Title
Integrating Specialist Judgments With and Without Mentalizing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Zhao, Danqin
Bramley, Neil R.

Publication Date
2025

Integrating Specialist Judgments With and Without Mentalizing

Danqin Zhao (d.zhao-24@sms.ed.ac.uk), Neil R. Bramley (neil.bramley@ed.ac.uk)
Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom

Abstract

In daily life, we frequently interact with specialists—individuals whose complementary insights augment our limited first-hand experiences in decision-making. Efficient as this may be, the cognitive demands of employing our theory of mind to determine and integrate the evidential contributions of diverse specialists may offset its benefits. In this study, we explore human judgments in a scenario where they must aggregate opinions from multiple social peers, each possessing expertise in a different aspect of the problem. We examine whether participants integrate specialist judgments with or without mentalizing using a normative Bayesian model and propose two heuristic approaches. Our results show the majority of participants across two experiments relied on heuristics, suggesting that people don't tend to or have limited ability to integrate specialist judgments through mentalizing.

Keywords: social inference; theory of mind; mentalizing; Bayesian modeling

Introduction

To form beliefs and make advantageous decisions, we often rely on social information in addition to our first-hand experiences. For instance, when deciding where to go on holiday, we might seek advice from friends, read travel blogs, and check local reviews on Google Maps. We refer to various social specialists who have useful and complementary experience or insight to evidence we lack—e.g. they have been there before, speak the language, etc (Gray et al., 2012). Similar considerations arise in government and commercial policy decisions. For example, when choosing a location for an infrastructure project, stakeholders will generally consult a variety of experts with different specialisms: architects, geological surveyors, and market researchers.

Integrating opinions rationally from diverse sources requires sophisticated Theory of Mind (ToM), also known as mentalizing, whereby we reverse engineer the evidential weight behind the beliefs and values that make sense of their behaviors or testimonies (Wellman et al., 1990). Aggregating evidence from diverse specialists with different knowledge bases presents an interesting and understudied facet of social reasoning, that we focus on in this paper.

Previous studies show that people are capable of considering an informant's expertise when evaluating advice (e.g., Hawthorne-Madell & Goodman, 2019). In early childhood, humans already preferentially learn from social agents who are more knowledgeable about relevant topics (Robinson et al., 2011; Koenig & Harris, 2005; Jaswal & Malone, 2007; Sabbagh & Baldwin, 2001). For example, when being asked what was inside a box, children tended to trust the suggestion of an informant who had seen the box (Robinson et al., 2011), indicating their capability of recognizing knowledgeability. Adults can even use indirect evidence of knowledgeability to guide social decisions, such as using an agent's past performances to infer their knowledgeability and selectively taking advice from well-performing agents (Hawthorne-Madell & Goodman, 2019; Yaniv, 2004). This evidence supports the human ability to use ToM to guide social information intake.

Our experiments explore a more minimal, abstract, yet inferentially rich scenario where participants must aggregate opinions from several social peers, whom they know to have expertise in different aspects of a problem. We examine how people aggregate social testimony holding this up to a Bayesian rationality standard. Concretely, we will compare participants' behavior to the predictions of a normative Bayesian model, and also examine several plausible heuristics, through which people might aggregate other agents' judgments at face value (Fränken & Pilditch, 2021; Fränken et al., 2024) using simple rules, such as counting votes and prioritizing agreement among more diverse sets of specialists.

Task

Our social learning scenario adapts the urn-and-ball task (Xie & Hayes, 2022; Whalen et al., 2018; Garety & Freeman, 1999). Our task involves two jars, containing balls of three colors: red, yellow, and blue. Crucially, we introduce different private knowledge (their 'expertise') about the content of the jars for the participants and their artificial peers through a cover story: In Experiment 1, participants observe both jars being filled, so they are aware of the full distribution of balls in both jars, while social agents Andy and Anna witness only the addition of blue balls, leaving the room while the red and yellow balls are added; similarly, Bella and Bryan observe only the addition of yellow balls (Figure 1a). Thus, when one of these social agents makes a guess about the jar a ball is drawn from, what the participant should conclude from this depends in subtle ways on the content of the jars, as we will unpack with our normative model below. In Experiment 2, participants' knowledge is also restricted. Participants only observe the addition of red balls, making the problem of interpreting, weighing, and integrating the guesses of their peers more difficult (Figure 1b).

Notably, the total number of balls in each jar is common

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

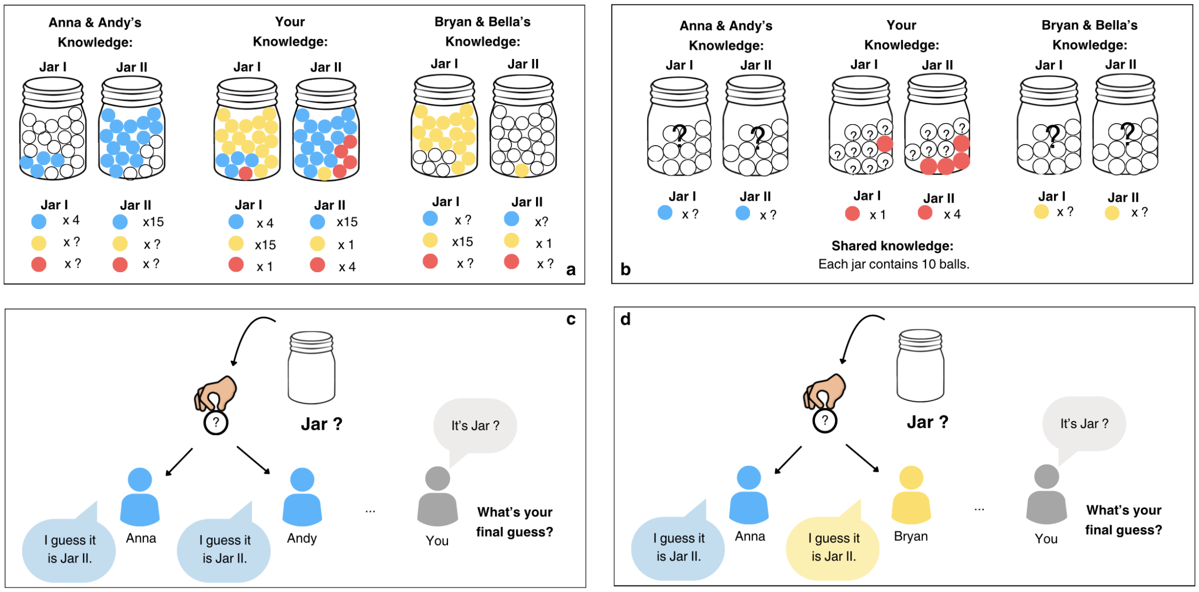

Figure 1: Experimental Interface: (a) In Experiment 1, participants have full knowledge of the content of both jars. In contrast, blue agents (Anna and Andy) only know the number of blue balls in each jar, whereas yellow agents (Bryan and Bella) only know the number of yellow balls in each jar. (b) In Experiment 2, participants' knowledge is also limited to the number of red balls in each jar. (c) In each trial, a ball is randomly drawn from an unknown jar and shown to two social agents, who independently guess whether it comes from Jar I or Jar II. Participants then integrate the agents' opinions to make a final judgment. In the no-diversity condition, participants collaborate with social agents share the same knowledge. (d) In the with-diversity condition, participants collaborate with social agents who possess knowledge about different components of the jars.

knowledge among both participants and all social agents. In Study 1, each jar contains 20 balls, while in Study 2, each jar contains 10 balls. These total numbers of balls were set to balance task complexity with model discriminability. A higher number of balls in Experiment 1 allows for greater variability in numbers of balls with different colors and thus clearer model discriminability. In contrast, Experiment 2, where participants have limited knowledge, is already cognitively demanding. Therefore, we reduced the number of balls to 10 in order to manage the difficulty of the task.

In each round, our participants collaborate with two other agents (e.g. Anna and Andy in Figure 1c). In the example, a ball is drawn from an unknown jar and presented to Anna and Andy, these two players will then infer which jar the ball came from (e.g. 'I think it is Jar I'). Since the participants did not see the ball directly, they must base their judgment on the guesses from the social agents they collaborated with. Because these guesses do not convey confidence level, a rational learner needs to consider the diagnosticity of these guesses in light of the social agents' knowledge and how they might have behaved depending on what ball they observed. There is no communication between the social agents and they are rewarded only if the participants answer correctly. Thus, participants are led to believe that the social agents acted independently and have no incentive to deceive.

To respond optimally in these tasks, participants need to reverse-engineer two players' responses to infer what ball they observed, and further, the more likely jar. In Experiment 2, where participants have even more limited information about Player 1 and Player 2's knowledge, they also need to consider and marginalize over the many possible combinations of jar contents that could lead to two players' responses, see Figure 2.

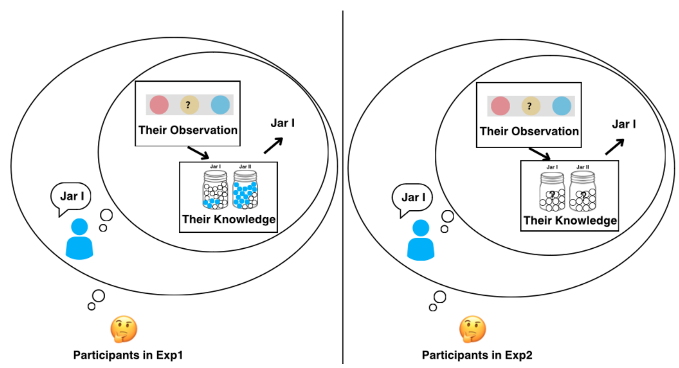

Figure 2: Normative inference: In Experiment 1 (left), rational reasoners need to infer other agents' observations based on their responses. In Experiment 2 (right), they need to jointly infer both agents' observations and knowledge.

Formal framework

Normative Bayesian Model

Let $U \in (u_1, u_2)$ be a random variable, such that $P(U)$ describes the rational social reasoner's subjective belief that jar u_i has been selected for a given trial. Before hearing any agents' judgments, it is equally likely that either jar will be selected, so the prior is $P(U) = [.5, .5]$.

The other agents a_1 and a_2 then make a shared observation of a single draw from selected jar $d \in \{\text{red, yellow, blue}\}$. They then make independent judgments $S_1, S_2 \in (s_1$:"I think it is Jar I", s_2 :"I think it is Jar II") allowing a rational reasoner to form a posterior belief $P(U|S_1, S_2)$.

We know each agent's judgment is based on the shared observation, plus their distinct prior knowledge about the content of the two jars determined by their earlier partial witnesses of the filling of the two jars. We write the agents' knowledge states as H^{a_1} $(H^{a_1}_{u_1}, H^{a_1}_{u_2})$ and H^{a_2} $(H^{a_2}_{u_1}, H^{a_2}_{u_2})$, where $H_{\text{jar}} \in \{h^1, \ldots, h^n\}$ is the set of all possible jar contents, such that $h^i \in [(n_{\text{red}}, n_{\text{yellow}}, n_{\text{blue}}) \in (0, 20)]$ where $n_{\text{red}} + n_{\text{yellow}} + n_{\text{blue}} = 20$ (or 10 in Experiment 2). Concretely, the agent who observes the blue balls being added can rule out all possibilities except those that match their witnesses. For simplicity, we assume that their judgment is a deterministic function (a hard maximum) over their subjective posterior belief about the jar:

$$P_{a_i}(S_i = u_1|d) = \begin{cases} 1, & \text{if } P_{a_i}(U_{a_i} = u_1|d, H^{a_i}) > .5, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Each agent's subjective posterior about the identity of the jar given their observation d is, in turn, proportional to the likelihood of their observing that the ball drawn from each jar, marginalized over their subjective prior beliefs about the possible content of that jar. The subjective posterior of agent a_i is:

$$P_{a_i}(U_i = u_1|d, H^{a_i}) = \Big(\sum_{h^j \in H^{a_i}} P_{a_i}(U_i = u_1|d, h^j) P(h^j|d), \\ \sum_{h^j \in H^{a_i}} P_{a_i}(U_i = u_2|d, h^j) P(h^j|d) \Big) \tag{2}$$

Since in Experiment 1 the rational social reasoner knows the true content of both jars $h^*_{u_1}$ and $h^*_{u_2}$, they know the probability of different draws conditional on the true jar $P(d|U)$. They also know which portion of $h^*_{u_1}$ and $h^*_{u_2}$ was observed by the agents a_1 and a_2, allowing them to infer $P_{a_i}(U|d)$ and hence $P(S_i|U)$ for each potential value of d. Combining these they can infer the posterior over possible draws given the agents' judgments

$$\begin{aligned} P(d|S_1, S_2) &\propto P(S_1, S_2|d)P(d) \\ &\propto P(S_1|d)P(S_2|d)P(d) \end{aligned} \tag{3}$$

allowing them to calculate posterior belief $P(U|S_1, S_2)$ as

$$P(U|S_1, S_2) = \sum_d P(U|d)P(d|S_1, S_2) \tag{4}$$

In Experiment 2, the rational social reasoner also does not know $h^*_{u_1}$ or $h^*_{u_2}$, but instead has their own hypotheses about the component of two jars $H^r(H^r_{u_1}, H^r_{u_2})$ resulting from their earlier partial observation of the jars being filled. This means they must additionally marginalize over H^r both when inferring $P(d|S_1, S_2)$ and again when using this to calculate $P(U|S_1, S_2)$

$$P(U|S_1, S_2) = \sum_{h^j \in H^r} \sum_d P(U|d, h^j) P(d, h^j|S_1, S_2) \tag{5}$$

$$P(d, h^j|S_1, S_2) \propto P(S_1|d, h^j)P(S_2|d, h^j)(d|h^j)P(h^j) \tag{6}$$

In order to translate participants' beliefs r about the identity of the jar into Likert-scale response predictions, we represented them by Beta distributions, which are centered around the participant's subjective posterior estimate, but that admit some second-order uncertainty controlled by confidence parameter $c \in (0, \infty)$:

$$r \sim \text{Beta}(\alpha = 1 + c \cdot P(U = u_2|S_1, S_2), \beta = 1 + c \cdot P(U = u_1|S_1, S_2)). \tag{7}$$

The Beta distribution is defined over the interval [0,1], where a value of 0 indicates certainty in favor of Jar I, and a value of 1 strong certainty in favor of Jar II. As $c \to 0$, the response becomes uniform across the response space. As $c \to \infty$, this function approaches a Dirac delta function centered at the subjective posterior.

Heuristics Models

As well as the Normative Bayesian Model (NB), we also consider several plausible heuristic simplifications. A maximally straightforward approach——the Majority Vote Model (MV)——is to eschew considering what agents' judgments imply about the ball that was drawn, and instead treat them as direct cues to the identity of the jar (i.e. if both agents think it is Jar I, then Jar I is a good guess). Furthermore, it might be sensible to value the guesses of agents with diverse expertise (i.e. one blue-knower and one yellow knower as more valuable than two blue-knowers or two yellow knowers), since their judgments are then also not redundant with one another.

The Majority Vote Model (MV) is based directly on the responses of two agents. Weighing each vote for each jar, V_{u_1} and V_{u_2} by a free parameter $n \in (0, \infty)$, we calculate the posterior as:

$$P_{MV}(U|S_1, S_2) = \left(\frac{1 + nV_{u1}}{2 + nV_{u1} + nV_{u2}}, \frac{1 + nV_{u2}}{2 + nV_{u1} + nV_{u2}}\right) \tag{8}$$

and use this in Equation 7.

In the Diversity Heuristic (DH) model, the strength of evidence is scaled when two agents of different knowledge agree, with the degree of scaling governed by a free parameter η. $\eta > 1$ indicates that participants treat two agents with different specialists as more informative; when $\eta < 1$, in contrast, participants think two agents with different specialists as less informative.

$$P_{DH}(U=u_1|S_1,S_2) = (\frac{1+nV_{u1}\cdot D(P_{a1},P_{a2})}{2+(nV_{u1}+nV_{u2})\cdot D(P_{a1},P_{a2})}, \frac{1+nV_{u2}\cdot D(P_{a1},P_{a2})}{2+(nV_{u1}+nV_{u2})\cdot D(P_{a1},P_{a2})})$$

$$\mathrm{D}(\mathrm{P}_{a1},P_{a2}) = \begin{cases} 1 & \text{If } P_{a1}{=}P_{a2} \\ \eta & \text{If } P_{a1} \neq P_{a2} \end{cases} \tag{9}$$

Experiment 1

Participants

We recruited 103 U.S. or U.K. residents using Prolific. We excluded 2 participants due to missing data. The final sample included 101 participants (53 female, 48 male), with an average age of 42.78 ± 13.66 years. Participants were compensated £1.20, and received an additional bonus between £0 and £0.80. The experiment was pre-registered here.

Design and Materials

In order to distinguish the predictions between normative and heuristic models, we designed 4 pairs of trials (see Figure 3), each consisting of a with-diversity and no-diversity conditions. In no-diversity condition, as all social agents were set to be rational, two agents of the same knowledge always provided the same results based on the same observation. In contrast, in the with-diversity condition, social agents with different knowledge might give different responses based on the same observation. Specifically, in Pair I and IV, two agents were designed to give opposing responses, while in Pair II and III, they provided the same answer. Besides, NB model predicted differently between with- and no-diversity conditions in Pair I and II, but not in III and IV.

Through this design, we are able to distinguish the predictions of these three models. If participants follow the normative Bayesian model, we should observe differences in judgments between the two conditions in Pairs I and II, but not in Pairs III and IV. In contrast, participants following the majority vote heuristic will respond differently between the two conditions in Pairs I and IV, while giving identical responses in Pairs II and III. Lastly, for participants using diversity heuristic, we expect an additional difference in Pair II and III, because the DH model assumes that agreement between two agents with varied knowledge provides stronger evidence than agreement between agents with identical knowledge.

These four conditions (eight trials in total) were implemented as an HTML webpage, available here.

Procedure

After completing a consent form, participants received instructions about the task. After reading the instructions (as described in the Task section), participants answered a quiz of six questions that tested their understanding of the task. If they failed any question, they would be redirected to review the instructions.

Before the main game, participants also completed a rating test about how informative in general they found different combinations of social agents. They rated both with-diversity combinations (e.g. Anna and Bryan), where social agents had different knowledge, and no-diversity combinations (e.g. Anna and Andy) where they shared same knowledge. Next, participants completed the main tasks. The four Pairs of trials mentioned above were presented in a random sequence, and trials within each pair were also randomized. Finally, participants completed a short demographic questionnaire, received feedback on their results, and were then redirected to Prolific for compensation. Notably, all data and code for modeling and analysis are available at github/social-specialist.

Computational Modeling

In our experiments, participants' beliefs were elicited using a 6-point Likert scale (k=6), chosen to avoid central tendency bias (Chomeya, 2010). Following the approach of Fränken et al. (2024), for each continuous belief distribution as $r \sim \mathrm{Beta}(\alpha,\beta)$, we first bin into 6 levels matching the increments of the response scale. $P(b_i)$ is given by the cumulative density of the Beta distribution:

$$P(b_i \mid r) = \int_{\frac{b_i-1}{6}}^{\frac{b_i}{6}} \mathrm{Beta}(x;\alpha,\beta)\,dx \tag{10}$$

We include a softmax function for all models with an additional free parameter β. After calculating the predicted probabilities for each judgment using the cumulative density functions, the softmax function is applied to modulate the sharpness of the probability distribution, simulating how participants' beliefs of which jars it is translate into their decision. A higher value of β accentuates the tendency to select options with higher probabilities, thereby reducing the likelihood of choosing less probable alternatives.

In summary, our analysis includes 3 models, NB (Normative Bayesian), MV (Majority Vote), and DH (Diversity Heuristic) as mentioned above along with a baseline model that assumes participants make choices randomly. For parameter optimization, we use the `optim()` function in R. Model performance is assessed using leave-one-out cross-validation. In each iteration, we train the model with data from one participant taken out. The likelihood of this held-out data is then computed to evaluate predictive accuracy. This method provides an unbiased estimate of model generalizability and is suitable for our dataset with a limited sample size.

Results

On average, 80.6% (SD = 17.1%) of participants' responses (Jar I/II) were directionally aligned with the NB model, which we took as the reference standard. The gray bars in Figure 3 show participants' judgments across conditions. Sixty-seven% of participants rated the diverse agent pairs as more informative, 18% rated them as equally informative and 15% as less informative.

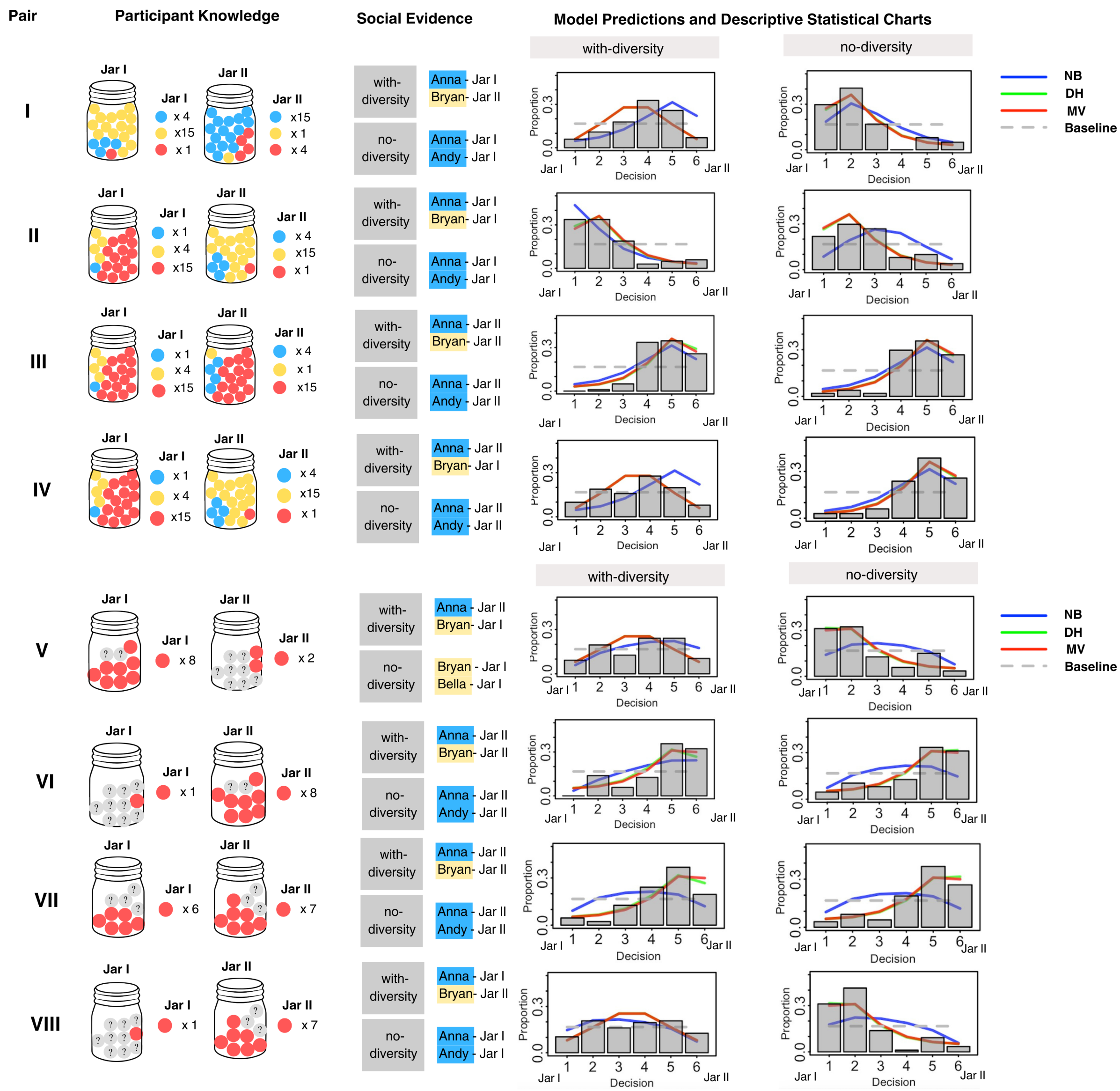

Figure 3: Model Predictions and Descriptive Statistical Charts of Participants' Judgments. Pairs I, II, III, and IV are from Experiment 1, while the remaining pairs are from Experiment 2.

Table 1: Model Fitting Results for Experiment 1. Models are fitted to individual judgments using log-likelihood. McFadden's pseudo-R² is also reported for the best fit participants.

Model	c	β	n	η	BIC	N fit	Rsq
NB	2.04	7.07			2623	24	0.18
MV	2.38	8.21	1.97		2523	29	0.23
DH	2.36	8.25	1.88	1.15	2529	33	0.22
Baseline					2895	15	

Table 1 shows the model fitting results. In the DH model, the fitted η is larger than 1, indicating that participants, overall, made directionally stronger judgments in conditions when agents with different specialisms agreed. However, MV had the best overall fit, with BIC=2523. To better understand individual-level behavior, we also computed the number of participants best predicted by each model, using log-likelihood for comparison. 29% of the participants are best predicted by MV, while 33% align with DH. Meanwhile, a notable 24% of participants are best predicted by NB.

In summary, Experiment 1 shows when combining opinions from social agents with different knowledge, most participants relied on heuristic strategies, although a nontrivial subset made judgments that reflected a sophisticated mentalizing-based Bayesian inference. In Experiment 2, we introduced a more complex scenario in which participants do not have full access to specialist knowledge. This more closely mirrors a real-life setting where people usually have limited knowledge of what others know. Our goal is to test whether participants maintain the same pattern as the cognitive demands of applying ToM increase.

Experiment 2

Participants

We recruited 99 U.S. or U.K. residents on Prolific. We excluded 12 participants with missing values. The final sample included 87 participants (41 female, 45 male, 1 nonbinary), aged 38.56 ± 12.63 years. Participants were compensated £1.50, with an additional bonus (0 - £0.8). The experiment was pre-registered here.

Design and Material

The design of Experiment 2 was similar to Experiment 1. We again designed 4 pairs of trials to distinguish predictions of normative and heuristic models (see Figure 3, and task is available here.

Results

On average, 73.7% (SD = 16.4%) of participants' responses directionally matched the predictions of the NB model. Figure 3 presents descriptive statistical charts of participants' judgments across different conditions. Fifty-four participants rated diverse pairs of agents as more informative than non-diverse, 17 rated them as equal, and 16 as less informative.

The results are shown in Table 2. The estimate of $\eta < 1$ indicates that, in Experiment 2, participants treated agreement between agents with different knowledge as less informative than agreement between agents with shared knowledge. MV again provides the best overall fit (BIC=2274). At the individual level, MV is the best-fitting model for 29% of participants, while DH provided the best fit for 40%. In comparison, NB is the best-fitting model for 22% of participants.

Table 2: Model Fitting Results for Experiment 2

Model	c	β	n	η	BIC	N fit	Rsq
NB	1.82e-05	3.12e05			2424	19	0.10
MV	2.64	5.73	2.64		2274	25	0.16
DH	2.71	5.66	2.89	0.74	2281	34	0.19
Baseline					2494	9	

General Discussion

In the two experiments, we explored how humans aggregate judgments from specialists with either equivalent or complementary knowledge about a domain. The majority of participants across both experiments behaved most consistently with reliance on heuristics that avoided the need to mentalize. This is not surprising given the high challenge of rationally reverse-engineering the most likely observations that caused peers to provide their responses—especially in Experiment 2.

Meanwhile, in Experiment 2, the normative Bayesian model only explains a modest proportion of the variance in responses (R_{sq} = 0.10). This may be because mentalizing in Experiment 2 was cognitively demanding and not particularly rewarding given the high residual uncertainty. Even if participants attempted to reason carefully about the agent's judgments, they likely relied on idiosyncratic priors and simplifications rather than precise calculations, which introduces variability that our normative model cannot not capture.

In both experiments, most of the participants rated the collaborators with different knowledge as more informative. However, although participants put slightly more weight on the advice of informants with different knowledge in Experiment 1, this tendency was reversed in Experiment 2, where informants of knowledge diversity were given less weight. With hindsight, this contradiction may arise from two competing beliefs: on the one hand, responses from different specialists are expected to be more informative because they provide additional evidence; on the other hand, agreements from same specialists offers stronger assurance of reliability (Alister et al., 2024; Xie & Hayes, 2022). Participants in Experiment 2 had uncertainty about the other agents' private knowledge, making it harder to judge whether they were reliable. This uncertainty may have meant they valued cases where similar agents agreed since this corroborated their reliability.

One limitation of our design was that it only weakly distinguishes the heuristics: predictions from the Majority Vote model differed only a little from the Diversity Heuristic and both fit very similarly to participants'. The Diversity Heuristic's extra nuance was not enough to pay for its additional parameter despite most participants reporting that they found diverse agents more informative. Future work can use studies designed to maximize these heuristics' predictive divergence and vary the normative role diverse specialism across the dimensions of complementarity of evidence and corroboration of reliability.

Despite the limitation, our findings suggest people don't tend to or have limited ability to integrate specialist judgments by considering the partial knowledge of others through mentalizing. In daily life, we frequently interact with peers who have different domains of expertise and are privy to different private information. This can easily lead to suboptimal or biased information aggregation, as people may overestimate or underestimate the contributions of informants and draw the wrong inferences from their agreements and disagreements. For example, when people read diverse and inconsistent claims about high-stakes topics like vaccination decisions, they may struggle to assess how to aggregate and integrate these perspectives. Moreover, from an organizational perspective, though including knowledge diversity can promote comprehensive views and innovative ideas (Gray et al., 2012), our results also point out the challenges stakeholders face when managing and integrating judgments from experts with differing perspectives.

References

Alister, M., Ransom, K., Desai, S. C., Soh, E. V., Hayes, B., & Perfors, A. (2024). Sensitivity to Online Consensus Effects Within Individuals and Claim Types. , *46*, 445–451.

Chomeya, R. (2010). Quality of psychology test between likert scale 5 and 6 points. *Journal of Social Sciences*, *6*, 399-403. doi: 10.3844/jssp.2010.399.403

Fränken, J.-P., & Pilditch, T. (2021). Cascades across networks are sufficient for the formation of echo chambers: An agent-based model. *Journal of Artificial Societies and Social Simulation*, *24*(3), 1. doi: 10.18564/jasss.4566

Fränken, J.-P., Valentin, S., Lucas, C. G., & Bramley, N. R. (2024). Naïve information aggregation in human social learning. *Cognition*, *242*, 105633. doi: 10.1016/j.cognition.2023.105633

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology*, *38*(2), 113–154. doi: 10.1348/014466599162700

Gray, S., Chan, A., Clark, D., & Jordan, R. (2012). Modeling the integration of stakeholder knowledge in social–ecological decision-making: Benefits and limitations to knowledge diversity. *Ecological Modelling*, *229*, 88–96. doi: 10.1016/j.ecolmodel.2011.09.011

Hawthorne-Madell, D., & Goodman, N. D. (2019). Reasoning about social sources to learn from actions and outcomes. *Decision*, *6*(1), 17–60. doi: 10.1037/dec0000088

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*, *8*(3), 263–283. doi: 10.1080/15248370701446392

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. doi: 10.1111/j.1467-8624.2005.00849.x

Robinson, E. J., Butterfill, S. A., & Nurmsoo, E. (2011). Gaining knowledge via other minds: Children's flexible trust in others as sources of information. *British Journal of Developmental Psychology*, *29*(4), 961–980. doi: 10.1111/j.2044-835X.2011.02036.x

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*(4), 1054–1070. doi: 10.1111/1467-8624.00334

Wellman, H. M., Carey, S., Gleitman, L., Newport, E. L., & Spelke, E. S. (1990). *The Child's Theory of Mind*. The MIT Press. doi: 10.7551/mitpress/1811.001.0001

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive Science*, *42*(1), 168–187. doi: 10.1111/cogs.12485

Xie, B., & Hayes, B. (2022). Sensitivity to evidential dependencies in judgments under uncertainty. *Cognitive Science*, *46*(5), e13144. doi: https://doi.org/10.1111/cogs.13144

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13. doi: 10.1016/j.obhdp.2003.08.002