency* and *switching mappings* games). The optimal mePOMDP agent and resource-limited mePOMDP agent selected an avatar by transitioning from centering to solving an extendedPOMDP. We plot the average number of steps taken to complete each level, averaging across participants for the human data and across random seeds for the model data. Error bars show 95% confidence intervals. mePOMDP = meta-extended partially observable Markov decision process. See the online article for the color version of this figure.

Study 2: Generalization Across Additional Game Variants

To provide further evidence for our hypothesis that human gameplay is best explained as solving a meta-ePOMDP under certain resource constraints, after replicating the experiments of De Freitas et al. (2023), we developed a larger and more diverse set of avatar games, creating eleven new games in total. This larger suite of games enables us to test the robustness of our models when centering becomes more challenging in various ways, such as when there are more characters to track, noisier mappings between keypresses and observable actions, and when there is uncertainty in the environment.

We expected participants to have similar relative solving times across the original four game types to those in De Freitas et al. (2023), though perhaps with lower solving times within game types due to the introduction of more explicit instructions and a step limit being in each game type. Importantly, we also expected both mePOMDP agents, but not the heuristic agent, to account for human solving times across the eleven game types, with the resource-limited variant doing so best.

Method

We included versions of the four original game variants from De Freitas et al. (2023) and also introduced seven new game variants. The *Contingency2*, *Contingency6*, and *Contingency8* games are identical to the *contingency* game but include two, six, and eight possible avatars instead of four. The *noisy contingency* game is identical to the *contingency* game, except that the avatar moves randomly with probability ⅓ at each timestep instead of responding to the player's keypress. The *contingency + goal uncertainty* and *switching mappings + goal uncertainty* games are variants of the *contingency* and *switching mappings* games, respectively, which have three decoy green goal squares in addition to the true goal, requiring players to resolve environmental uncertainty regarding where the true goal is in order to solve the game. Finally, the *switching embodiments (every 10)* game is a variant of the *switching embodiments* game in which the avatar switches realizations every 10 moves instead of every 7. Each of the 11 game variants, including the original 4, allowed a maximum of 150 moves per level before the level would restart, with the number of remaining moves displayed to participants. Correspondingly, in our analyses, we cap model solving moves on each level at 150.

Model Extension. The mePOMDP agents are straightforwardly extended to play each of the new game types. In the *noisy contingency* game type, the noise parameter α is set to 1/3, the ground truth noise level in this game, instead of 0.01 as we used for the noise-free games previously. In the *switching embodiments (every 10)* game, P_{switch} is set to the new embodiment switch probability of 1/10, analogously to how it was previously set to 1/7 in the original *switching embodiments* game when the avatar switched realizations every seven steps. Finally, in the game types with multiple possible goals, solving each ePOMDP now requires the model to represent uncertainty over which goal is the true goal on a given level. This form of partial observability is natural to incorporate in the core POMDP formalism. While the mePOMDP agents in Study 1 simply navigated to the goal when solving a particular ePOMDP, here they monitor the probability that each green square is the true goal, starting with a uniform prior and updating their beliefs at each timestep, based on their assumption that the game level ends only when the avatar reaches the true goal. The estimated probability that a green square is the true goal therefore decreases when the agent reaches the square, and the level does not end. When solving an ePOMDP, the mePOMDP agents navigate the avatar to the currently most probable goal, using proximity as a tiebreaker.

Human Participants. We recruited distinct sets of 20 participants from Prolific to play 40 levels of each of the 11 game variants. We thus recruited 220 participants in total (M_{age} = 38.3, 52% female, 1.8% nonbinary), who took 14.4 min to complete the study and were paid $14.43/hr on average. Since we are interested in formalizing how people play these games *after* having learned the game dynamics, participants were told the rules of the game before playing (unlike in De Freitas et al., 2023). For example, participants playing the *switching embodiments* game were told:

> You will soon begin playing a game using the arrow keys on your keyboard. There are 40 levels. On each level, you will control a red square with the arrow keys. The square you are controlling may change within a level. You win the level by moving to a green goal square before your moves run out.

This approach allows us to compare human performance more fairly to that of the mePOMDP agents, which make inferences and decisions based on knowledge of the true game dynamics or "rules." It also allows us to compare human performance with and without explicit instructions, to see whether this made a difference in the original experiments. As in Study 1, we only analyzed participant data after the average participant performance reached an asymptote within each game type (cf. De Freitas et al., 2023), from level 16 to 40.

Model Parameters. The resource-limited agent has several parameters, meant to capture particular cognitive constraints, that we earlier set to fixed values to model our original four avatar games. We hypothesized that allowing two of these parameters to vary might better describe human behavior across our new and larger range of games. Specifically, we now fit the action mapping forgetting factor *ff* to human data, along with an additional attention parameter $P(att)$ which allows the resource-limited agent to attend to more than one character at each timestep. Each additional character is attended to with probability $P(att)$, with unattended characters having no effect on the belief update as before. When $P(att) = 0$, the resource-limited mePOMDP agent is at its most limited and attends to only one character at each timestep; this is the version tested above for our original four games. When $P(att) = 1$, the resource-limited mePOMDP agent attends to all characters at each timestep as in the optimal mePOMDP agent, and intermediate values of $P(att)$ between 0 and 1 interpolate between these two agents.

We preregistered an analysis to fit these two model parameters in a way that could reliably estimate their contributions to the agent's overall performance and be expected to generalize to new, held-out participants (https://osf.io/5tcze/). We fit parameters over 1,000 runs of Monte Carlo cross-validation, finding the maximum-likelihood values of $P(att)$ and *ff* using grid search over the range of 0 to 1 for $P(att)$ and 0 to .8 for *ff* (as higher values of *ff* made learning prohibitively slow) in steps of .05. On each cross-validation run, human participants were randomly and evenly split within each game type between a train and test set, and we selected the parameter

combination which maximized the likelihood of the human solving times in the training set under the resource-limited agent. We then computed the held-out likelihood of solving times for participants in the test split under both the best fitting resource-limited agent and the optimal agent, providing us with a distribution across runs of likelihoods of human data under both agents, as well as likelihood-maximizing parameters.

Results and Discussion

We compared average solving times of the optimal and resource-limited mePOMDP agents and human participants on the 11 game types. In a preliminary analysis, we first asked whether the distribution of human solving times across the original four game types replicated the findings of De Freitas et al. (2023). We found very similar solving times in three of four games and the same overall ordering of solving times as in Study 1, with average solving time increasing across games as follows: *logic*, *contingency*, *switching mappings*, *switching embodiments* (Figure 11). However, participants in the replication had significantly shorter solving times in one game, *switching embodiments*, $M_{\text{old}} = 53.9$ (22.4), $M_{\text{new}} = 37.4$ (7.07), $t(20.0) = 2.91$, $p = .009$, $d = 0.99$, which may be explained by the clearer instructions and fewer levels played in the replication. We also continued to see this difference when capping solving times from De Freitas et al. at 150 steps, the maximum number of steps allowed per level in the new experiments: $M_{\text{old}} = 48.8$ (13.7), $t(36.0) = 3.18$, $p = .003$, $d = 1.03$.

Based on our preregistered analysis, we then asked whether the resource-limited agent provided a significantly better fit to human solving times on heldout test data than the optimal agent, where $ff = 0$, $P(att) = 1$, and if so, which parameters for the resource-limited agent fit behavior most robustly. Across 1,000 cross-validation runs, the mean estimated difference between the log likelihood of held-out human data under the resource-limited agent with fit parameters and under the optimal agent was 15,323 (95% bootstrapped confidence interval: [15,111, 15,535], where a confidence interval including 0 would fail to indicate that the log likelihoods under the two agents differed significantly). The parameter combination which most often maximized likelihood across cross-validation runs was $ff = 0.4$, $P(att) = .05$ (see Appendix C for average likelihoods under alternative parameter combinations). Under this parameterization, the resource-limited mePOMDP agent has a .05 probability of attending to each character in addition to the single randomly selected character at each timestep in each game type. Additionally, in the *switching mappings* game types, at each timestep its posterior over action mappings is averaged with a distribution biased toward the typical action mapping, with weights 0.6 and 0.4, respectively. We used this parameterization of the resource-limited agent for the remainder of our analyses.

We then compared human solving times across game types to that of the mePOMDP agents and the proximity heuristic. Across the 11 game types, mean solving times for both the resource-limited and optimal mePOMDP agents correlated very well with mean human solving times (Figure 11, $r = 0.96$, $p < .001$, and $r = 0.90$, $p < .001$, respectively). In contrast, the proximity heuristic agent did not significantly correlate with human solving times (Figure 11, $r = 0.16$, $p = .634$).[6] Additionally, we computed the absolute difference in solving times on each game type between each model and humans and ran paired t tests comparing these differences between each model pair. The optimal mePOMDP agent differed less from humans than the proximity heuristic agent did, that is, the difference in absolute differences was significant, $t(10.0) = 4.13$, $p = .002$, $d = 1.73$, as did the resource-limited mePOMDP agent, $t(10.0) = 4.35$, $p = .001$, $d = 1.98$. Within the mePOMDP agents, the resource-limited variant differed less from humans than the optimal variant, $t(10.0) = 2.87$, $p = .017$, $d = 0.92$. Thus both mePOMDP agents account for trends in human solving times across game types (with the resource-limited variant fitting best), while the proximity heuristic agent does not.

Qualitatively, both mePOMDP agents accounted for specific patterns in human solving times across game types (Figure 12), such as the increase in solving time across the original four game types (*logic, contingency*, *switching mappings*, *switching embodiments*), as in Study 1. They also accounted for the increased difficulty of the new game types due to goal uncertainty (*contingency* vs. *contingency + goal uncertainty*, and *switching mappings* vs. *switching mappings + goal uncertainty*), additional characters (*contingency* vs. *contingency6* vs. *contingency8*), and environmental noise (*contingency* vs. *noisy contingency*).

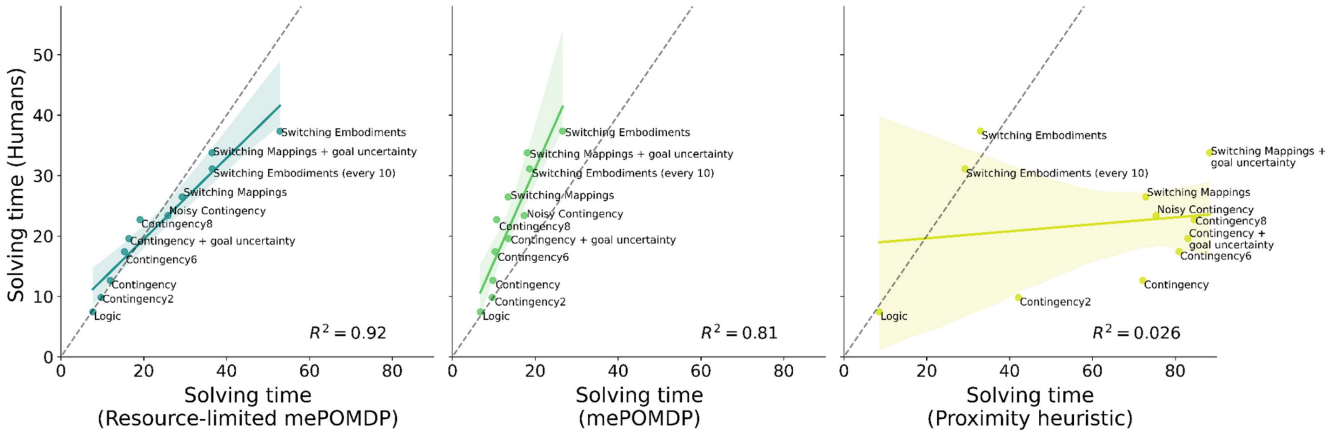
At the same time, and as already suggested by the correlation results, the resource-limited agent captured differences in human solving times that the optimal agent did not. For example, humans required fewer steps to complete the contingency game with two characters than the original version with four characters, $t(38.0) = -4.07$, $p < .001$, $d = 1.29$. While this difference was mirrored by the resource-limited mePOMDP agent, $t(38.0) = -15.2$, $p < .001$, $d = 4.81$, it was absent in the optimal mePOMDP agent, which performed similarly in both games, $t(38.0) = -1.39$, $p = .174$, $d = 0.44$, suggesting that limited attention better accounts for human solving times between games. The resource-limited mePOMDP agent also captured the particular difficulty that humans had with the *switching mappings* games, in which the optimal mePOMDP agent tracks 96 different ePOMDPs to determine the most likely action-mapping, and memory constraints likely factor into human gameplay. For example, while humans took significantly longer to solve the *switching mappings* game than the *contingency + goal uncertainty* game, $t(38.0) = 2.14$, $p = .039$, $d = 0.68$, as did the resource-limited mePOMDP agent, $t(23.2) = 11.4$, $p < .001$, $d = 3.60$, the optimal agent did not, $t(38.0) = -0.09$, $p = .928$, $d = 0.03$. Similarly, humans and the resource-limited agent took significantly longer to solve the *switching mappings + goal uncertainty* game than the *noisy contingency* game $t(38.0) = 6.44$, $p < .001$, $d = 2.04$ and $t(38.0) = 7.54$, $p < .001$, $d = 2.38$ respectively, unlike the optimal agent, $t(38.0) = 1.48$, $p = .148$, $d = 0.47$. Thus, by implementing cognitive constraints on attention and memory, the resource-limited agent accounts for additional nuances in human solving times across the different game types that the optimal mePOMDP agent "misses."

In fact, the resource-limited agent even matched absolute human solving time well within most game types (see Figure 12). In nine of the 11 game types, t tests between average model and human solving times across game types did not allow us to reject the null hypothesis

[6] Recall that we capped human and model solving times at 150. When removing this cap for model solving times, these results still hold: Mean human solving times correlated strongly with the optimal mePOMDP agent's solving times ($r = 0.90$, $p < .001$) and with the resource-limited mePOMDP agent's solving times ($r = 0.95$, $p < .001$), but not with the proximity heuristic agent's solving times ($r = -0.098$, $p = .776$).

Figure 11

Comparison of Study 2 Solving Times by Game Type, Between Humans and Models

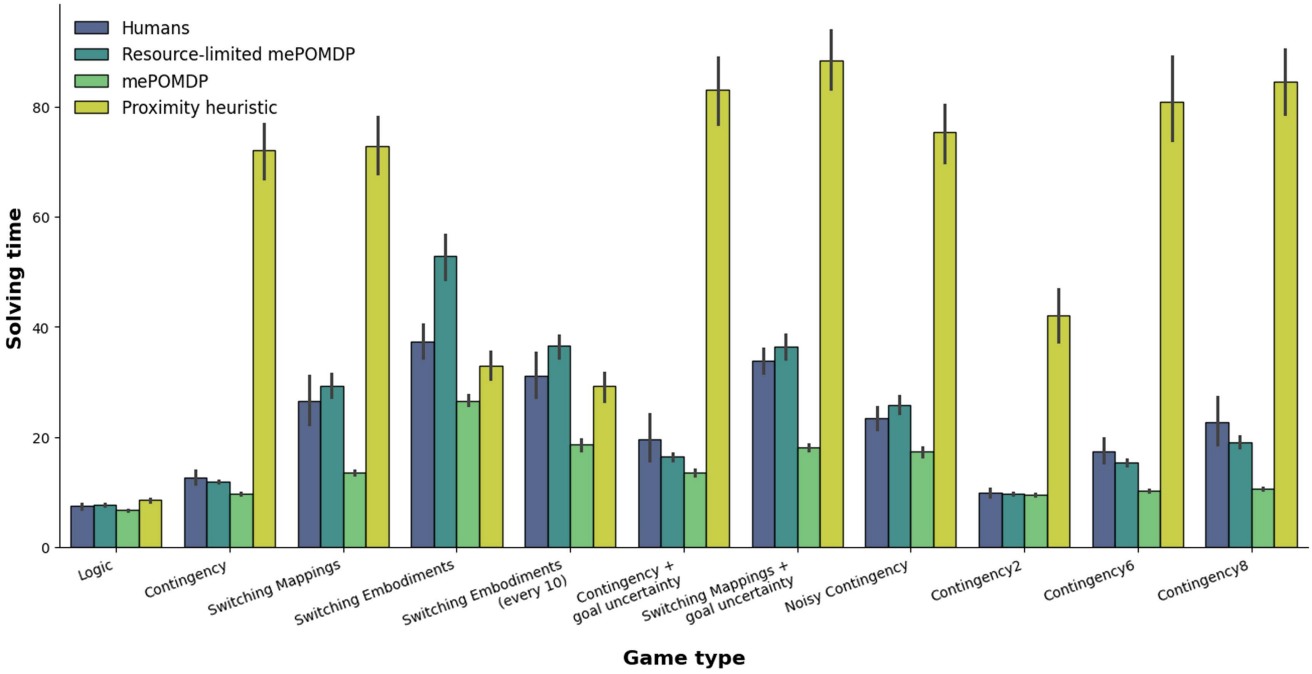


Note. Scatter plots comparing average solving times within each game type, between humans and the resource-limited mePOMDP agent (left), the optimal mePOMDP agent (middle), and the proximity heuristic agent (right). The number of steps per level for both humans and models is capped at 150. Gray dashed lines indicate $y = x$. mePOMDP = meta-extended partially observable Markov decision process. See the online article for the color version of this figure.

that the means of the distributions underlying the two samples are equal (see Appendix C for test results with .05 significant levels). The two exceptions were the *switching embodiments* and *switching embodiments (every 10)* game types, in which people solved the games in fewer steps than the resource-limited agent. One potential explanation for this difference is that participants recognized the particularly taxing nature of these games and so attended to more characters simultaneously. Indeed, human data for this game type were most often best fit across cross-validation runs when $P(att) = .35$ (see Appendix C), meaning that nonselected characters are attended to with probability .35, rather than .05, as fit best across all game types. Participants might also have had a lower confidence

Figure 12

Solving Times for the Extended Set of Avatar Games (Study 2)



Note. The average number of steps humans and models took before completing a level. We plot the average number of steps taken to complete each level, averaging across participants for the human data, and across random seeds for the model data. The number of steps per level for both humans and models is capped at 150. Error bars show 95% confidence intervals. mePOMDP = meta-extended partially observable Markov decision process. See the online article for the color version of this figure.

threshold for centering on a particular realization than our model; since the avatar realization changes regularly, it may be beneficial to begin attempting to move the avatar to the goal before being sure of its identity. Finally, participants might have incorporated other strategies not fully captured by our simple meta-ePOMDP solver, such as having a bias for attending to or optimistically moving characters which are closer to the goal, as better accounted for by the proximity heuristic strategy. That said, while the proximity heuristic is indeed a simple and effective strategy that does not involve centering, human solving times on the *switching embodiments* game are not well described by this model, $t(38.0) = -2.14$, $p = .039$, $d = 0.68$.

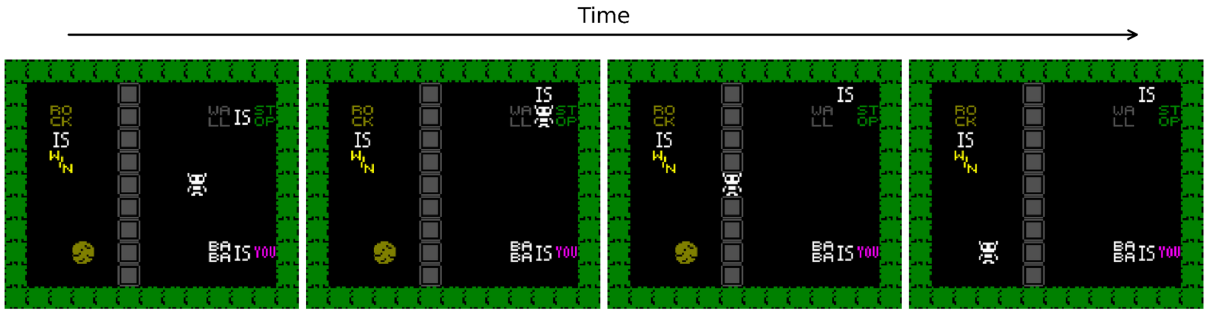
Modeling Problem Identification in "Baba Is You"

With the avatar games we demonstrated how our framework can capture people's tendency to center and flexibly recenter themselves with respect to a problem, identifying the location from and manner in which they can exert their agency to determine "Who am I (within this world)?" A key aspect of our theory not captured by these games is that people solve problems by centering themselves in a more abstract manner as well, by representing and selecting from a broader space of possible problem specifications in which to locate themselves. This conceptual task requires people to realize that, for example, the rules that govern their world might be different from what they originally thought they were. We described this above as the task of centering as problem identification: People solve the meta-problem of "which problem am I solving" by centering on a particular problem specification. When they recenter, they choose a new problem specification, perhaps one with different rules for governing the world. In the second set of games, we will discuss, demonstrate, and test this aspect of our theory by studying the way people change the problem they are solving using a set of games inspired by the puzzle video game "Baba Is You" (Hempuli, 2019).

In these games, players navigate a 2D grid world by controlling an avatar and trying to reach a goal object. The rules of the game are spelled out by words within the grid. For example, the game shown in Panel 1 of Figure 13 has three rules: *Baba Is You* (the titular "Baba" is the player's original avatar), *rock is win* (the goal object is a rock), and *wall is stop* (the avatar cannot move through walls). Notice that given these rules, the game appears to be unwinnable: An impassable wall separates the avatar, Baba, from the goal, the rock. To win the game, the player must recognize that they need to change the problem they are solving. They must reason not only *within* the rules of the game but *about* the rules of the game, by taking advantage of a unique twist in the game mechanics: The player can change the rules by interacting with the game elements that specify them. For example, if the player manages to misalign the three blocks with the words "wall," "is," and "stop," the rule no longer applies (see Figure 13). Players can thus change the rules of the game as they play by pushing around these "rule blocks" with their avatar. Winning the Baba Is You games requires determining what possible ruleset, or problem specification, will make the game solvable, and instantiating that ruleset, before solving that particular problem. By requiring players to physically construct the problem they wish to solve, these games usefully highlight and externalize the generally internal, cognitive process of centering in conceptual space.

The flexible problem-solving skills required to play Baba Is You make it a compelling challenge for AI systems (Charity & Togelius, 2022). Evaluations of recent large language model (LLM) agents have shown that they struggle to propose actions that will change the rules in useful, context-appropriate ways, especially in more complex games (Cloos et al., 2024; Paglieri et al., 2024); though Ahmed et al. (2025) found that integrating LLMs with more structured reasoning and abstract representations of rules in a neurosymbolic architecture improves performance. While this work suggests the usefulness of hierarchically reasoning about possible rulesets, it is unknown to what extent humans play in this way. Here, our goal is not simply to develop an agent that plays Baba Is You games but to further demonstrate our framework and test whether it can explain signatures of human problem-solving revealed by these games. Below we introduce the game environments we used in our study, describe a mePOMDP agent that

Figure 13
Example Baba Is You Game Progression



Note. Selected snapshots of a single participant's gameplay on Game Type 5. The initial game rules include *baba is you, wall is stop,* and *rock is win* (Panel 1). This means that the player controls the white baba object, cannot move through the gray wall objects, and must reach the brown rock object to win. The participant breaks the *wall is stop* rule (Panel 2), moves through the wall (Panel 3), and reaches the rock object to win the game (Panel 4). Graphics from "Baba Is y'all: Collaborative Mixed-Initiative Level Design," by M. Charity, A. Khalifa, and J. Togelius, *2020 IEEE Conference on Games (CoG)* (pp. 542–549), 2020, Institute of Electrical and Electronics Engineers. CC BY 4.0; Charity et al. (2022). See the online article for the color version of this figure.

plays the games, and present empirical findings comparing human gameplay to the mePOMDP agent as well as a noncentering alternative approach.

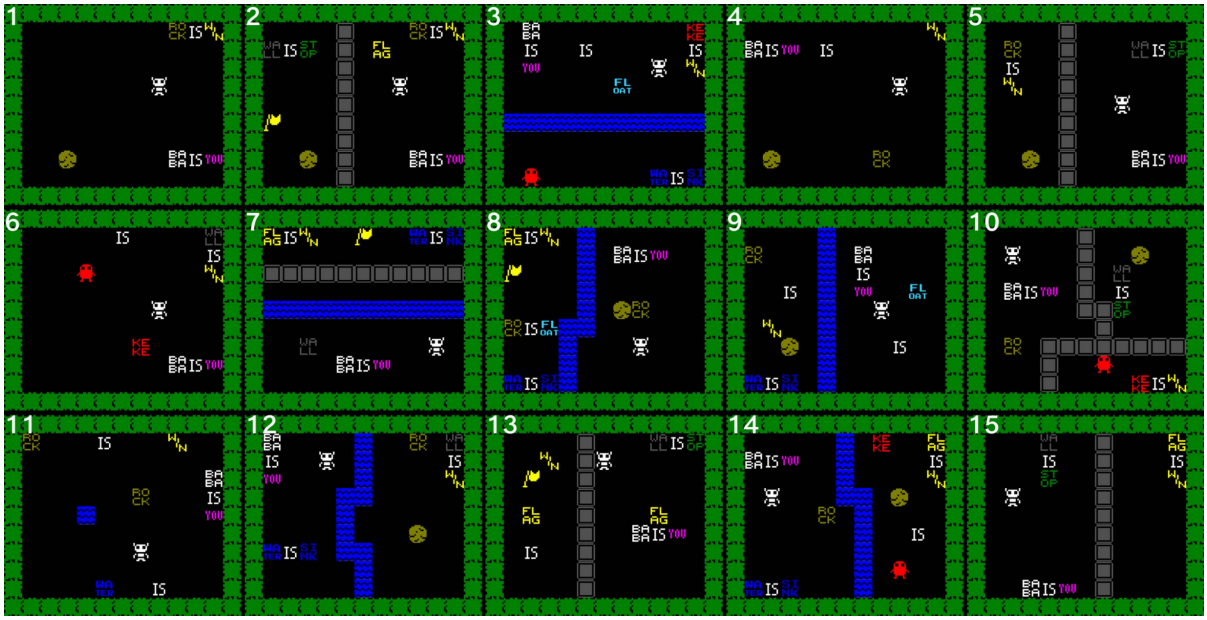
We designed and implemented a variety of Baba Is You games, building on the game implementation developed by Cloos et al. (2024) and the interface from Charity et al. (2020, 2022). In the Baba Is You games we tested, rules work as follows: Rules are always made up of three words, with one word per rule block, arranged adjacently left to right or top to bottom. The first word in a rule is an object type (baba, keke, flag, rock, water, or wall). Objects of each type may or may not be present in any given game, for example, the game shown in Figure 13 contains a baba object and a rock object. The second word in a rule is always "is." The third word in a rule can either be a property (you, win, stop, sink, or float) or another object type. If an object type *is you*, the player can control objects of that type. If an object type *is win*, the avatar must reach an object of that type to win the game. If an object type *is stop*, other objects cannot move through objects of that type. If an object type *is sink,* objects that cannot float will disappear if they collide with that object type. If an object type *is float*, objects of that type will not sink. Finally, if an object type *is* another object type, for example, "flag is rock," all objects of the first type are transformed into objects of the second type.

We designed 15 unique game types that highlight different aspects of centering by requiring players to change the rules in a variety of ways (Figure 14). For example, while Game Types 2 and 5 are superficially similar, their solutions are quite different; in Game Type 2 players must change their avatar by pushing the "flag" rule block into the "baba" rule block to simultaneously form *flag is you* and break *Baba Is You*, while in Game Type 5 players must change the fact that they are in a world in which they cannot move through walls by breaking *wall is stop.* We also varied the number of rule changes required to make the game solvable: Game Type 1 does not require any rule changes, as the player simply must move baba to the rock object, while Game Types 2–8 require one rule change, and Game Types 9–15 require two rule changes. For each of the 15 base game types, we designed a high distractor variant of the game, with additional distracting rule blocks and objects that are not necessary for solving the game (these game variants are shown in Figure D1).

Meta-ePOMDP Implementation of Baba Is You

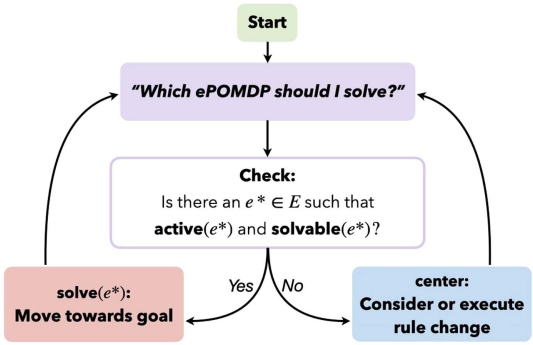
We now describe a candidate model of how people play the "Baba Is You" games. As in the avatar games, the mePOMDP agent constructs a space of possible ePOMDPs, centers on a particular ePOMDP in this space, and attempts to solve this ePOMDP, recentering when necessary (see Figure 15). In doing so, the mePOMDP agent figures out what problem to solve by jointly specifying which of many possible worlds it might be located in (and who it is in that world)—analogously to Pooh solving his problem of who was leaving the tracks. Though this agent and the mePOMDP agent that solved the avatar games are both instances of the more general mePOMDP framework (Figure 5), they differ

Figure 14
All Baba Is You Game Types



Note. Snapshots of the 15 base Baba Is You games as seen by human participants. Different games are separated by a green hedge border, which acts as a barrier in all games. Participants saw one game at a time. Numbers at the top left corner of each game are used to refer to that game type throughout the text. In each game, the words (rule blocks) specify the rules when arranged three in a row and can be manipulated by the player. Graphics from "Baba Is y'all: Collaborative Mixed-Initiative Level Design," by M. Charity, A. Khalifa, and J. Togelius, *2020 IEEE Conference on Games (CoG)* (pp. 542–549), 2020, Institute of Electrical and Electronics Engineers. CC BY 4.0; Charity et al. (2022). See the online article for the color version of this figure.

Figure 15
Meta-ePOMDP Agent Schematic for Baba Is You



Note. The agent begins to decide which ePOMDP to solve by checking whether E, the space of ePOMDPs it has considered, contains an ePOMDP that is both solvable and currently active (instantiated by the game). If so, it takes an action toward solving this ePOMDP e^* by moving the avatar object to the goal object. If not, it continues centering by either continuing to search through the space of possible ePOMDPs for a solvable one (by checking whether it could execute a particular rule change and evaluating the resulting ePOMDP), or executing a previously considered rule change to construct a solvable ePOMDP. Bold text denotes agent actions and decisions, while italic text denotes branch labels and mathematical variables and quantities. ePOMDP = extended partially observable Markov decision process. See the online article for the color version of this figure.

in the specifics of solving a particular ePOMDP, the process by which they center on an ePOMDP, and in the decision procedure for transitioning between solving and centering.

Constructing a Space of ePOMDPs

The mePOMDP agent perceives the locations of all game elements, including all objects, rule blocks, and the game border. It identifies the active ePOMDP as determined by the initial set of active rules, which may specify the agent's avatar embodiment (e.g., *baba Is you*), the agent's goal (e.g., *flag is win*), qualities and abilities of the avatar or other objects (e.g., *baba is float*), and other key features of the world (e.g., *wall is stop*). The agent constructs a space of alternative possible ePOMDPs by considering what other rulesets might be formed from the rule blocks present in the game.

Determining Which ePOMDP to Solve

When outlining our general framework, in *a computational formalization of centering and recentering*, we proposed that the expected value of solving a particular ePOMDP might be decomposed into (a) whether the ePOMDP is in line with relevant preferences, and (b) whether the ePOMDP correctly describes the situation. In the avatar games, the relative values of ePOMDPs largely depend on (b), as they vary primarily with respect to the accuracy with which they represent the world. In contrast, in the Baba Is You games, ePOMDPs may vary substantially in (a), the degree to which they are in line with relevant preferences. Here, players should prefer to solve ePOMDPs that are easier to solve, or ones that are at least *solvable*. A solvable ePOMDP is one in which the player can move the avatar to a goal object. In all game types other than Game Type 1, players are initially presented with an unsolvable ePOMDP and must change the ePOMDP to a solvable one in order to win. Different ePOMDPs might also vary with respect to (b), their alignment with what is achievable in the world of the game, as some ePOMDPs can be constructed by the player while others cannot. For example, just as one might fruitlessly wish to be Julius Caesar at the battle of Alesia, a player in Game Type 2 might wish to solve the ePOMDP in which baba can move through walls—but this is not a useful ePOMDP to entertain because the player cannot initially break the rule *wall is stop*.

The mePOMDP agent's goal is thus to find and construct an ePOMDP that is solvable (aligned with relevant preferences) and achievable. Since constructing a new ePOMDP requires the agent to take actions in the game, this amounts to a planning problem: At a high level, the agent must find a sequence of rules to break or form to construct a solvable ePOMDP. To do so, the mePOMDP agent runs a breadth first search (BFS) over sequences of rule-changing actions.

The mePOMDP agent searches through rule-changing actions and their resulting game states, adding new game states to a frontier to be processed, until it finds a game state that meets the termination criterion. A game state meets the termination criterion if the ePOMDP instantiated by that game state is solvable. During search, the agent considers possible ways of changing the rules from each game state in the frontier given the rule blocks present in the game, by either forming a possible rule that is not currently active or breaking a rule that is currently active. It checks whether each of the possible actions from each state is achievable, first considering rule changes that might result in a solvable ePOMDP (one in which a goal object and an avatar object are at least present in the game, a prerequisite for being able to move the avatar to the goal). Within each partition, it considers actions in a random order. To check if a particular rule change is achievable, the mePOMDP agent calls a low-level subroutine that runs BFS over low-level block pushing actions: When its goal is to break a rule, it attempts to push one of the three rule blocks in that rule away from the other two, and when its goal is to form a rule, it attempts to push the three rule blocks next to each other. This low-level subroutine is essentially a sequence of calls to an A* pathfinding algorithm which the agent uses to check whether it can move to a certain location or push a certain block to a certain location. Each time the agent checks whether it can achieve a rule change with the low-level subroutine it calls this pathfinding algorithm several times. The agent searches through the space of possible rule changes until it finds an achievable and solvable ePOMDP or exhausts the state space. If it finds an appropriate ePOMDP, the agent then constructs the ePOMDP by executing the corresponding sequence of rule changes.

Solving an ePOMDP

As in the avatar games, solving the selected ePOMDP is quite simple. The agent solves the game, holding the rules as fixed. At each timestep, it simply computes the shortest path between the avatar object and the goal object using the A* algorithm and takes the first step in that path.

Recentering: Choosing a New ePOMDP to Solve

When solving an ePOMDP, if the agent realizes that the active ePOMDP is no longer solvable (it cannot find a path between the avatar and a goal object), it begins the centering process again by restarting the

BFS algorithm. If the algorithm explores the entire state space without finding a solvable ePOMDP, the agent can select the "reset" action which resets the game to its initial state. Though all games are initially solvable, the agent may make the game unsolvable through its actions (e.g., by accidentally sinking an important rule block in the water).

Model Variants

Under our theory, we expect people to solve the Baba Is You games by centering: Representing and reasoning about possible ePOMDPs to solve, or rulesets that could be formed. However, there are other ways that these games might be played that do not involve centering. We describe one of particular theoretical importance below.[7]

Flat Solver

To solve the Baba Is You games, an agent might plan its actions not by considering possible rulesets, but by only considering more low-level actions, such as what arrow key to press next, or, at a slightly higher level of abstraction, where to move the avatar or what block to push where. Such an agent need not distinguish between constructing an ePOMDP and solving an ePOMDP; it simply takes actions at one "flat" level of abstraction—until the game has been solved. It might still change the rules of the game by moving rule blocks to new locations through a series of keypresses or block pushes, but it does not represent the space of possible actions in terms of possible rule changes. Rule changes happen as a byproduct of its actions, rather than as the goal of its actions.

We implement a version of this type of flat solver, which searches through a space of actions at a level of abstraction below that of rule changes. Specifically, the actions it considers include moving the avatar to specific locations and pushing blocks to specific locations. In order to highlight the most theoretically interesting differences between the flat and mePOMDP agents, we build in several additional domain-specific biases to the flat agent. Since the actions required to win each game include moving the avatar to the goal and breaking and forming rules (by pushing rule blocks next to or away from other rule blocks), the agent only considers avatar movements that move the avatar to the locations of goal objects, and only considers pushing rule blocks to locations adjacent to another rule block or to a single one of the many locations away from all other rule blocks.

The flat agent runs the same general BFS algorithm as the mePOMDP agent but through a lower-level search space. Since the flat agent does not go through the two-stage process of centering and then solving an ePOMDP, it simply terminates search once it has found a way to win the game (a game state in which the avatar has reached a goal object). The possible actions it considers taking from a given state include moving the avatar to a goal object (if present) and pushing rule blocks to new locations (adjacent to another rule block or away from all other rule blocks). When checking whether each of these possible actions is achievable in order of preference, it partitions actions into more and less preferable actions, where more preferable actions result in solution states in which the avatar has reached a goal object. Within each partition, it considers actions in a random order. To check if a particular action is achievable, the flat agent calls the A* pathfinding algorithm a single time.

Although the mePOMDP and flat agent will both eventually arrive at the solutions to all of the Baba Is You games we tested and often take identical actions in their ultimate solutions, they differ in the way that they arrive at those solutions and thus in the relative costliness of solving different games. While the flat agent searches through actions at a single level of abstraction below that of rule changes until the game is solved, the mePOMDP agent considers changing the ePOMDP by changing the rules, invoking a lower level planner to test whether the rule change can be made and whether the resulting ePOMDP would be solvable. The flat agent thus offers an alternative to the distinctively hierarchical, modular approach of the mePOMDP agent.

Study 3: Modeling Baba Is You Gameplay

Method

We test whether we can explain human gameplay of Baba Is You as centering in conceptual space by comparing human participants to the mePOMDP and flat agents on 30 different Baba Is You games.[8] The different games afford different spaces of possible rulesets, some of them quite large, and require different types and numbers of rule changes in order to be solved (Figure 14). Since these are puzzle games, often with unique solutions, we focus not on the specific actions taken by humans or models but on the underlying process by which solutions are found. Accordingly, we evaluate the models based on how well they capture the relative difficulty of solving each game. For human players, we use time spent solving each game as a proxy for difficulty. For the models, we derive an analogous metric by measuring the total number of low-level pathfinding calls made during search. These calls form the core computational primitive of search in both models, though they are structured differently: The flat agent performs one call per considered action, while the mePOMDP agent performs several calls within the low-level subroutine for each considered rule change.

We assume that for humans, checking whether a path exists between objects is cognitively cheap and is performed in close to constant time for the short paths within our 11 × 13 Baba Is You grids. We therefore use the number of pathfinding calls as a common unit of solving time between the models, and expect that if humans solve the games by searching through a similar space of actions to that considered by either model, then the relative number of pathfinding calls made by that model should predict the relative difficulty of games as reflected in human solving times.[9]

Participants. We recruited 61 participants from Prolific to play 15 Baba Is You games, with one distractor variant of each game type randomly assigned to each participant to play in a random order. We excluded participants who gave up on more than 25% of games, leaving us with 50 participants (23–27 participants per game).[10] Before playing the games, participants completed a tutorial in which they learned the game mechanics and practiced changing rules of

[7] While for the avatar games, we also included a resource-limited mePOMDP agent with constraints on attention and memory, we do not implement an analogous model here, since the game state is fully observable at all times, and players do not need to attend to or recall changes in the environment.

[8] This study was not preregistered.

[9] Data, study materials, and analysis code are available at https://osf.io/48gav/.

[10] All analyses replicate and are statistically significant in the same direction without this exclusion.

each type. After passing a comprehension test, participants completed each of their 15 assigned games. Participants used arrow keys to move their avatar and could click a reset button to reset the game at any point. They moved on to the next game only after winning the current game or giving up after a minimum of 10 min. For our analyses, we thus cap all human solving times at 10 min. Participants saw a timer that tracked their solving time as they played and were told that they would receive a bonus of $0.25 for each game they solved faster than half of other participants. The median time to complete the experiment was 74 min, and mean compensation was $13.88.

Models. We ran the mePOMDP and flat agents on each of the 30 games and recorded the number of pathfinding calls made during search. When computing whether the avatar can move to a particular location, or whether the avatar can push a block to a particular location, both agents call the pathfinding algorithm once, unless the existence of a path is automatically ruled out by any of three logical rules. First, it is impossible to move or push a block to a new location if there is no avatar; second, it is impossible to push a block away from a border to which it is adjacent; and third, it is impossible to push a block that was determined to be unreachable by the avatar during a previous pathfinding call from the current game state. We include this basic game-specific logic to make the agents more realistic models of human players.

The flat agent is deterministic apart from the order in which it checks whether actions are achievable, and so we ran this model once on each game and analytically computed the expected number of pathfinding calls before finding a solution. The mePOMDP agent has additional randomness in the order in which low-level block pushes are considered in the low-level subroutine, and so we estimated the expected number of pathfinding calls before finding a solution for each game by computing the mean across 50 random seeds.

Results and Discussion

Human participants had highly variable average solving times across the 30 games, solving the low distractor variant of Game Type 1 most quickly, in about 15 s on average, and the high distractor variant of Game Type 14 most slowly, in about 5 min and 50 s on average (including six of 24 participants who gave up after 10 min). Similarly, the mePOMDP agent took the fewest pathfinding calls to solve the low distractor variant of Game Type 1 (in which no rule changes are required), and the most pathfinding calls to solve the high distractor variant of Game Type 14 (in which two rule changes are required, and the space of possible rulesets is quite large). More generally, across games, the number of pathfinding calls made by the mePOMDP agent strongly predicts human solving times (Figure 16A, $r = 0.75$, $p < .001$). In contrast, the number of pathfinding calls made by the flat agent when solving each game does not correlate significantly with human solving times (Figure 16A, $r = 0.16$, $p = .393$). As shown in Figure 16A, the flat agent finds some games far costlier to solve than others, in a way that does not scale with human solving times. The flat agent takes the most pathfinding calls to solve the high distractor variants of Game Types 9 and 11, both of which require the agent to push four different rule blocks to new locations before going to the goal object. In Game Type 9, the agent must form the rules *baba is float* and *rock is win* before going to the rock object, and in Game Type 11 the agent must form the rules *rock is win* and *water is rock* before going to the rock object. Though these games are challenging for humans as well, with average solving times in the top third of all games, solving times fall more in distribution with other game types with distracting game elements and multiple rule changes (the high distractor variants of Game Types 9–15)—a pattern captured by the mePOMDP agent. The fact that the mePOMDP agent captures patterns in absolute human solving times suggests that people may be searching through a similar state space when solving the Baba Is You games—specifically, reasoning about alternative possible ePOMDPs.

We also assess how well each model accounts for relative human solving times—even if not their absolute scale—by comparing human solving times and model pathfinding calls on a log scale. The correlation between the mePOMDP agent and humans is strong (Figure 16B, $r = 0.82$, $p < .001$). While the flat agent also shows a positive relationship with human solving times (Figure 16B, $r = 0.65$, $p < .001$), this relationship disappears when controlling for the predictions of the mePOMDP agent in a partial correlation ($r_{\text{partial}} = 0.063$, $p = .744$). In contrast, there is a strong partial correlation between humans and the mePOMDP agent when controlling for the predictions of the flat agent ($r_{\text{partial}} = 0.66$, $p < .001$). This suggests that the mePOMDP agent explains variance in human solving times that the flat agent does not, while the flat agent explains human solving times only insofar as its predictions correlate with those of the mePOMDP agent.

It is not surprising that log-scaled pathfinding calls might be correlated between the two agents, since changing more rules will generally require pushing more rule blocks, and games with more rule blocks will also generally afford larger spaces of possible rulesets. However, the relative costliness of solving different games for the two agents diverges in certain cases. For example, while the high distractor variant of Game Type 12 requires players to break the rule *water is sink* and form the rule *rock is wall* in order to win, this can be done by moving a single rule block before going to the goal. This makes it one of the least costly games for the flat agent, but one of the most costly games for the mePOMDP agent, which uniquely captures the relatively high solving times of human participants in this game.

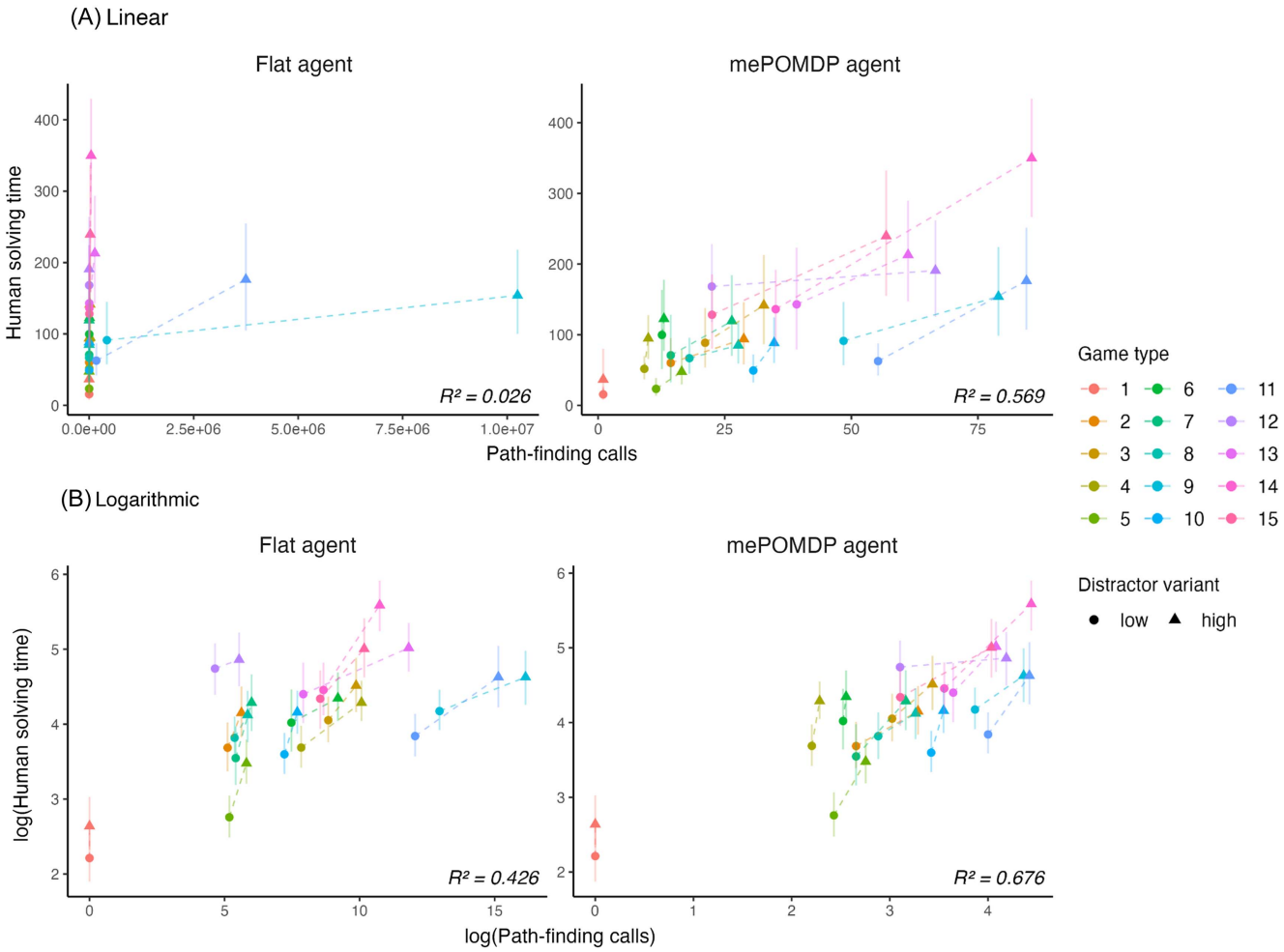
In summary, the number of pathfinding calls made by the mePOMDP agent when solving different games strongly predicts absolute human solving times and explains substantial variance in human solving times not explained by the flat agent even when log-transforming the variables to allow for differences in scale. Though the flat agent employs domain-specific abstractions and plans rationally over sequences of high-level actions, it does not explicitly construct and reason about a space of possible ePOMDPs, and thus, the results suggest, fails to capture patterns in human solving times explained by the mePOMDP agent. These results provide evidence that people, like the mePOMDP agent, solve the Baba Is You games by centering in conceptual space: representing and orienting themselves within a space of possible problems to be solved and identifying a particular problem to solve as the way to win the game (*this problem is the one I need to solve*).

General Discussion

In this article, we develop a conceptual and empirical framework for a distinctive kind of *centered* problem-solving where people

Figure 16

Comparison of Model and Human Solving Times (Study 3)



Note. Mean human solving times versus the mean number of pathfinding calls made by each model for each game and distractor type (A: linear scale, B: log–log scale). Error bars indicate 95% confidence intervals for human solving times. Low distractor variants of each game type are shown as circles and high distractor variants as triangles. The two distractor variants of a given game type are connected by dashed lines. mePOMDP = meta-extended partially observable Markov decision process. See the online article for the color version of this figure.

solve problems from their perspective as an agent embedded within but distinct from a given environment, and describe this type of problem-solving as relying on the ability to "center" oneself. Our work is inspired by the many interesting insights and discussions of the philosophical debate over centering and perspectival representation, a debate focusing on how, in certain contexts of inquiry, people need to determine which possible world is actual, and to represent themselves as occupying a certain position in that world in order to understand themselves as effective agents.

Our empirical studies test a computational model of this type of perspectival representation and problem-solving. We explore centered problem-solving in terms of self-location, where people center themselves on representations of themselves in an environment and solve problems from their perspective as agents within their environment. We also explore centered problem-solving in terms of problem identification, where the agent faces the meta-problem of "which problem do I need to solve?" as a version of identifying which kind of possibility they inhabit, including what type of world they need to model, and interpreting this as identifying the particular problem they need to solve from their perspective. The idea combines a novel philosophical interpretation of centering (as deciding which problem to solve in a space of possible problems) with a rigorous computational framework. Finally, we describe recentering as involving the agent-centered ability to change which particular problem one is solving (either by changing who, where, or when one locates oneself in a world or by changing which kind of world one takes oneself to be in).[11]

[11] As noted above, we take recentering to be changing which problem one is solving only when the conceptual realignment is substantial. Merely adjusting your position on the map by a few points is simply refining your problem rather than replacing your problem, even if in broader philosophical and linguistic contexts it can be called "recentering."

A major contribution of this article is interdisciplinary and theoretical: We bring the techniques and insights of computational cognitive science to bear on a central philosophical debate in modal epistemology about the nature of "de se" or "self-involving" attitudes and argue that the capacity to center and recenter is a key computational building block that human intelligence relies on to solve perspective-dependent problems.

Perhaps most importantly, we also develop and test a novel thesis within this interdisciplinary framing: The thesis that self-location can be interpreted as a meta-problem of deciding which problem to solve, understood as specifying a representation of the world and of ourselves within the world. In this way, we provide an empirical underpinning for philosophical explorations of centered world reasoning while making a new philosophical move: we interpret the meta-problem of "which problem do I need to solve?" as a problem of locating yourself on a map of conceptual possibilities. Our work opens up an entirely new way to explore the thesis that

> [T]he goal of inquiry is to determine which possible world is actual, which possible world we are in. Even stronger, for beliefs and desires to motivate us to action, they must not merely represent the world, but represent us as occupying a certain position in the world. Perception essentially presents the world as received from a perspective. (Cappelen & Dever, 2013, p. 1)

Turning to our empirical contribution, in the first set of studies, we investigated how people center on an avatar in a controlled computer game setting as a simplified model of how they locate themselves within a world—that is, how they gain agency by aligning with a representation of themselves embedded in an environment. We find that our agent-centering models, unlike alternatives, perform all tasks at humanlike levels of efficiency and also predict relative task difficulty, consistent with our hypothesis that humans use similar inferential mechanisms to determine what problem they are solving.

Building on our earlier work suggesting that centering and recentering on self-representations underlies human behavior in dynamic, uncertain avatar-based games (De Freitas et al., 2023; Paul et al., 2023), we formalized this capacity using a principled Bayesian framework. We developed the meta-ePOMDP (mePOMDP) agent, which selects among possible ePOMDPs—each encoding an environment, a goal, and an embodied self-representation—based on Bayesian inferences drawn from observations and background knowledge. This allows the agent to flexibly center and recenter in response to changes in self and world.

To test the model, we introduced a novel and diverse suite of avatar games. The optimal mePOMDP agent solved these tasks in a manner that closely mirrored the distinctive patterns of human performance across all eleven game types. By contrast, a heuristic agent that did not center or recenter solved some games efficiently but failed to generalize. A resource-limited variant of the mePOMDP agent, incorporating plausible constraints on attention and memory, provided the best overall fit to human behavior: It reached goals in a comparable number of steps and centered at rates matching those of participants explicitly reporting when they had found their avatar.

In our second set of studies, we focus on engineering a computational model for the task of recentering with respect to the problem being solved. That is, we propose a computational model for how people solve the meta-problem of "which problem do I need to solve?" in conceptual space. Interpreting a choice of problem specification, as we argued above, as a choice of a conceptual possibility, we explore, using a variant of the game Baba Is You, how people swap one ePOMDP for another as a means of swapping one problem specification (with one set of rules) for a different problem specification (with a new set of rules). Our mePOMDP agent solves 30 different Baba Is You games, which afford different spaces of possible rulesets and require different ways of changing the rules, in a way that captures patterns in human solving times. In contrast, human solving times were not well-explained by the flat agent, a strong alternative approach which solves the games without centering by reasoning only about lower-level actions such as moving or pushing blocks to particular locations.

We take our results to provide support for our thesis that centering and recentering allows people to flexibly realign themselves with the world in order to solve a new problem, and further, that this can be captured as rational inference and planning in an appropriate hierarchical model spanning the ePOMDP and meta-ePOMDP levels.

However, our work leaves open several directions for future study. While the avatar and Baba Is You games highlight key behaviors of interest, they differ from real-world problem-solving in many respects. For example, in Baba Is You, possible problems are determined by the set of available actions in the game, whereas people generally face the more difficult task of generating their own candidate problems from a much larger and only implicitly defined space of possibilities. Accordingly, we emphasize that the models we tested should be taken as prototype instantiations of (only some aspects of) the broad theoretical framework we are proposing in this article; many of the most interesting components of the theory have not yet been implemented.

In short, what we have built illustrates the core inference machinery for centering and recentering, but is likely neither necessary nor sufficient for self-centering. A fully general computational model for how people center and recenter themselves in order to identify and solve new problems, for all the problems that human beings identify and potentially solve this way, remains a much larger project for future work. In this future work, we see great promise in building on recent hybrid architectures (e.g., Ahmed et al., 2025; Wong et al., 2023, 2025) for cognition that combine distributed representations and language-based associative learning (such as those instantiated in large language models) with explicit, structured hypothesis generation, planning and probabilistic inference like we have used here.

We would now like to close our discussion by suggesting that centering is intuitively related to "having a self," and, in this sense, propose that our computational models provide a framework for reverse engineering a centered self, in the sense of reverse engineering a type of perspectival thought and representation intuitively describable as "having a point of view." Below, we describe a way that our model of self-location in terms of avatar selection could function as a natural model for the way people center on themselves as physical, embodied agents in the real world, by relying on proprioception and their first-person perspective to "select" their physical bodies as the representation of themselves. We conclude with open directions and broader questions.

Our Bodies, Our Selves

We suggest that humans exploit their ability to center on themselves as physical, embodied agents in the real world in a way that parallels the way they center on an avatar when playing a computer game. We take this proposal to amount to the idea that, in

real life, human agents perform a type of computational task that is isomorphic to what they do when choosing an avatar.

In our avatar games, people gain agency through solving the avatar problem, that is, by centering themselves on an avatar. In the real world, we suggest that people "solve" a similar problem, gaining agency by centering themselves on their physical bodies, by triangulating their perceptual and proprioceptive inputs with different third-person representations of themselves in order to represent themselves as an agent in the world. In each situation, the problem to solve is: "Who am I?" As with locating oneself on a map or solving the avatar problem, perhaps locating yourself in this sense involves a sophisticated computational feat that, in ordinary circumstances, is so effortless that one does not even realize a problem is being solved.

Set in this context, the way we use an avatar to represent ourselves in first-person games where the camera is aligned with an avatar's first-person perspective is of particular interest (see Figure 17, especially D).

In the first-person avatar task, the "line of sight" and other first-person sensory inputs presented to the player represent the perceptual and proprioceptive "within-the-boots" inputs of the avatar. These inputs are taken by default (by the game) to define the avatar's first-person perspective in the game.

A player can solve the avatar problem in first-person games by triangulating changes in these perceptual "within-the-boots" inputs with causal manipulations of a particular avatar. To do this, they can center themselves on their avatar by locating themselves "within" the avatar, taking on the first-person perspective of the avatar as their "own" (even if conceptually at one remove, i.e., while knowing it is their perspective as taken within the game). They then coordinate this perspective with different third-person representations of the avatar. The task can be performed in a variety of ways: For example, the player can link to their avatar by coordinating the given avatar-first-person-point-of-view with representations such as different "over the shoulder" camera angles on the avatar or with "bird's-eye" visual inputs of the avatar's location on the map (Paul, 2016).

By triangulating the perceptual and proprioceptive "within-the-boots" inputs of the avatar with different third-person representations of the avatar, a player represents themselves, to some extent, as virtually embodied by this avatar (as it is located in the virtual space). Interestingly, people can exploit this structure in order to explore new domains; for example, they can explore the environment of a video game through centering themselves on an avatar and using their virtual embodiment to explore the environment represented in the game. Such an experience is familiar to anyone who has played a first-person game or used a virtual reality headset.

As we propose at the start of this section, we think that in real life, people may solve a version of this problem by relying on proprioception and their first-person perspective to "select" their bodies as the representation of themselves. In particular, we suspect people triangulate their perceptual and proprioceptive inputs with different third-person representations of themselves to construct a representation of themselves, and then (usually) center it on their physical body.[12] As we think of it, people coordinate first-person representations of themselves from a "within the boots" point of view with third-person representations of themselves, aligning these to center themselves on their bodies as agents. From this embodied perspective, people can orient and reorient themselves as they shift their perceptual and cognitive center to new locations in new spaces.

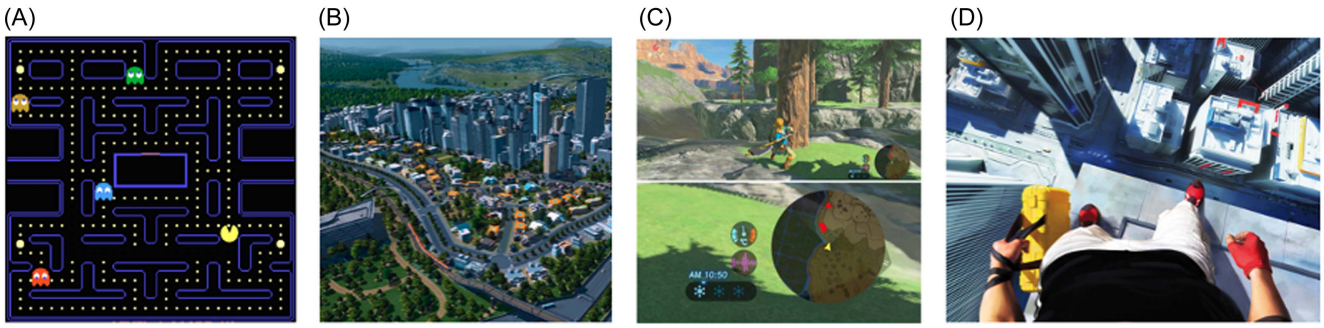
Our idea is that this sort of representational coordination and centering is, at least intuitively, related to how we represent ourselves as an "I" or as a "self" in centered thought and action. Our suggestion is that our avatar games provide an interesting experimental implementation of this type of perspectival representation and use of an icon to represent oneself as an agent, and that our Baba Is You games model a type of metacognitive "ascent" that is characteristic of reflective self-awareness and self-consciousness in a strange or new environment. While we are simply suggesting rather than arguing for this claim, it is strongly consistent with the spirit of the philosophical literature on the perspectivality of belief and self-involving attitudes. Perspectival representation, self-location, and centering more generally are often taken to be elements of the metaphysics and epistemology of personal identity, the self, self-consciousness, and self-involving perceptual content (see Bermúdez, 2017, 2018; Ismael, 2007; Paul, 2014; Paul & Quiggin, 2018; Perry, 1979; Recanati, 2007; Stalnaker, 2008; among many others). These discussions use phrases like "self-locating beliefs" and "de se" to describe a distinctive type of content that is intended to constitute a metaphysical or epistemological ground for the existence of a self, that is, a "thinking I." While we have not built the bridge between this philosophical content and our computational framework, as this would require an article of its own, we do think that our model gives a candidate structure for the psychological self at the level of the thinking agent, which is why we take our project, ultimately, as an attempt to "reverse engineer a centered self," or, at least, to provide the computational ground for such a self. Our hope is that our work will provide a scaffolding for further interdisciplinary research and discussion between psychologists, computational cognitive scientists, and philosophers on the nature of the self and its relation to perspectival thought and action.

We recognize that our approach, focused as it is on centering and problem-solving and drawing explicitly on the philosophical tradition of centering, departs substantially from many extant interpretations of "the self" in related areas (such as those of Cassam, 1994; Chisholm, 1969; Fekete & Edelman, 2011; Gallagher, 2000; Garfield, 2022; Metzinger, 2004; Starmans & Bloom, 2012; Strawson, 2013, 2017; Taylor, 2008; Thompson, 2014; Zahavi, 2008). While we recognize the importance of these literatures, we reinterpret and redefine the notion for our own purposes here, linking our computational framework to the different, and distinct, literature in analytic philosophy of language and mind grounded on indexicality, agency, and self-location, as described in Appendix A.[13]

[12] We note that there is an empirical literature on the coordination of first- and third-person representations, and such coordination is seen in nonhuman animals as well as humans. While we do not have the space in this article to expand on the connection, interested readers may wish to refer to related discussions on the construction of body ownership in nonhuman animals (e.g., Fang et al., 2019) as well as the mirror self-recognition task (Chang et al., 2015, 2017).

[13] Our notion of "having a self" relates to the engineering aim of those who want to develop things like a "self-driving" car or a machine that "thinks for itself." The possessive connotation of "have" is slightly misleading here, as "having a self" in our sense is more about realizing a process of generation and inference conducted from its intellectual center than "having" something. By analogy, consider an agent that can see or that "has vision" through a process of inverse graphics (see, e.g., Baumgart, 1974; Yuille & Kersten, 2006; and more recently Kulkarni et al., 2015). The agent's vision is not to be located in any one specific successful parsing of a visual scene, but in its general ability to do so, by inverting a generative process that could in principle produce myriad scenes.

Figure 17
Examples of Different Types of Avatars in Games



Note. Example frames from different kinds of games, showing different (visual) viewpoints, avatars, and representations that need to be linked to and coordinated. (A) Frame from Pac-Man, a classic two-dimensional game in which one of the agents (Pac-Man, in yellow) is the avatar, with relevant controls and goals, but with no first-person perspective built into the game. (B) Frame from City Skylines, a town-building game with a bird's-eye three-dimensional perspective but no specific avatar. (C) Frame from Zelda: Breath of the Wild, showing the player's avatar from an over-the-shoulder perspective (top), as well as relevant information about their location and heading in the form of a mini-map (bottom). (D) Frame from Mirror's Edge, the avatar visual viewpoint is presented from its first-person, within-the-boots perspective. See the online article for the color version of this figure.

Philosophical Implications: From Indexicals to Meta-ePOMDPs

As we state above, one of the contributions of this article is to bring the techniques and insights of computational cognitive science to bear on philosophical debates about centered worlds reasoning and perspectival belief and action. That said, as we noted in our introductory sections, our objective is simply to provide a starting point for the joint philosophical and computational exploration of centering and its relation to agent-centered problem-solving. We have no wish to endorse a deflationary interpretation of indexical questions such as "Who am I?" as they are used by human agents or to claim that our approach to self-location and its relation to problem-solving dissolves the interest in, or need for, the exploration of such questions. With this in mind, here we describe some potential implications of our work for discussions about indexicals and self-locating beliefs.

Philosophers since Perry (1979) have argued that certain first-person thoughts—"I am here now," "I'm over there," "I am the one who did it"—carry irreducibly indexical content: you cannot paraphrase "I" away without loss of meaning and loss of explanation. Our computational framework can be seen as exploring an experimental counterpart to this idea, that there is a psychologically important type of agent-centered problem-solving that involves self-location. In these cases, people, when they are inquiring (or solving problems) need to determine which possible world is actual, or which possible world they are in, and they need to represent themselves as occupying a certain position in that world, that is, they need to self-locate.

Our work shows how a computational model (the meta-ePOMDP) can implement precisely this kind of self-locating inference. Where the philosopher asks "which world am I in?" and "who, where, and when am I in this world?" our agent literally entertains multiple centered ePOMDPs—each a candidate *de se* representation—and uses Bayesian evidence and expected utilities to single out the correct one. In doing so, it provides a formal analogue to the self-locating beliefs that Perry (1979) and Lewis (1979) identified. Unlike traditional POMDPs, which handle uncertainty about the world, our extended POMDPs additionally separate *agent* and *world* states, modeling beliefs like "I am *that* red block" or "I am that agent solving that problem" (as defined by a particular ruleset). When our model updates its beliefs about avatar-identities or rulesets, it is performing a formal analogue of the very kind of belief revision that Perry and others argue is necessary for action.

We note that while, within philosophical circles, the notion of agent-centered problem-solving is uncontroversial, a debate rages over whether *de se* content or "essentially indexical" self-locating beliefs are necessary to explain these cases of human behavior. In this debate, phrases like "self-locating beliefs" and "de se" are used to describe a distinctive type of content that could constitute a metaphysical ground for the existence of a self, that is, a "thinking I," and advocates of the *de se* argue that such content is necessary for capturing certain types of action-ascriptions. We remain neutral on the metaphysics but point out that *computationally*, self-location by centering on "which avatar *I* control" is not the same as identifying the nonindexical "which block is closest to the goal"—and this difference emerges naturally in our Bayesian hierarchy. The fact that such a model better fits patterns of human behavior in our studies suggests the mind may indeed deploy a specialized mechanism for *de se* representation and inference (even if there is ultimately no such thing as *de se* content, but instead merely a *de se* mode of presentation, as developed in Paul, 2017b).

Ultimately, by embedding classic puzzles about "who am I?" and "where am I?" into a concrete, testable framework, we offer philosophers a novel tool for refining and evaluating competing theories of indexical thought, for building on centering in theories of the self and self-involving action, and for developing modal epistemology more broadly. Rather than relying solely on armchair examples, one can now specify a particular centering hypothesis in Bayesian terms, implement it in an interactive game environment, and observe whether human performance matches those formal predictions. This approach is intended as an initial approach toward transforming abstract debates about self-locating belief into empirically tractable hypotheses without losing sight of the acknowledged gap between our models and the rich and puzzling meaning of

indexicals such as "I" in natural language. Again, we have no wish to deflate the interpretation of questions such as "Who am I?" or "Who are you?" or to suggest that our version of the problem addresses all of the philosophical dimensions of treatments of the first person, the *de se*, and related debates. We hope, rather, that we are opening the door to systematic comparison among philosophical accounts (e.g., file-theoretic vs. character/content semantics) and providing a rigorous empirical framing for philosophical projects that wish to incorporate such computational insights. As these paradigms grow richer—incorporating real-time sensorimotor cues alongside rule-space inference—they promise to give philosophers additional perspectives on long-standing discussions about the nature of perspectival thought, belief, and action.

Further Directions

The current work also suggests future empirical and computational modeling investigations of triangulation, more "radical" forms of centering and recentering, more biologically inspired computational models of centering, and exploration of more realistic settings and other forms of centering.

Triangulation

While in the current work, we have focused on implementing a simpler version of centering that involves just one (third-person) perspective, we hope future work will attempt to implement more complicated versions of centering that entail triangulating multiple perspectives. As a starting point, empirical and modeling paradigms could present participants with first-person perspectives that become misaligned with the world, then test when and how participants are able to leverage third-person representations of the self or its position in order to relocate (or find) the self at a new location.

Alternative Computational Foundations of Centering

One might imagine alternative computational accounts of centering on a problem representation that do not implement all aspects of the hierarchical meta-ePOMDP framework as we have formalized and implemented it. In particular, it would be interesting to consider departures from the explicitly Bayesian computations we adopt. For example, instead of selecting a problem representation through explicit, online probabilistic inference and expected value decision-making, aspects of this decision process might be cached over time, through a sort of "metacognitive reinforcement learning" (as described by Callaway et al., 2017; Lieder & Griffiths, 2017). In the course of everyday experience, people likely come to rely on a range of strategies for efficiently selecting and updating problem representations that approximate the computational level theory we have presented here. Though such approximate meta-ePOMDP solvers might still include factored, hierarchical representations of problems to be solved and ways of solving those problems, one can also imagine alternative models—purely model-free or model-based reinforcement learning algorithms, symbolic planners built atop graph search, or hybrid neurosymbolic architectures—that do not explicitly instantiate a space of self-world models or perform metacognitive recentering. However, as De Freitas et al. (2023) demonstrated, even sophisticated deep-RL agents trained end to end on the avatar games fail to capture the speed and flexibility of human centering, suggesting that mere experience-driven learning is insufficient. In contrast, people arrive with rich inductive biases—what one might call "start-up software"—that bootstrap rapid self-location and problem identification and that (we posit) allow them to solve both the avatar games and the Baba Is You games much more quickly and robustly. This raises an open question: What minimal set of built-in representations or neural mechanisms would allow a biologically plausible learner to acquire humanlike centering and recentering capabilities?

Realism and Other Forms of Centering

We have illustrated our framework as applied to the avatar problem, where the agent must center on a particular avatar realization in a particular game world. In a simple grid world, the first-person data and representations one must condition on are far less rich (e.g., simple colors of grid-world squares), as is the relevant self-knowledge one needs to solve the game (e.g., you are one of the red squares, and have four available actions). These simplified representations constrain the space of possible ePOMDPs the meta-agent might construct, making the task of synthesizing and evaluating ePOMDPs that appropriately integrate this information computationally tractable. However, the avatar problem is still just a simplified version of what we take the real world computational task to be. When people center in the real world, the relevant representations and self-knowledge involved are far more complex, bringing many additional ways of integrating these representations and self-knowledge.

Other Problems of the Self

Finally, we hope that our computational model might lay the groundwork for connecting and illuminating several other aspects of self-representation. In particular, literature spanning psychology and neuroscience has characterized many common challenges, illusions, and mistakes involved in self-representation which may be usefully understood as instances or aspects of agent-centered problem-solving. These include first-person self-localization (Bertoni et al., 2023; Botvinick & Cohen, 1998; Schettler et al., 2020), sense of agency (Haggard et al., 2002; Legaspi & Toyoizumi, 2019; Wegner, 2004), spatial navigation (Collett & Graham, 2004; Ekstrom & Isham, 2017; Epstein et al., 2017; Freas & Cheng, 2022), and even disorders of the self (Bertoni et al., 2023; Franck et al., 2000; Frith, 2005; Frith et al., 2000; Leptourgos & Corlett, 2020; Schwabe & Blanke, 2008). We leave the development of these connections for future work.

Conclusion

We take our approach to constitute a first step toward understanding the computational way that people represent themselves in order to act, plan, and self-correct when they engage in agent-centered problem-solving. We recognize that we have formalized and tested quantitatively only some aspects of our theoretical proposal, and we do not claim that even fully realized mePOMDP agents would completely capture the complex tasks of constructing, coordinating, and updating the representations needed to implement agent-centered problem-solving. However, we think the initial work presented here opens up a range of possibilities for modeling hallmark features of this type of perspectival problem-solving, and moreover, provides a robust computational framework that psychologists, cognitive

scientists, and philosophers can draw on in order to continue to explore interdisciplinary questions about the nature of having a self.

References

Ahmed, Z., Tenenbaum, J. B., Bates, C. J., & Gershman, S. J. (2025). *Synthesizing world models for bilevel planning*. arXiv. https://doi.org/10.48550/arXiv.2503.20124

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences*, *117*(47), 29302–29310. https://doi.org/10.1073/pnas.1912341117

Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, *9*(2), 93–106. https://doi.org/10.1007/s10339-007-0188-5

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), Article 0064. https://doi.org/10.1038/s41562-017-0064

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18327–18332. https://doi.org/10.1073/pnas.1306572110

Baumgart, B. G. (1974). *Geometric modeling for computer vision* [Technical report]. Defense Technical Information Center. https://apps.dtic.mil/sti/citations/ADA002261

Berlucchi, G., & Aglioti, S. (1997). The body in the brain: Neural bases of corporeal awareness. *Trends in Neurosciences*, *20*(12), 560–564. https://doi.org/10.1016/S0166-2236(97)01136-3

Bermúdez, J. (2017). *Understanding "I": Language and thought*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198796213.001.0001

Bermúdez, J. (2018). *The bodily self: Selected essays*. MIT Press. https://mitpress.mit.edu/9780262551083/the-bodily-self/

Bertoni, T., Mastria, G., Akulenko, N., Perrin, H., Zbinden, B., Bassolino, M., & Serino, A. (2023). The self and the Bayesian brain: Testing probabilistic models of body ownership through a self-localization task. *Cortex*, *167*, 247–272. https://doi.org/10.1016/j.cortex.2023.06.019

Bi, W., Lin, Q., Peng, K., Shah, A., & Yildirim, I. (2024). Visual processing of soft objects automatically activates physics-based representations in the human brain. *Journal of Vision*, *24*(10), Article 835. https://doi.org/10.1167/jov.24.10.835

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge. https://doi.org/10.4324/9780203508527 (Original work published 1990)

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391*(6669), Article 756. https://doi.org/10.1038/35784

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Bratman, M. E. (2018). *Planning, time, and self-governance: Essays in practical rationality*. Oxford University Press. https://doi.org/10.1093/oso/9780190867850.001.0001

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, *4*(3), 349–355. https://doi.org/10.1111/j.1467-8640.1988.tb00284.x

Braun, D. (2017). Indexicals. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 ed.). Stanford University. https://plato.stanford.edu/archives/sum2017/entries/indexicals/

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. https://doi.org/10.1016/j.tics.2006.10.005

Butterfield, J., Jenkins, O. C., Sobel, D. M., & Schwertfeger, J. (2009). Modeling aspects of theory of mind with Markov random fields. *International Journal of Social Robotics*, *1*(1), 41–51. https://doi.org/10.1007/s12369-008-0003-1

Callaway, F., Gul, S., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2017). *Learning to select computations*. arXiv. https://doi.org/10.48550/arXiv.1711.06892

Campbell, J. (1994). *Past, space and self*. Clarendon Press. https://doi.org/10.7551/mitpress/5262.001.0001

Cappelen, H., & Dever, J. (2013). *The inessential indexical: On the philosophical insignificance of perspective and the first person*. Oxford University Press. https://doi.org/10.1093/analys/anv077

Cassam, Q. (Ed.). (1994). *Self-knowledge*. Oxford University Press.

Chandra, K., Chen, T., Tenenbaum, J. B., & Ragan-Kelley, J. (2025). A domain-specific probabilistic programming language for reasoning about reasoning (or: a memo on memo). *Proceedings of the ACM on Programming Languages*, *9*, 784–814. https://doi.org/10.1145/3763078

Chang, L., Fang, Q., Zhang, S., Poo, M. M., & Gong, N. (2015). Mirror-induced self-directed behaviors in rhesus monkeys after visual-somatosensory training. *Current Biology*, *25*(2), 212–217. https://doi.org/10.1016/j.cub.2014.11.016

Chang, L., Zhang, S., Poo, M. M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(12), 3258–3263. https://doi.org/10.1073/pnas.1620764114

Charity, M., Dave, I., Khalifa, A., & Togelius, J. (2022). Baba is y'all 2.0: Design and investigation of a collaborative mixed-initiative system. *IEEE Transactions on Games*, *16*(1), 75–89. https://doi.org/10.1109/TG.2022.3223527

Charity, M., Khalifa, A., & Togelius, J. (2020). Baba is y'all: Collaborative mixed-initiative level design. *2020 IEEE Conference on Games (CoG)* (pp. 542–549). Institute of Electrical and Electronics Engineers.

Charity, M., & Togelius, J. (2022). Keke AI competition: Solving puzzle levels in a dynamically changing mechanic space. *2022 IEEE Conference on Games (CoG)* (pp. 570–575). Institute of Electrical and Electronics Engineers.

Chen, T., Cheyette, S., Allen, K., Tenenbaum, J., & Smith, K. (2026). *"Just in time" world modeling supports human planning and reasoning*. arXiv. https://arxiv.org/abs/2601.14514

Chisholm, R. M. (1969). On the observability of the self. *Philosophy and Phenomenological Research*, *30*(1), 7–21. https://doi.org/10.2307/2105917

Chrysikou, E. G. (2006). When shoes become hammers: Goal-derived categorization training enhances problem-solving performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 935–942. https://doi.org/10.1037/0278-7393.32.4.935

Cloos, N., Jens, M., Naim, M., Kuo, Y. L., Cases, I., Barbu, A., & Cueva, C. J. (2024). *Baba is AI: Break the rules to beat the benchmark* (arXiv preprint arXiv:2407.13729). https://doi.org/10.48550/arxiv.2407.13729

Collett, T. S., & Graham, P. (2004). Animal navigation: Path integration, visual landmarks and cognitive maps. *Current Biology*, *14*(12), R475–R477. https://doi.org/10.1016/j.cub.2004.06.013

Colombo, M., Lai, J., & Crupi, V. (2018). Sleeping beauty goes to the lab: The psychology of self-locating evidence. *Review of Philosophy and Psychology*, *10*, 173–185. https://doi.org/10.1007/s13164-018-0381-8

Cusumano-Towner, M. F., Saad, F. A., Lew, A. K., & Mansinghka, V. K. (2019, June 22–26). Gen: A general-purpose probabilistic programming system with programmable inference. *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation* (pp. 221–236). Association for Computing Machinery.

De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., Paul, L. A., Tenenbaum, J., & Ullman, T. (2023). Self-orienting in human and machine learning. *Nature Human Behavior*, *7*, 2126–2139. https://doi.org/10.1038/s41562-023-01696-5

Dennett, D. (1987). *The intentional stance*. Cambridge University Press.
Dennett, D. (1991). Consciousness explained, "the origin of selves". *Cogito*, *3*(3), 163–173.
Dennett, D. (1992). The self as a center of narrative gravity. In F. S. Kessel, P. M. Cole, D. L. Johnson, & M. D. Hakel (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103–115). Psychology Press.
Devitt, M. (2013). The myth of the problematic de se. In A. Capone & N. Feit (Eds.), *Attitudes de se: Linguistics, epistemology, metaphysics* (pp. 133–162). CSLI Publications.
Duncker, K. (1945). *On problem-solving* (Psychological Monographs, No. 270.). American Psychological Association.
Ekstrom, A., & Isham, E. (2017). Human spatial navigation: Representations across dimensions and scales. *Current Opinion in Behavioral Sciences*, *17*, 84–89. https://doi.org/10.1016/j.cobeha.2017.06.005
Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*, 1504–1513. https://doi.org/10.1038/nn.4656
Evans, O., Stuhlmüller, A., Salvatier, J., & Filan, D. (2018, November 18). *Modeling agents with probabilistic programs*. https://agentmodels.org
Fang, W., Li, J., Qi, G., Li, S., Sigman, M., & Wang, L. (2019). Statistical inference of body representation in the macaque brain. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(40), 20151–20157. https://doi.org/10.1073/pnas.1902334116
Fekete, T., & Edelman, S. (2011). Towards a computational theory of experience. *Consciousness and Cognition*, *20*(3), 807–827. https://doi.org/10.1016/j.concog.2011.02.010
Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417
Franck, N., Rouby, P., Daprati, E., Daléry, J., Marie-Cardine, M., & Georgieff, N. (2000). Confusion between silent and overt reading in schizophrenia. *Schizophrenia Research*, *41*(2), 357–364. https://doi.org/10.1016/S0920-9964(99)00067-5
Freas, C., & Cheng, K. (2022). The basis of navigation across species. *Annual Review of Psychology*, *73*, 217–241. https://doi.org/10.1146/annurev-psych-020821-111311
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732
Frith, C. D. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition*, *14*, 752–770. https://doi.org/10.1016/j.concog.2005.04.002
Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *355*(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734
Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5
Garfield, J. L. (2022). *Losing ourselves: Learning to live without a self*. Princeton University Press. https://doi.org/10.1515/9780691220291
German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science*, *16*(1), 1–5. https://doi.org/10.1111/j.0956-7976.2005.00771.x
Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076
Goodman, N. D., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2008). Church: A language for generative models. *UAI'08: Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence* (pp. 220–229). Association for Computing Machinery.
Goodman, N. D., Tenenbaum, J. B., & the ProbMods Contributors. (2016). *Probabilistic models of cognition* (2nd ed.). http://probmods.org/
Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004
Griffiths, T. L., Chater, N., & Tenenbaum, J. B. (Eds.). (2024). *Bayesian models of cognition: Reverse engineering the mind*. MIT Press.
Haber, N., Mrowca, D., Fei-Fei, L., & Yamins, D. L. (2018). *Learning to play with intrinsically-motivated self-aware agents*. arXiv. https://doi.org/10.48550/arXiv.1802.07442
Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385. https://doi.org/10.1038/nn827
Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, *34*(2–3), 102–254. https://doi.org/10.1093/brain/34.2-3.102
Hempuli. (2019). *Baba is you* [Video game]. Hempuli Oy. https://hempuli.com/baba/
Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature*, *606*(7912), 129–136. https://doi.org/10.1038/s41586-022-04743-9
Ismael, J. T. (2007). *The situated self*. Oxford University Press.
Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011
Jern, A., & Kemp, C. (2015). A decision network account of reasoning about other people's choices. *Cognition*, *142*, 12–38. https://doi.org/10.1016/j.cognition.2015.05.006
Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X
Kaplan, D. (1979). On the logic of demonstratives. *Journal of Philosophical Logic*, *8*(1), 81–98. https://doi.org/10.1007/BF00258420
Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.
Kim, K., Sano, M., De Freitas, J., Haber, N., & Yamins, D. (2020). Active world model learning with progress curiosity. *International Conference on Machine Learning* (pp. 5306–5315). Proceedings of Machine Learning Research.
Kleiman-Weiner, M., Saxe, R., & Tenenbaum, J. B. (2017). Learning a commonsense moral theory. *Cognition*, *167*, 107–123. https://doi.org/10.1016/j.cognition.2017.03.005
Kulkarni, T. D., Whitney, W., Kohli, P., & Tenenbaum, J. B. (2015). *Deep convolutional inverse graphics network*. arXiv. https://doi.org/10.48550/arXiv.1503.03167
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, Article e253. https://doi.org/10.1017/S0140525X16001837
Legaspi, R., & Toyoizumi, T. (2019). A Bayesian psychophysics model of sense of agency. *Nature Communications*, *10*(1), Article 4250. https://doi.org/10.1038/s41467-019-12170-0
Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, *317*(5841), 1096–1099. https://doi.org/10.1126/science.1143439
Leptourgos, P., & Corlett, P. R. (2020). Embodied predictions, agency, and psychosis. *Frontiers in Big Data*, *3*, Article 27. https://doi.org/10.3389/fdata.2020.00027
Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*(4), 513–543. https://doi.org/10.2307/2184843
Li, H. H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*(1), Article 2004. https://doi.org/10.1038/s41467-020-15581-6
Liao, S. Y. (2012). What are centered worlds? *The Philosophical Quarterly*, *62*(247), 294–316. https://doi.org/10.1111/j.1467-9213.2011.00042.x

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762–794. https://doi.org/10.1037/rev0000075

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

List, C. (2023). The many-worlds theory of consciousness. *Noûs*, *57*(2), 316–340. https://doi.org/10.1111/nous.12408

Magidor, O. (2015). The myth of the de se. *Philosophical Perspectives*, *29*(1), 249–283. https://doi.org/10.1111/phpe.12065

Marr, D., & Poggio, T. (1976). *From understanding computation to understanding neural circuitry* [AI Memos]. http://hdl.handle.net/1721.1/5782

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing*. MIT Press.

Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. MIT Press.

Millikan, R. G. (1990). The myth of the essential indexical. *Noûs*, *24*(5), 723–734. https://doi.org/10.2307/2215811

Milne, A. A., & Shepard, E. H. (1926). *Winnie-the-pooh*. Methuen & Co.

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Nagel, T. (1989). *The view from nowhere*. Oxford University Press.

Ninan, D. (2013). Self-location and other-location. *Philosophy and Phenomenological Research*, *87*(2), 301–331. https://doi.org/10.1111/phpr.12051

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*(8), 349–357. https://doi.org/10.1016/S1364-6613(00)01699-5

Ongchoco, J. D. K., Davis, I. M., Jara-Ettinger, J., & Paul, L. A. (2024). When new experience leads to new knowledge: A computational framework for formalizing epistemically transformative experiences. *Open Mind*, *8*, 1291–1311. https://doi.org/10.1162/opmi_a_00168

Paglieri, D., Cupiał, B., Coward, S., Piterbarg, U., Wolczyk, M., Khan, A., Pignatelli, E., Kuciński, Ł., Pinto, L., Fergus, R., Foerster, J. N., Parker-Holder, J., & Rocktäschel, T. (2024). *BALROG: Benchmarking agentic LLM and VLM reasoning on games*. arXiv. https://doi.org/10.48550/arXiv.2411.13543

Paul, L. A. (2014). *Transformative experience*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198717959.001.0001

Paul, L. A. (2016). Who am I? The immersed first personal view. In A. M. Battro & S. Dehaene (Eds.), *Power and limits of artificial intelligence* (pp. 106–113). Pontificiae Academiae Scientiarum Scripta Varia.

Paul, L. A. (2017a). First personal modes of presentation and the structure of empathy. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, *60*(3), 189–207. https://doi.org/10.1080/0020174X.2017.1261991

Paul, L. A. (2017b). De se preferences and empathy for future selves. *Philosophical Perspectives*, *31*, 7–39. https://doi.org/10.1111/phpe.12090

Paul, L. A., & Quiggin, J. (2018). Real world problems. *Episteme*, *15*(3), 363–382. https://doi.org/10.1017/epi.2018.28

Paul, L. A., Ullman, T. E., De Freitas, J., & Tenenbaum, J. (2023). *Reverse-engineering the self*. Open Science Framework. https://doi.org/10.31234/osf.io/vzwrn

Peacocke, C. (1999). *Being known*. Oxford University Press. https://doi.org/10.1093/0198238606.001.0001

Peacocke, C. (2008). *Truly understood*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199239443.001.0001

Perry, J. (1977). Frege on demonstratives. *Philosophical Review*, *86*(4), 474–497. https://doi.org/10.2307/2184564

Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*(1), 3–21. https://doi.org/10.2307/2214792

Perry, J. (1996). The self. In E. Craig (Ed.), *Routledge encyclopedia of philosophy*. Routledge.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLOS ONE*, *3*(12), Article e3832. https://doi.org/10.1371/journal.pone.0003832

Pollock, J. L. (2006). *Thinking about acting: Logical foundations for rational decision making*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195304817.001.0001

Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S. M., & Botvinick, M. (2018). *Machine theory of mind*. arXiv. https://doi.org/10.48550/arXiv.1802.07740

Recanati, F. (2007). *Perspectival thought*. https://doi.org/10.1093/acprof:oso/9780199230532.001.0001

Recanati, F. (2012a). Immunity to error through misidentification: What it is and where it comes from. In *Immunity to error through misidentification: New essays* (pp. 180–201).

Recanati, F. (2012b). *Mental files*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659982.001.0001

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson.

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach*. Pearson.

Savage, L. J. (1954). *The foundations of statistics*. Wiley.

Schettler, A., Holstead, I., Turri, J., & Barnett-Cowan, M. (2020). Visual self-motion feedback affects the sense of self in virtual reality. *Multisensory Research*. Advance online publication. https://doi.org/10.1163/22134808-bja10043

Schwabe, L., & Blanke, O. (2008). The vestibular component in out-of-body experiences: A computational approach. *Frontiers in Human Neuroscience*, *2*, Article 17. https://doi.org/10.3389/neuro.09.017.2008

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118. https://doi.org/10.2307/1884852

Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199545995.001.0001

Starmans, C., & Bloom, P. (2012). Windows to the soul: Children and adults see the eyes as the location of the self, *Cognition*, *123*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.02.002

Strawson, G. (2013). Self-intimation. *Phenomenology and the Cognitive Sciences*, *14*(1), 1–31. https://doi.org/10.1007/s11097-013-9339-6

Strawson, G. (2017). *The subject of experience*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198777885.001.0001

Taylor, C. (2008). *Sources of the self*. Harvard University Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Thompson, E. (2014). *Waking, dreaming, being: Self and consciousness in neuroscience, meditation, and philosophy*. Columbia University Press.

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive Psychology*, *104*, 57–82. https://doi.org/10.1016/j.cogpsych.2017.05.006

Von Neumann, J., & Morgenstern, O. (1994). *Theory of games and economic behavior*. Princeton University Press.

Ward, T. B., Smith, S. M., & Vaid, J. E. (1997). *Creative thought: An investigation of conceptual structures and processes*. American Psychological Association. https://doi.org/10.1037/10227-000

Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, *27*(5), 649–659. https://doi.org/10.1017/S0140525X04000159

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, *85*(1), 5–19. https://doi.org/10.1037/0022-3514.85.1.5

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480–492. https://doi.org/10.1037/0003-066X.54.7.480

Wong, L., Collins, K., Ying, L., Zhang, C., Weller, A., Gersternberg, T., O'Donnell, T., Lew, A., Andreas, J., Tenenbaum, J., & Brooke-Wilson, T. (2025). *Modeling open-world cognition as on-demand synthesis of probabilistic models* (arXiv preprint arXiv:2507.12547). https://doi.org/10.48550/arXiv.2507.12547

Wong, L., Grand, G., Lew, A. K., Goodman, N. D., Mansinghka, V. K., Andreas, J., & Tenenbaum, J. B. (2023). *From word models to world models: Translating from natural language to the probabilistic language of thought*. arXiv. https://doi.org/10.48550/arXiv.2306.12672

Yildirim, I., & Paul, L. A. (2024). From task structures to world models: What do LLMs know? *Trends in Cognitive Sciences*, *28*(5), 404–415. https://doi.org/10.1016/j.tics.2024.02.008

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zahavi, D. (2008). *Subjectivity and selfhood: Investigating the first-person perspective*. MIT Press. https://doi.org/10.7551/mitpress/6541.001.0001

Appendix A

Indexicality, Agency, and Self-Locating Beliefs

Following a tradition in the philosophy of language and mind, the content of a psychological attitude can be characterized by a set of possibilities. This insight serves as a foundation for philosophical models of indexical attitudes, including self-involving beliefs, desires, and other attitudes using "centered possible worlds."

In modal metaphysics, a possible world represents a way a world could be. A centered possible world represents a way an environment could be from a particular "center" or "perspective" in that world, that is, the way the world might be for a given individual at a given spatiotemporal location in that world.

Indexical expressions are expressions that include terms such as "I" or "now" whose referents are determined by the context in which they are tokened. Indexical properties are properties of place, time, individual, and world. They are captured by indexical truths expressed by sentences such as "I am in the Stanford library" or "The meeting starts now." They vary as an agent's context varies. We can see the need for a semantics of variability when we compare "I am in the Stanford library" to "Rudolf is in the Stanford library": The truth of the first varies from speaker to speaker, the truth of the second does not. Recanati (2012b) argues that the use of mental files can capture the variable nature of indexical thought, and for a classic treatment of the semantics of indexical expressions, see Kaplan (1979, 1989). For a thorough introduction, see Braun (2017).

Many discussions of the importance of centering and its role in agency and perspectival attitudes stem from examples suggesting there is self-locating content. These examples were first put forward by Perry (1979) in his groundbreaking article, "The Problem of the Essential Indexical." In this article, Perry introduced the terms "essential indexicality" and "locating beliefs" and took a "(self) locating belief" to be an indexical belief about where I am, when I am, and who I am. On Perry's view, such a belief is *essentially* indexical, that is, if the indexical "I" is replaced, this perspectival, indexical belief is replaced with a different nonperspectival belief. Relatedly, Lewis (1979) defends a view about "de se attitudes." In both approaches there is a commitment to irreducibly perspectival *de se* content, which is the content that is replaced when the indexical is replaced. In their skeptical treatise on the *de se*, Cappelen and Dever (2013) summarized the contemporary version of the view thus:

> Perspectival representations are representationally essential: perspectivality makes available ways of representing the world that are not available with non-perspectival devices. And perspectival representations are metarepresentationally essential: many of the philosophical roles that we want representations to play can be played only by perspectival representations. The goal of inquiry is to determine which possible world is actual, which possible world we are in. Even stronger, for beliefs and desires to motivate us to action, they must not merely represent the world, but represent us as occupying a certain position in the world. Perception essentially presents the world as received from a perspective. (Cappelen & Dever, 2013, p. 1)

Cappelen and Dever go on to reject the essentialist versions of these claims, and more generally, to reject *de se* content, and in the broader philosophical literature, whether or not *de se* content is necessary for action is hotly debated. Some endorse the thesis that we employ a distinctive "de se" semantics for the indexical "I," or the "essential indexical," to capture distinctive metaphysical and epistemic features of perspectival, first-person thought and action. They argue that this semantics is needed to adequately represent perspectival features of reality, in particular, features concerning ourselves as thinkers and our relationship to other parts of the world. Others demur.

We take no stand on this important philosophical question. While we embrace the importance of perspectival thought and action for psychology and philosophy, especially in the context of modeling agent-centered problem-solving, we make no commitment to any essentialist theses about *de se* content (though we are happy to grant that consciousness may endow a person with a special kind of subjective presentation of truths about themselves, a kind of *de se* phenomenal mode, see Paul, 2017b for discussion).

That is, the crucial element of Perry's idea that we endorse is the point that, when a person centers or recenters themselves, they can do so in part by adopting new representations of themselves in their world (they gain a "self-locating" belief), and this brings an increase in self-understanding. Such understanding can be cognitively significant when it involves a discovery about oneself as an agent in the world (*I am the person who is making the tracks*) and even more so when it brings a new representation (or presentation) of the nature of the world itself (*there is no Woozle*).

Readers interested in the debate about indexicality and the *de se* should consult Bermúdez (2017, 2018), Devitt (2013), Campbell (1994), Cappelen and Dever (2013), Ismael (2007), Lewis (1979), Liao (2012), Magidor (2015), Millikan (1990), Ninan (2013), Paul (2016, 2017b; Paul & Quiggin, 2018), Peacocke (2008), Perry (1977, 1979), Recanati (2012a), Stalnaker (2008), and the formalization of the account of the self and of self-knowledge given in Perry (1996). Readers interested in extant empirical work on self-locating beliefs should consult Colombo et al. (2018).

(*Appendices continue*)

Appendix B

Creative Problem-Solving Without Recentering

Our framework also invites comparison with literatures on creative problem-solving, particularly those that distinguish between "thinking outside the box" and "thinking inside the box" (Ward et al., 1997). These paradigms explore how people shift their conceptual frames or reconfigure their understanding of the problem space—often by loosening or redefining assumed constraints. A central construct in this tradition is *functional fixedness*, the cognitive bias to perceive objects only in terms of their typical uses, which can hinder problem-solving by overly constraining the space of affordances (Duncker, 1945). Overcoming functional fixedness—such as recognizing that a box might serve as a platform rather than merely as a container—requires relaxing one's assumptions about the roles of objects, and thus amounts to a transformation of the world model (Chrysikou, 2006; German & Barrett, 2005).

This kind of reframing also appears in abstract reasoning tasks. For instance, in the well-known *bat and ball* problem from the cognitive reflection test (Frederick, 2005), people often arrive at the intuitive but incorrect answer by applying a default heuristic. Solving the problem correctly requires inhibiting that default and recentering one's interpretation of the linguistic structure of the question—treating it not as a straightforward arithmetic problem but as one requiring recursive relational reasoning. In this sense, even abstract puzzles can require a form of cognitive recentering, in which the agent shifts perspective not spatially, but representationally, to locate themselves differently within the problem space. Another relevant contribution comes from Boden (2004), who characterized a distinctive type of transformational creativity in terms of changing generative rules or constraints such that previously "impossible" ideas become generable.

However, what these creative and logical paradigms typically lack is the notion of *self-location*—of centering the problem in relation to a situated agent and solving the problem from this perspectival approach. Our model of agent-centered problem-solving emphasizes not just representational flexibility, but perspectival and self-locating reasoning. It is not merely that the problem shifts, but that the agent must reposition themselves as a particular entity, at a particular moment, within a dynamic world, choosing an entirely new perspective. This distinction highlights how our model captures a uniquely self-referential and situated mode of cognition—one that complements, but goes beyond, the scope of classical creativity and reasoning frameworks. It opens the door to understanding cognitive constraints not only as fixed assumptions about the world but also as fixed assumptions about the self as it is situated in the world (Ismael, 2007).

Appendix C

The Avatar Games Model Implementation Details

The Meta-ePOMDP Agent: Estimating Informative Actions for Centering

Here, we provide more detail on how the meta-ePOMDP agent selects informative actions while centering. Specifically, at each timestep, it selects the action a that maximizes the one-step expected information gain (*ig*):

$$Eig|a, o_{1:t-1} = \sum_{o_t} JSP_t(E|o_{1:t}), P_{t-1}(E|o_{1:t-1}) \cdot Po_t|a, o_{1:t-1}, \quad \text{(C1)}$$

where JS(.,.) is the Jensen–Shannon between prior and posterior distributions and $P(o_t|a, o_{1:t-1})$ is the distribution over possible next observations. This can be computed by integrating over possible next hidden states s_t:

$$Po_t|a, o_{1:t-1} = \int_{s_t} Po_t|s_t \cdot Ps_t|a, o_{1:t-1}, \quad \text{(C2)}$$

where the distribution of possible next hidden states is integrated over possible transitions:

$$Ps_t|a, o_{1:t-1} = \int_{s_{t-1}} Ts_t|s_{t-1}, a \cdot Ps_{t-1}|o_{1:t-1} \quad \text{(C3)}$$

To make this process tractable, we approximate these integrals with Monte Carlo sampling. For each candidate action, the agent simulates 10 possible next world states and their corresponding posterior updates. It then computes the expected information gain as the average Jensen–Shannon distance between posterior and prior over these 10 simulations. This leads to efficient exploration. For example, in an environment in which each character has a wall directly to its right and no wall to its left, moving right will be expected to provide less information than moving left. When moving left, any character that moves left will either be the avatar or a nonavatar character that happened to move left. When moving right, any character that stays put will be the avatar, or a nonavatar character that *either* attempted to move right or to stay put. Moving left is more likely to lead to an observation that will be highly improbable for more possible avatar identities, allowing the meta-ePOMDP agent to infer the agent realization more quickly.

Proximity Heuristic Model: Game-Specific Details

In the *switching mappings* game, the mapping between agent actions and avatar movements is initially unknown. The proximity heuristic uses the action mapping which has most often been consistent with past movements of selected characters when selecting actions in this game type. Specifically, at each timestep, when attempting to move the closest character to the goal, the heuristic model selects an action and observes the movement of the selected character. It checks whether each possible action mapping is consistent with the resulting direction of movement. When selecting an action, it determines the desired direction of movement with the A* algorithm and picks the action which corresponds to that direction in the action mapping which has most often been consistent with selected character movements.

(*Appendices continue*)

Supplementary Results

Table C1
Independent t Tests Comparing Model and Study 1 Human Solving Times on the Avatar Games

Game type	mePOMDP	Resource-limited mePOMDP	Proximity heuristic
Logic	$t(20.0) = -3.99$ $p < .001$ $d = 1.26$	$t(25.3) = 1.20$ $p = .240$ $d = 0.38$	$t(38.0) = 4.97$ $p < .001$ $d = 1.57$
Contingency	$t(19.5) = -5.56$ $p < .001$ $d = 1.75$	$t(21.7) = -0.84$ $p = .410$ $d = 0.27$	$t(19.0) = 12.0$ $p < .001$ $d = 3.81$
Switching mappings	$t(19.1) = -4.80$ $p < .001$ $d = 1.52$	$t(20.3) = 0.44$ $p = .666$ $d = 0.14$	$t(19.2) = 6.54$ $p < .001$ $d = 2.07$
Switching embodiments	$t(17.3) = -4.91$ $p < .001$ $d = 1.68$	$t(19.7) = 1.23$ $p = .233$ $d = 0.42$	$t(18.3) = -3.58$ $p = .00209$ $d = 1.22$

Note. Independent t tests for each game type and model, between the solving times of humans and each model. Solving times averaged across participants (for the human data) and seeds (for the models). p values indicate the probability of obtaining a test-statistic at least as extreme as the one observed under the null hypothesis that the means of the distributions underlying the two samples are equal. mePOMDP = meta-extended partially observable Markov decision process.

Table C2
Independent t Tests Comparing Model and Study 2 Human Solving Times on the Avatar Games

Game type	mePOMDP	Resource-limited mePOMDP	Proximity heuristic
Logic	$t(20.3) = -3.11$ $p = .00540$ $d = 0.98$	$t(22.8) = 1.08$ $p = .291$ $d = 0.34$	$t(38.0) = 4.40$ $p < .001$ $d = 1.39$
Contingency	$t(19.9) = -4.63$ $p < .001$ $d = 1.47$	$t(20.4) = -1.20$ $p = .242$ $d = 0.38$	$t(21.1) = 22.6$ $p < .001$ $d = 7.16$
Switching mappings	$t(19.2) = -5.54$ $p < .001$ $d = 1.75$	$t(26.7) = 1.10$ $p = .283$ $d = 0.35$	$t(38.0) = 13.3$ $p < .001$ $d = 4.21$
Switching embodiments	$t(11.3) = -6.36$ $p < .001$ $d = 2.01$	$t(38.0) = 5.84$ $p < .001$ $d = 1.85$	$t(38.0) = -2.14$ $p = .0392$ $d = 0.68$
Switching embodiments (every 10)	$t(21.1) = -5.44$ $p < .001$ $d = 1.72$	$t(26.7) = 2.20$ $p = .0368$ $d = 0.70$	$t(38.0) = -0.724$ $p < .474$ $d = 0.23$
Contingency + goal uncertainty	$t(19.7) = -2.73$ $p = .0129$ $d = 0.86$	$t(20.0) = -1.43$ $p = .169$ $d = 0.45$	$t(38.0) = 17.0$ $p < .001$ $d = 5.38$
Switching mappings + goal uncertainty	$t(21.2) = -12.5$ $p < .001$ $d = 3.96$	$t(38.0) = 1.61$ $p = .116$ $d = 0.51$	$t(26.3) = 18.2$ $p < .001$ $d = 5.76$
Noisy contingency	$t(25.0) = -5.32$ $p < .001$ $d = 1.68$	$t(38.0) = 1.83$ $p = .0750$ $d = 0.58$	$t(24.9) = 18.2$ $p < .001$ $d = 5.77$
Contingency2	$t(21.5) = -0.643$ $p = .0527$ $d = 0.20$	$t(21.4) = -0.328$ $p = .746$ $d = 0.10$	$t(19.7) = 12.6$ $p < .001$ $d = 4.00$
Contingency6	$t(19.3) = -6.36$ $p < .001$ $d = 2.01$	$t(21.7) = -1.80$ $p = .0852$ $d = 0.57$	$t(22.6) = 16.9$ $p < .001$ $d = 5.35$
Contingency8	$t(19.0) = -5.38$ $p < .001$ $d = 1.70$	$t(21.7) = -1.56$ $p = .133$ $d = 0.49$	$t(38.0) = 15.9$ $p < .001$ $d = 5.02$

Note. Independent t tests for each game type and model, between the solving times of humans and each model. Solving times averaged across participants (for the human data) and seeds (for the models). p values indicate the probability of obtaining a test-statistic at least as extreme as the one observed under the null hypothesis that the means of the distributions underlying the two samples are equal. mePOMDP = meta-extended partially observable Markov decision process.

(*Appendices continue*)

Table C3

Best Fitting Parameters of Resource-Limited Meta-Extended Partially Observable Markov Decision Process Agent by Game Type

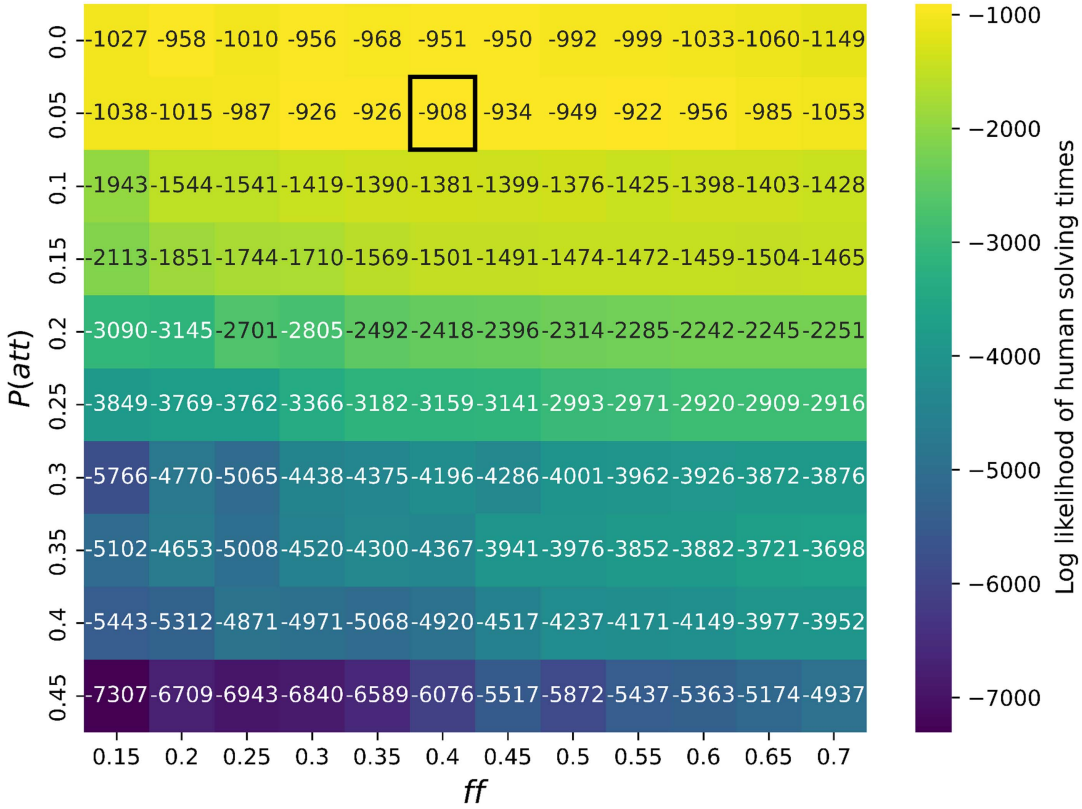
Game type	Best fitting parameter		Difference in LL	
	P(*att*)	*ff*	*M*	95% CI
Logic	0.0	0.4	12.87	[10.87, 14.43]
Contingency	0.0	0.4	103.4	[100.8, 106.1]
Switching mappings	0.2	0.6	−9.380	[−11.34, −8.034]
Switching embodiments	0.35	0.4	9.068	[8.778, 9.345]
Switching embodiments (every 10)	0.15	0.4	1.673	[1.154, 2.112]
Contingency + goal uncertainty	0.2	0.4	53.90	[51.07, 56.56]
Switching mappings + goal uncertainty	0.25	0.65	−0.1396	[−0.3707, 0.04249]
Noisy contingency	0.1	0.4	−1.240	[−1.585, −0.9619]
Contingency2	0.4	0.4	7.901	[−1.585, −0.9619]
Contingency6	0.0	0.4	19.59	[18.87, 20.33]
Contingency8	0.0	0.4	0.001332	[−0.1364, 0.1333]

Note. For each game type, we ran 1,000 Monte Carlo cross-validation runs, in which the human data for that game type was split evenly into train and test sets, and the parameter combination which maximized the likelihood of the training data under the resource-limited model was selected. The best fitting parameters column shows the most frequently selected parameter combination across runs. The mean difference in LL shows the mean difference across runs between the log likelihood of the testing data under the resource-limited model with parameters fit across game types, $P(att) = 0.05$, $ff = 0.4$, and the log likelihood of the testing data under the resource-limited model with the selected parameter combination for that run. It also shows the 95% bootstrapped CI over the mean difference, where an interval with all values greater than 0 suggests that fitting the model specifically on data from the given game type provides a significantly better explanation of the data than fitting the model across game types. CI = confidence interval; *ff* = forgetting factor; *att* = attention; LL = log likelihood.

(*Appendices continue*)

Figure C1

Resource-Limited Meta-ePOMDP Agent Parameter Grid Search Results



Note. Results of grid search for selected resource-limited meta-ePOMDP agent parameter combinations, where $P(att)$ is the probability of attending to any additional character other than the single selected character at each timestep, and ff is the action mapping forgetting factor. Cell colors indicate the log likelihood of human solving times under the resource-limited model with the given parameter combination, averaged over 1,000 split-half cross-validation runs. The cell for the likelihood-maximizing parameter combination used in main analyses, $P(att) = 0.05$, $ff = 0.4$, is outlined in black. ePOMDP = extended partially observable Markov decision process. See the online article for the color version of this figure.

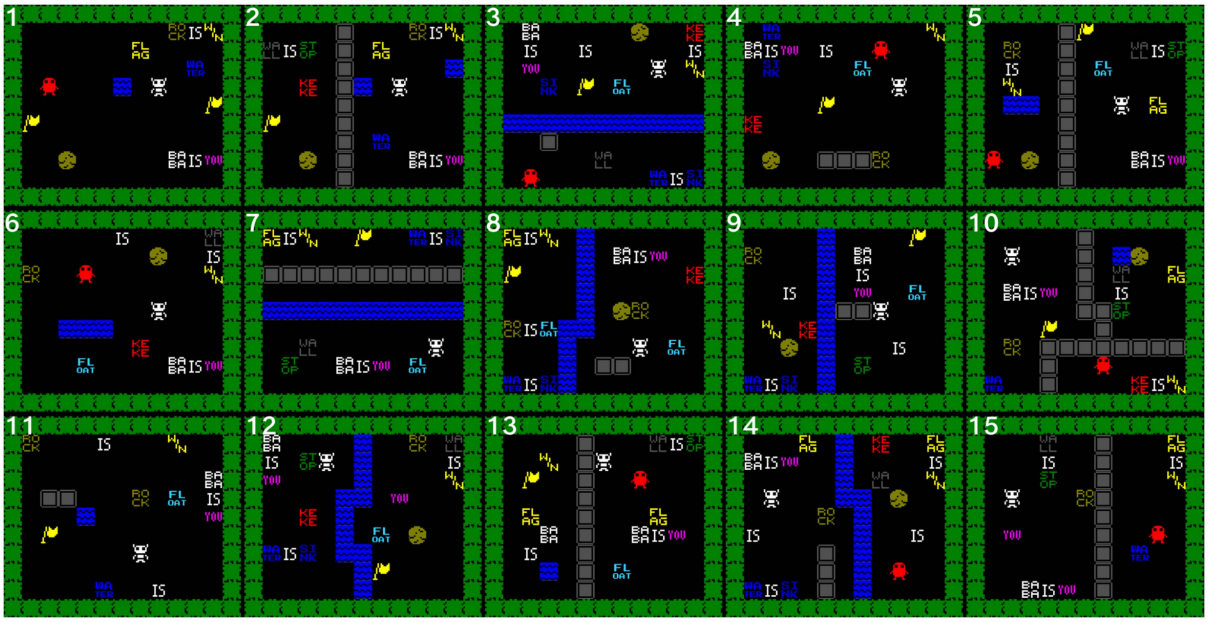
(*Appendices continue*)

Appendix D

Baba Is You Game Details

Figure D1

All Baba Is You Game Types, High Distractor Variants



Note. Snapshots of the 15 high distractor variant Baba Is You games as seen by human participants. Each game includes distracting objects and rule blocks not necessary to solve the game. Different games are separated by a green hedge border, which acts as a barrier in all games. Participants saw one game at a time. Numbers at the top left corner of each game are used to refer to that game type throughout the text. In each game, the words (rule blocks) specify the rules when arranged three in a row and can be manipulated by the player. Graphics from "Baba Is y'all: Collaborative Mixed-Initiative Level Design," by M. Charity, A. Khalifa, and J. Togelius, *2020 IEEE Conference on Games (CoG)* (pp. 542–549), 2020, Institute of Electrical and Electronics Engineers. CC BY 4.0; Charity et al. (2022). See the online article for the color version of this figure.

Received May 21, 2024
Revision received January 7, 2026
Accepted February 12, 2026 ■