

I can tell you know a lot, although I'm not sure what: Modeling broad epistemic inference from minimal action

Rosie Aboody (rosie.aboody@yale.edu), Isaac Davis (isaac.davis@yale.edu)
Yarrow Dunham (yarrow.dunham@yale.edu), Julian Jara-Ettinger (julian.jara-ettinger@yale.edu)
Department of Psychology, Yale University

Abstract

Inferences about other people's knowledge and beliefs are central to social interaction. In many situations, however, it's not possible to be sure what other people know because their behavior is consistent with a range of potential epistemic states. Nonetheless, this behavior can give us coarse intuitions about how much someone might know, even if we cannot pinpoint the exact nature of this knowledge. We present a computational model of this kind of broad epistemic-state inference, centered on the expectation that agents maximize epistemic utilities. We evaluate our model in a graded inference task where people had to infer how much an agent knew based on the actions they chose. Critically, the agent's behavior was always underdetermined, but nonetheless contained information about how much knowledge they possessed. Our model captures nuanced patterns in participant judgments, revealing a quantitative capacity to infer amorphous knowledge from minimal behavioral evidence.

Keywords: Computational modeling; Theory of Mind; Epistemic inference

Introduction

Imagine going to your friend's house for dinner and, as you're cooking together, realizing that you'll need more flour. As the two of you head out, you notice that your friend immediately starts walking in the direction of a large supermarket, rather than her usual go-to bodega around the corner. From this simple decision you might quickly suspect that she knows something you don't. Perhaps the bodega doesn't carry flour; maybe it's cash only and your friend intends to use her credit card; or the supermarket might be the only place that's open late. Inferences like these not only enable us to make sense of others' behavior, but also help us decide when to share what we know, and from whom to learn what we don't, forming a cornerstone of complex social action.

The ability to interpret other people's behavior in terms of mental states, called a *Theory of Mind*, has its origins in early childhood. From infancy, we interpret other people's behavior as goal-directed (Woodward, 1998) and infer others' goals and preferences by assuming that agents act to maximize utilities—the difference between the costs they incur and the rewards they obtain (Csibra, 2003; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Liu, Ullman, Tenenbaum, & Spelke, 2017). Throughout our life, this expectation enables us to make a variety of judgments, such as inferring what others like (Lucas et al., 2014; Jern, Lucas, & Kemp, 2017), predicting how they might behave (Jara-Ettinger, Schulz, & Tenenbaum, 2020), and determining their

social affiliations (Jern & Kemp, 2014; Ullman et al., 2009; Jara-Ettinger, Tenenbaum, & Schulz, 2015).

As the example above shows, however, inferences about others' minds are not restricted to goals and preferences: they also include judgments about what others may or may not know. Consistent with this, research in computational social cognition has found that people can make quantitative inferences about others' beliefs based on their behavior (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). This work showed that a computational model of joint belief-desire attribution, embedded in a Bayesian framework for action understanding, captures how people determine what an agent believes about their environment given their behavior (e.g., if an agent looking for lunch walks towards the end of the block, peeks around the corner to see a Mexican food truck, and then turns around, we can infer that the agent was hoping to see a different food truck there).

While this work shows that people can make targeted belief inferences, such as determining whether an agent knew the type of food a vendor might be selling based on their behavior, these inferences often require access to a relatively constrained hypothesis space and key actions that reveal the agent's beliefs. In many everyday situations, however, we may not immediately know the most relevant epistemic hypothesis to consider, and other people's behavior may not contain the amount of information needed to disambiguate between different degrees of knowledge. In cases like these, our representations of other people's epistemic states appear to consist of amorphous estimates of how much others know, without being sure exactly what it is that they know. Returning to the example in the introduction, when your friend chose to go to the supermarket it was easy to infer that she knew more than you did, despite not knowing exactly what she knew. What computations underlie such epistemic inferences?

Research investigating the ability to estimate and quantify how much others know—intuitions about how much people know without specifying exactly what they know—has generally focused on children. By early in preschool children can represent how much others know about a domain, without needing to list the full contents of their knowledge (Landrum & Mills, 2015; Lutz & Keil, 2002). However, to our knowledge, no work has explored our capacity to infer knowledge magnitude from others' actions, or specified the computations that might underlie such inferences.

Here we propose that such inferences are supported by an expectation that agents maximize utilities, through a sensitivity to the apparent costs that agents choose to incur. Specifically, we suggest adults understand that the costs agents incur often depend on the knowledge they possess. Thus, an ability to infer the subjective costs that an agent appears to act under can reveal the amount of knowledge they might have. In the example above, for instance, the fact that your friend chose to incur a seemingly higher cost (walking to a place that was farther away) for the same reward (getting flour), suggests that she possessed privileged information—leading her to conclude that the large supermarket was a better option than you’d originally assumed.

In this paper we present a computational model of epistemic quantification through an expectation that agents maximize utilities, and we test its performance on a task where participants must infer how much someone knows based on their behavior. Our work shows that people can seamlessly make graded quantitative estimates of how much someone knows, and that these inferences can be explained through an expectation that agents maximize utilities (the difference between the costs they incur and the rewards they obtain), and an understanding that the costs agents incur depend on the knowledge they possess.

Computational Framework

Our computational framework builds on a recent family of computational models of mental-state inference structured around an expectation that agents act rationally—formalized as a generative model of utility maximization, combined with a mechanism for inverting this causal model via Bayesian inference (Lucas et al., 2014; Jern et al., 2017; Baker et al., 2017; Jara-Ettinger et al., 2020). We extend this framework by proposing that adults often expect agents’ costs to be mediated by their knowledge state—and can thus infer others’ epistemic states from observing the apparent costs they choose to incur.

In this project, we consider scenarios in which an agent’s knowledge state affects the expected cost of an action plan. In general, for a given action plan a , we can write $E[Cost(a)|k]$ to mean “the expected cost of implementing action plan a , given knowledge state k ”. If the reward of completing action plan a is $R(a)$, then the total expected utility of action plan a , given knowledge state k , is $R(a) - E[Cost(a)|k]$.

Suppose now that the agent has knowledge state k and set of possible action plans $\{a_1, \dots, a_n\}$. A standard assumption for utility-based agent models is that the agent will compute the expected utility of each action plan $R(a_i) - E[Cost(a_i)|k]$, and probabilistically select the best option through softmaxing. By placing a prior distribution $P(k)$ over possible knowledge states $k \in K$, a Bayesian observer can make inferences about the agent’s knowledge state k , given the observed actions a according to Bayes’ rule:

$$P(k|a) \propto P(a|k)P(k) \quad (1)$$

This model allows observers to infer a posterior distribution over what an agent knows given their behavior. As the example in the introduction shows, however, many situations are under-determined, and in these cases it is not possible to infer exactly what an agent knows. In our framework, this situation arises when an agent’s behavior is consistent with a range of different knowledge states, and it is thus impossible to determine under which exact knowledge state the agent was acting. But even when we can’t infer the precise contents of others’ knowledge representations, we may still be able to infer approximately how much they know (getting a rough sense of how knowledgeable they are). Thus, given a posterior distribution over what the agent might know, we formalize the quantity of amorphous knowledge Q as the expected quantity of knowledge encoded in the probable epistemic states that the agent has, given by

$$Q = \sum_{k \in K} |k|p(k|a) \quad (2)$$

where K is the set of all possible epistemic states, $|k|$ is a quantification of how much the agent knows in that state, and $p(k|a)$ is the posterior probability of that knowledge state (Eq. 1). Naturally, precisely defining the measure $|k|$ may be highly context-sensitive. Here we focus on its application in a particular experimental context but return to the idea of how this might generalize in the discussion.

To evaluate this framework, we considered cases like those shown in Figure 1. Here, an agent is deciding which of two fields to go on an easter egg hunt in, knowing that each field has only one egg with a prize inside (and that the reward is always the same in every field). In this paradigm, a knowledge state k consists of a subset of eggs that the agent might be aware of. The cost the agent incurs depends on their search trajectory. We assume that agents navigate efficiently in space (Csibra, Bíró, Koós, & Gergely, 2003), and search only locations where they think they may find the prize (that is, agents should not visit a location they are sure does not contain the prize). This implies that when the agent’s knowledge state includes information about the contents of the prize egg, the agent will always move directly towards the prize. Otherwise, the agent will search the eggs in a way that minimizes the expected search time. Finally, to compute equation 2, we treat the amount of knowledge in an epistemic state as $1 -$ the proportion of eggs the agent is still uncertain about (if the agent knows where the prize is, they know the rest of the eggs are empty, and thus the proportion known is 1; if the agent is unsure about half of the eggs, the proportion known is .5; etc.).

To produce a prediction for a given scenario, our generative model sampled 10,000 knowledge states over the contents of each field, predicted which field the agent would visit (by softmaxing the expected utility of each field given the agent’s knowledge), and recorded both the agent’s knowledge state over each field, and the agent’s predicted choice. To perform epistemic inference, when given an agent’s actual choice of field in this scenario, the model returned the average propor-

tion known about each field under cases where the predicted choice was consistent with the observed choice.

Our model has four parameters: the reward of obtaining the prize (set as a constant $R(a_i) = 100$), the cost of checking an egg’s contents upon reaching it (randomly chosen for each knowledge sample from the continuous uniform distribution $[1, 3]$), a prior over the probability the agent knew the contents of each egg (0.5; this prior was also communicated to participants in our task), and the softmax parameter ($\tau = 3$). All parameter values and model predictions were preregistered prior to data collection (see project OSF page: https://osf.io/q3szu/?view_only=6e23c6b79dc24e21acb5a332b0a190ed).

Alternate Model

Our main model assumes that people quantify the cost of obtaining the prize in each field under different degrees of knowledge, and then reason about the knowledge states under which the agent’s actions would have been utility-maximizing. However, it is possible that adults generally do not apply such complex computations when inferring others’ knowledge states, and instead rely on simpler rules or heuristics. Such heuristics could get things right most of the time, while requiring less effort to apply.

To address this possibility, our alternate model encoded the simple heuristic that agents tend to choose options they know more pieces of information about. Critically, this alternate model did not consider agents’ knowledge states in a full mentalistic way: it did not compute the utility of each field based on the agent’s knowledge state, and did not expect agents to navigate directly to an egg if they knew it contained the prize. It simply considered the proportion of eggs with known contents in each field, and expected the agent to always choose the field where this proportion was larger (or choose randomly when this proportion was equal across fields). We then generate predictions from this alternate model using the same sampling procedure as in the main model. This model was not preregistered, but uses only one parameter: the same knowledge prior as in our main model.

Experiment

To test our model, we designed a task where an agent’s behavior (and its costs) could reveal approximately how much they knew—but was too impoverished to reveal precisely what they knew. Participants watched an agent choose which of two fields to search for a prize hidden in an easter egg. The cost of locating the prize in any given field was determined by the number of eggs, their spatial distribution, and the true location of the prize. By manipulating all three variables, we test if participants infer how much others know by quantifying and comparing their expected costs—or whether participants rely on a simpler heuristic that does not require them to track or reason about others’ costs when inferring epistemic states. Our procedure, stimuli, sample size, and analysis plan for our main model were preregistered (see OSF page).

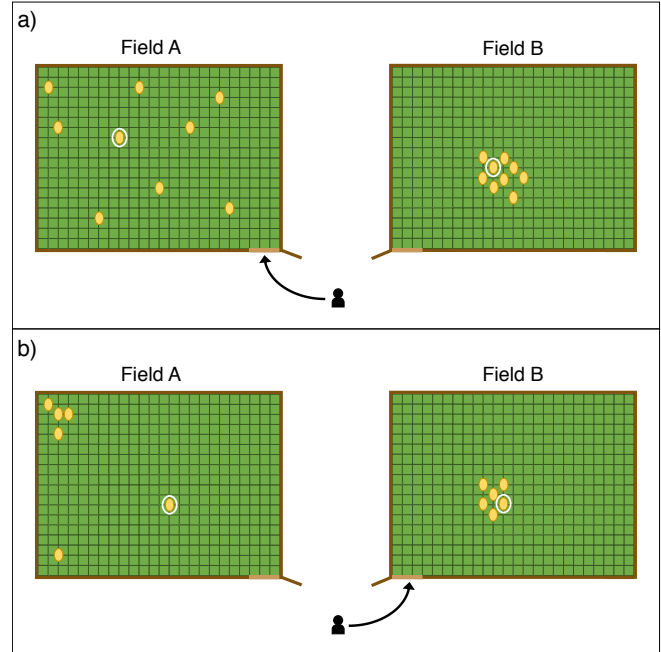


Figure 1: Example of the experimental stimuli. The arrow indicates the agent’s chosen field; eggs containing a prize are circled. Panel A depicts a strong epistemic contrast: here, you might infer that the agent knows approximately where the prize is located in their chosen field, and very little about the other field. Panel B depicts a more graded contrast: here, you might suspect that the agent knows more about the prize’s location in their chosen field, but may be less certain they know a lot (because their chosen field is also much less costly to search).

Participants

40 adult participants with U.S.-based IP addresses were recruited via Amazon Mechanical Turk ($M = 35.05$ years, $SD = 9.23$). 7 additional participants were recruited but excluded from the study for failing a preregistered inclusion trial.

Stimuli

Stimuli consisted of 19 test trials, plus one inclusion trial. The test trials were presented in a randomized order, and the inclusion trial was always presented last. Each trial showed an agent, and two fields. The fields each had easter eggs placed inside, and one egg in each field contained a hidden prize. This egg was circled for participants. An arrow indicated the agent’s path to their chosen field, thus showing which field the agent chose to visit on each trial (see Figure 1).

Stimuli were based on three scenarios (pairs of fields) we thought could elicit a range of model ratings. To manipulate the cost of searching each field, eggs in the first field (field A) were always wide-spread. The second field (field B) contained the same number of eggs, but these eggs were instead clustered near the middle of the field. The first scenario is

shown in Figure 1a. The second scenario was based on the first: we selected a subset of 6 eggs from each field, thus varying the number of eggs but not their position. The third scenario was in turn based on the second, but here we instead varied the position of the eggs in field A (capturing a case where most of the eggs in field A were extremely costly; see Figure 1b).

To select the final locations of the prize in field A, we provided each scenario as input to the model, but systematically varied which egg in field A contained the prize, yielding 42 trials (21 unique scenarios x 2 choices per scenario).¹ We selected 24 trials (12 unique scenarios x 2 choices per scenario) that both produced a range of model responses, and were not too similar to each other. In preparation to present stimuli to participants, some trials were mirrored, and we slightly varied the position of the prize in field B amongst similar scenarios (to prevent participants from noticing similarities between trials).² We then obtained final model predictions, and excluded any trials where the model’s knowledge predictions were based on less than 500 samples (that is, where the predicted choice of field was consistent with the observed choice in less than 5% of cases). This yielded 19 final trials; this criterion and our final set of stimuli was preregistered.

Procedure

Participants were introduced to an agent going on easter-egg hunts in a two-dimensional grid-world. Participants learned that a farmer had placed easter eggs in his fields, hiding a prize inside one egg in every field. This prize (one silver token) was always the same in every field, and the prize egg was always circled for participants.

Participants learned that because the grass in the fields was quite short, the agent could always see where the eggs were located in a field before entering it. But while the prize egg was circled for participants, the agent didn’t necessarily know which egg contained the prize. Participants learned that the agent had seen the farmer set up some of the eggs; it was unclear what prior over knowledge participants would bring to the task, so we specified that the agent had a 50/50 chance of knowing the contents of any given egg. And participants were explicitly instructed that the agent did not always know the same amount about every field; the amount she knew about the location of the prize in each field could differ.

¹We did not expect the location of the prize in field B to strongly affect the model’s predictions; to test if this was the case, we did also replicate one scenario given a different prize location in field B, yielding an additional 18 additional trials. The location of the prize in field B indeed had little effect (as all of these eggs are so close to each other), and thus we selected our final stimuli by considering primarily the location of the prize in field A.

²Despite slightly varying the prize’s location in field B across similar trials in our preregistered stimuli, our model predictions were not updated accordingly prior to preregistration. Because we collected our data using the preregistered stimuli, we obtained new model predictions for any trials where the location of the prize in the stimuli did not match the coordinates originally used to generate the preregistered model predictions. We used the same preregistered parameters.

Participants learned that the agent always had to choose between two fields, and could only search the field she chose. An arrow indicated which field the agent had chosen to search (see Figure 1). Participants were oriented to factors that might affect the agent’s search decision: they were told that the agent always wanted to find the prize as quickly and easily as possible, and that the difficulty of finding the prize was determined by the number of eggs in a field, their distance from the entrance, and the amount the agent already knew about the location of the prize. Note that while this tutorial ensured participants were attentive to the main features of our task, we are interested in how participants combine these different pieces of information and reason over them to infer what others know. The tutorial did not specify how participants should weight or use any of these features in their judgments.

To access the task, participants then completed a preregistered inclusion quiz that assessed their understanding of the task instructions. Participants were given two chances to pass the inclusion quiz; those who failed on their first attempt were required to review the task introduction before trying again. Participants who failed both attempts were not given access to the task. Upon passing the inclusion quiz, participants then completed the 19 test trials (presented in a randomized order), plus one inclusion trial at the end. For each trial, participants were asked to rate, on a sliding scale from 0 - 100, how much the agent knew about the location of the prize in each field. Critically, participants rated how much the agent knew about both fields, not just the field she had chosen. The preregistered inclusion trial always came last. It was similar to the test trials, but presented an extreme contrast where we could make a strong prediction about the pattern of judgments an attentive participant should make. Participants whose judgments differed from our preregistered criteria were excluded. Finally, participants were asked what they thought the point of the task had been, and were given an opportunity to provide feedback or note any technical difficulties.

Results

Participants rated the agent’s knowledge about both fields in 19 test trials, yielding 38 ratings. As preregistered, participant responses were averaged by question, and then z-scored; the corresponding model predictions were also z-scored.

Figure 2 shows the overall results, revealing that our model was highly correlated with participant judgments, $r = 0.94$ (95% CI: 91.78, 98.84). And this correlation did not reflect only cases where both the model and participants inferred a lot of knowledge or very little knowledge. Critically, it included cases where both the model and participants were equally uncertain, in a graded manner, about how much the agent knew. Figure 3 plots the trial-by-trial correspondence between model and participant ratings, showing that participants’ judgments were not bi-modal, but rather graded in a way that closely tracked our model’s predictions.

To ensure that these results could not be the product of a simple heuristic, we implemented an alternate model. Rather than performing full mental-state inference, our alternate

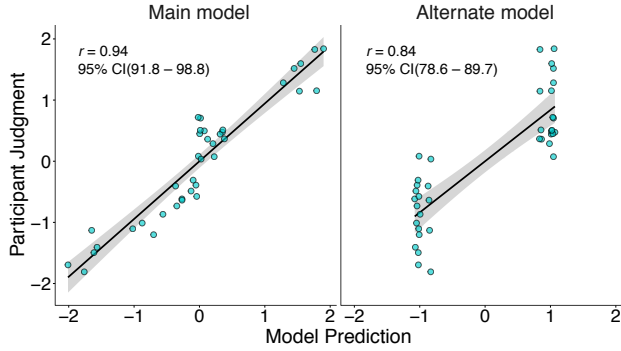


Figure 2: Comparison between our model and the alternate model, with linear regressions fit to each dataset. Each point represents one knowledge rating, with model predictions on the x axis and participant judgments on the y axis. Gray bands show 95% confidence intervals in the regression.

model simply assumed that agents always choose fields where they know about a greater proportion of eggs. Note that we only preregistered an analysis plan for our main model, but test the performance of the alternate model in the same way. The alternate model showed a weaker correlation with participant judgments, $r = 0.84$ (95% CI: 78.58, 89.77), demonstrating that the amount of locations an agent knows about in each field does matter, but that predictions made on the basis of this one factor (without considering costs) do not capture the full pattern of participant judgments. A bootstrap over the correlation difference revealed that the main model was reliably better correlated with participant judgments than the alternate model (correlation difference, alternate model – main model = -0.11 , 95% CI: $-17.34, -4.36$; not preregistered). As Figure 2 reveals, although the correlation between the alternate model and participant judgments was still high, this is only because the alternate model categorized every judgment into two rough bins. These predictions were approximately correct, but lack the nuance that participants’ epistemic inferences showed, and that our model was able to capture.

Discussion

Here we presented a computational model of amorphous epistemic inference. Our model sought to explain people’s capacity to infer how much others might know in situations where inferring their precise knowledge states would be difficult (due to a lack of diagnostic actions, and large hypothesis spaces over potential epistemic states). Our computational framework was based on a growing body of research showing that mental-state inference is structured around an assumption that agents act to maximize utilities—the difference between the costs that agents incur and the rewards they obtain (Jara-Ettinger et al., 2016; Gergely & Csibra, 2003; Lucas et al., 2014; Jern et al., 2017). Our model builds on these ideas, and extends them by explicitly modeling the idea that the costs agents expect to incur often depend on the knowledge they

possess. Consequently, others’ choices can reveal the costs they appear to be acting under, providing indirect insight into the nature and quantity of their privileged knowledge.

The results of our experiment suggest that people can derive graded estimates of how much others know based on limited observable action, and their inferences matched our model predictions in a quantitative manner. Moreover, an alternative model that inferred knowledge through an assumption that agents always pursue options they know more about (without considering how this knowledge would affect agents’ subsequent search behavior, or their costs) failed to capture the graded structure of participants’ judgments.

Related work has developed computational models that explain how people infer each other’s beliefs about the world (Baker et al., 2017) as well as beliefs about their own competence and preferences (Jara-Ettinger et al., 2020). These inferences, however, often depend on access to a limited set of epistemic hypotheses, and to observable behavior that is diagnostic of the agent’s epistemic state. While these inferences are undoubtedly critical for social interaction, many everyday social behaviors lack the information needed to make such precise and targeted epistemic inferences. We show that, in such situations, people can nonetheless derive quantitative estimates of how much knowledge someone might possess.

To make these broad epistemic inferences, our model considered a large space of possible epistemic states and weighted the amount of knowledge expressed in each epistemic hypothesis by its posterior probability (Equation 2). This highlights two critical assumptions that our model makes. First, observers must have access to a range of epistemic hypotheses that they can evaluate; second, they must have a way to quantify how much knowledge is expressed in each hypothesis. While the first assumption may seem plausible in common situations, there are many cases where we do not know what other people’s epistemic states could look like (e.g., we can represent that pilots know how to fly planes while having no idea precisely what a pilot knows). Similarly, the second assumption (that it is possible to quantify the knowledge encoded in each hypothesis) was easy to formalize in the experimental context that we considered. But this is not always the case; and, as illustrated by the pilot example, sometimes we may not even know what there is to know.

These assumptions highlight an important limitation of our findings. Here, our goal was to describe epistemic inference at a computational level (Marr, 1982), not an algorithmic one. It is quite possible participants did not consider the same kinds of hypotheses our model did. And if participants did sample knowledge states in the same way our model did, they may have come to the same conclusions based on far fewer samples (Vul, Goodman, Griffiths, & Tenenbaum, 2014); future work should test this possibility. Furthermore, while our alternate model demonstrates participants were not using a simple cost-insensitive heuristic, it is an open question whether other heuristics could explain participant judgments. Future work will investigate whether this is the case.

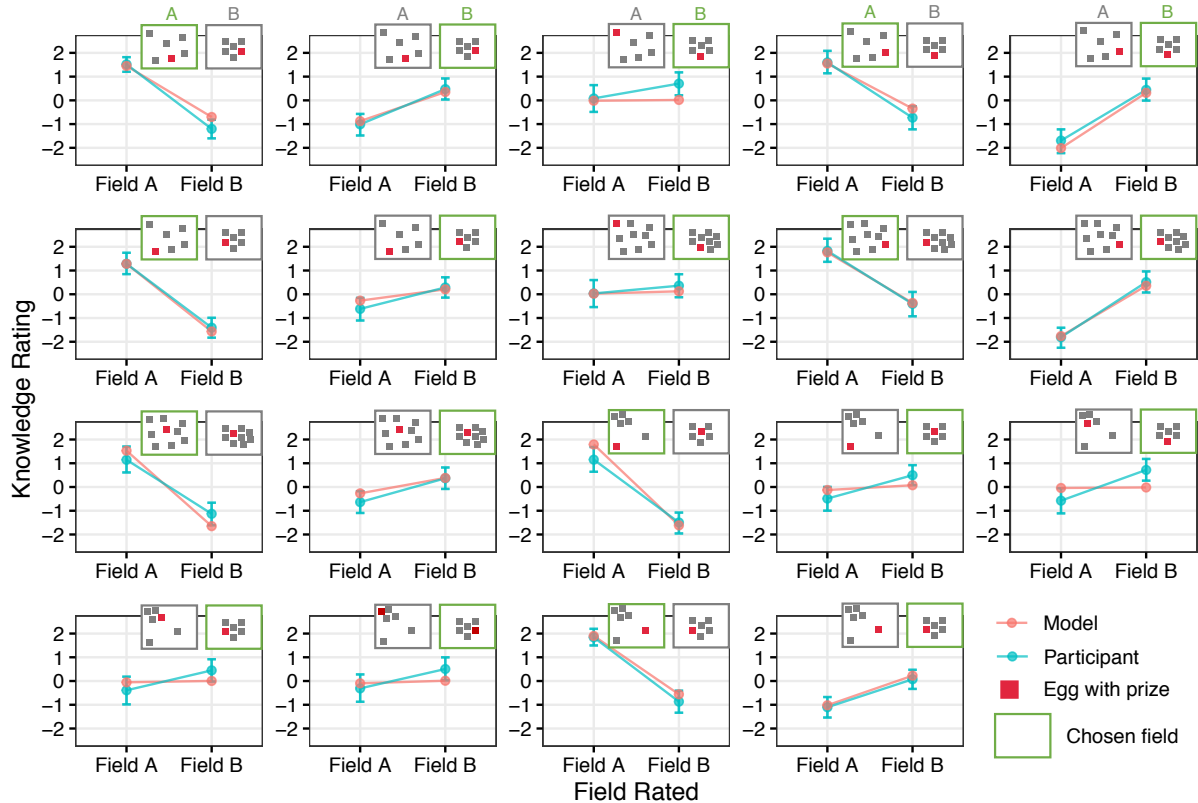


Figure 3: Detailed results for the experiment. Each panel presents one trial, with results split by the field rated (Field A or Field B, indicated on the x axis). The y axis indicates standardized knowledge ratings. Participant judgments are plotted in blue; model predictions are plotted in red. Vertical bars show 95% confidence intervals over participant judgments. The schematics show the position and number of eggs in each field, the egg with the prize, and the field the agent ultimately chose in each trial.

Our findings raise two additional open questions. First, because each field contained only one prize, it was not necessary for the agent to see inside every egg to have full knowledge. Specifically, if the agent knew which egg contained the prize, they could conclude that the rest of the eggs in the field were empty by default. Thus, while the agent’s knowledge was often graded (if she didn’t know where the prize was, she could still avoid opening eggs she knew were empty), it wasn’t always (if she knew where the prize was, she also knew where it was not). While such situations are commonplace (if you find the silverware drawer in a new kitchen, you don’t need to keep searching), knowledge is not always all-or-none, and future work should continue to explore adult epistemic inference in a wider variety of contexts. Relatedly, in future work it will be important to investigate whether there are any general measures, such as entropy, that might support quantifying knowledge independent of context.

Finally, we proposed that our task elicited, and model captured, amorphous knowledge inferences. However, we did not test whether participants in fact made specific inferences in this task. While many knowledge states could have given rise to each observed choice, it is still true that knowledge over certain eggs affects the expected utility of an action plan

more than others. Participants may in fact be able to make specific inferences in such cases (for instance, inferring that an agent who chooses an apparently more costly field may know that some of the furthest eggs are empty). However, participants should be more uncertain when many knowledge states affect the expected utility of an action plan in the same way. Our model already infers the probability that each egg’s contents are known by the agent; future work could test whether these predictions capture participant judgments as well.

To conclude, in the current work we explain a common everyday epistemic inference: the ability to infer how much others know even when specific inferences about what they know are under-determined. But as we navigate the world, these are not the only epistemic inferences we make. From observing the outcome of others’ goal-directed actions we infer what they thought they knew (and what they actually did); from observing their choices to seek information (and at what cost) we infer how much they believed they could learn; and from observing their interactions with other agents, we infer how much they think others know. The work presented here serves as an initial step towards modeling the full scope of epistemic inferences people make in their everyday lives.

Acknowledgments

We thank the members of the Yale Computational Social Cognition Lab for helpful conversations and advice. This work was supported by NSF award BSC-2045778, and also by Yale's Franke Program in Science and the Humanities, via a Franke Interdisciplinary Graduate Award to RA.

References

- 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), 0064.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philos Trans R Soc Lond B Biol Sci*, *358*(1431), 447–458.
- Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.
- 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.
- Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.
- Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science*, *26*(5), 633–640.
- Jern, A., & Kemp, C. (2014). Reasoning about social choices and social relationships. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.
- Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition*, *168*, 46–64.
- Landrum, A. R., & Mills, C. M. (2015). Developing expectations regarding the boundaries of expertise. *Cognition*, *134*, 215–231.
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*(6366), 1038–1041.
- Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., . . . Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PloS one*, *9*(3), e92160.
- Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*(4), 1073–1084.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.
- Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems* (pp. 1874–1882).
- Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599–637.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.