Title

Ad Hoc Theories: How Social Interaction Helps Us Make Sense of the World

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Jonusaite, Izabele
Perez, Karla E
Siegel, Max
et al.

Publication Date
2024

Peer reviewed

Ad Hoc Theories: How Social Interaction Helps Us Make Sense of the World

Izabelė Jonušaitė (izabelej@mit.edu) **Karla E. Perez (perezke@mit.edu)**
Max Siegel (maxs@mit.edu) **Laura E. Schulz (lschulz@mit.edu)**
Department of Brain and Cognitive Sciences, MIT

Abstract

In three experiments, we investigated the effect of repeated exposure and social interaction on adults' tendency to make sense of novel events. Specifically, we examined whether, across trials, participants' observations shifted from descriptive to explanatory, from specific to generic, became more inclined to reference causes, and more evaluative. We found that while there was an effect of repeated exposure on generalization and of social interaction on both explanation and generalization, the intervention that was most likely to shift adults' sense-making behavior was a communicative context of small groups in which each participant had partial and different knowledge. We suggest that this is because social contexts inherently motivate individuals to integrate new information, reconcile discrepancies, and forge efficient, generalizable concepts.

Keywords: sense-making; intuitive theories; ad hoc theories

Introduction

Most research in cognitive science has focused on learning as a process primarily driven by the pursuit of accuracy in representation and reasoning (Fiser, Berkes, Orbán, & Lengyel, 2010; Gershman, 2015; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Sarnecka & Carey, 2008; Spelke & Kinzler, 2009; Xu & Kushnir, 2013). However, when interpreting complex and ambiguous scenes, people craft narratives that make sense of the data in a way that is not explained solely by a drive for accuracy: They impose structures that facilitate generalization beyond the immediate data, identify more and less important elements, hypothesize latent constructs like causal relationships, and construct coherent explanations (Gopnik & Wellman, 2012a; L. Schulz, 2012; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Moreover, the structures of ad hoc theories that individuals develop often vary from person to person and do not necessarily converge on a ground truth (Hirstein, 2005). Yet, these cognitive structures provide a framework through which people can decide what questions to ask, what interventions to perform, and what might count as relevant data in ways that might eventually lead to accurate understanding. In this paper, we consider what it might mean for adults to "make sense" of novel events and we present three experiments that investigate two factors that might lead people to move beyond mere descriptions of their observations towards representations that are more abstract and explanatory.

A caveat before we begin: The ideas we propose here are an attempt to make sense of how humans make sense of the world in ad hoc ways—before we have the kind of rich, structured knowledge that supports intuitive theories and deeper understanding. We propose several features of sense-making, and several factors that might contribute to it. This is a preliminary proposal and neither the list of features or factors is intended to be exhaustive. Instead, our aim here is pragmatic: We propose ideas that we believe we could test and either confirm or falsify. What follows is our preliminary proposal and three experiments that investigate some of the factors that might contribute to humans' ability to develop ad hoc theories.

What do we mean by making sense?

Consider a popular board game: Dixit. In the game, players select cards with rich, evocative but ambiguous drawings. The person initiating each round chooses one of their cards and gives a phrase or brief account intended to capture at least some aspect of what they see. The other players then select a card from their own hands that, in their judgment, is also a good fit with the starting player's depiction. All players lay their chosen cards face down in the middle, the cards are shuffled and then turned over. All players then cast their vote for the card that they believe launched the round; that is, the card that best corresponds to the initial player's depiction.

There are many interesting features of this game, but of most interest here is the fact that people are able to play it. Starting from nothing more than an evocative scene, players are able to generate rich, informative accounts abstract enough that other players can select a different scene from their own hand that could fall within the same, conceptual space. Moreover, people create this rich representation on the fly; the representation has sufficient structure for other people to use it to account for their own observations, and people are able to do this even though there is no fact of the matter or ground truth to appeal to.

Dixit is a game. However, we suggest that it exemplifies a critical but relatively neglected aspect of human cognition: the cognitive process by which individuals actively construct meaning from their surroundings and develop ad hoc theories: narratives and frameworks that render complex and ambiguous data comprehensible. Studies of inference, reasoning, and learning in cognitive science have focused primarily on how we go from observations of specific entities and surface features to intuitive theories that enable prediction, plan-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ning, intervention, and converging explanations (Gopnik & Wellman, 2012b; Lombrozo, 2006; Tenenbaum et al., 2011). Here we suggest that there is a critical intermediate stage that involves setting up the kind of cognitive structure that allows us to establish what kinds of hypotheses we might entertain and decide what might count as relevant evidence in the first place. We refer to this intermediate stage as sense-making and we suggest that it is characterized by several features that also characterize intuitive theories but in the case of ideas that merely "make sense" and do not yet reflect real understanding, these features largely refer to the internal structure of the cognitive representation rather than to events in the world.

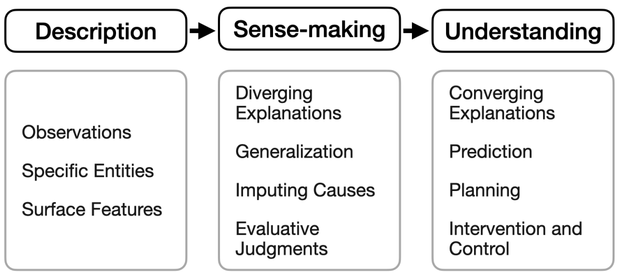

Figure 1: A schema of the relationship and differences between description, sense-making, and understanding. Even though it is represented as a progression, we do not necessarily claim that it is strictly progressive–we might be able to make sense of a phenomenon but not predict or control it.

Features of making sense

Explanation Like intuitive theories, ad hoc theories are explanatory but not in the sense that they provide the kind of information likely to support prediction or intervention (Lombrozo & Carey, 2006). When we make sense of things, we try to provide explanatory links from some aspects of our observations to others such that the whole is greater than the sum of the parts. However, the explanatory links in our ad hoc theories are internal to our representations; they connect some facts to others and pick out patterns but they may be largely subsumptive or formal explanations ("it picked up the eggs because it is an alien"). Such explanations are relatively free of content and may not, yet, support learning or control (Gelman, Cimpian, & Roberts, 2018; Lombrozo, 2006; Williams & Lombrozo, 2010).

Generalization Making sense also involves generalization. Ad hoc theories invoke kinds and relations, not just specific entities and concrete events. Again however, in contrast to intuitive theories which involve sustained commitments to ontological kinds with distinct properties and relations (Carey & Spelke, 1996; Gelman, 2003; Wellman & Gelman, 1992) the kinds we posit when we are in the process of making sense of things may be minimal groupings, with almost no extension beyond the observations themselves; they are likely to be provisional and support only minimal inductive inferences.

Inferring causes When we try to make sense of events, we also invoke causes, both physical causes and psychological ones (reasons, motivations, etc.) In our intuitive theories (let alone our scientific theories), we often have substantial evidence for the causes we invoke, and we can use our knowledge of the causal structure of events both to generate and prevent outcomes (Gopnik et al., 2004; Pearl, 2009). When we are merely making sense of things however, we may invoke causes primarily because we are inclined to believe in a deterministic universe in which all events have causes and causes reliably generate their outcomes (L. E. Schulz & Sommerville, 2006). Like formal explanations, such causes are effectively "placeholder" causes (Cimpian & Salomon, 2014). They may support inferences that do not rely on evidence external to the representation (e.g., counterfactual reasoning) but that are unlikely to support prediction, planning, or intervention.

Evaluative claims We also suggest that when we try to make sense of events we may often impose evaluative judgments on our observations, deciding what is good or bad, easy or difficult, desirable or upsetting. Doing so, we believe, is a preliminary step en route to the grounded understanding of costs and rewards that obtain in environments for which we can more accurately compute the expected utilities of our own and others' actions.

Factors that lead to making sense

We believe that many factors lead us to move from mere redescriptions of the data towards making sense of our observations, including repeated exposure to evidence, social interaction, and pedagogical interaction including instruction or testimony and pedagogical problems or queries. Here we focus on just the first two of these factors.

Repeated exposure Human minds have limited working memory; we are unable to hold onto very many isolated, discrete observations. Thus, we may naturally try to compress observations into more general, abstract, connected representations because these may be easier both to store and access; in this sense, making sense is, if not an optimal use of our limited computational resources (Lieder & Griffiths, 2020), at least perhaps en route to such representations. There has been very little previous work on sense-making, but what literature exists has suggested that compressed, efficient representations are a main feature of this process (Chater & Loewenstein, 2016).

To the degree that this is the case, mere repeated exposure to events may suffice for us to begin making sense of them: that is, to move from the descriptive to the explanatory, the specific to the generic, the associative to the causal, and the neutral to the evaluative. Repeated exposure provides both additional time to process observations and an opportunity to see the evidence again and re-evaluate one's interpretation. We test the effect of repeated exposure on participants' tendency to make sense of events in Experiment 1.

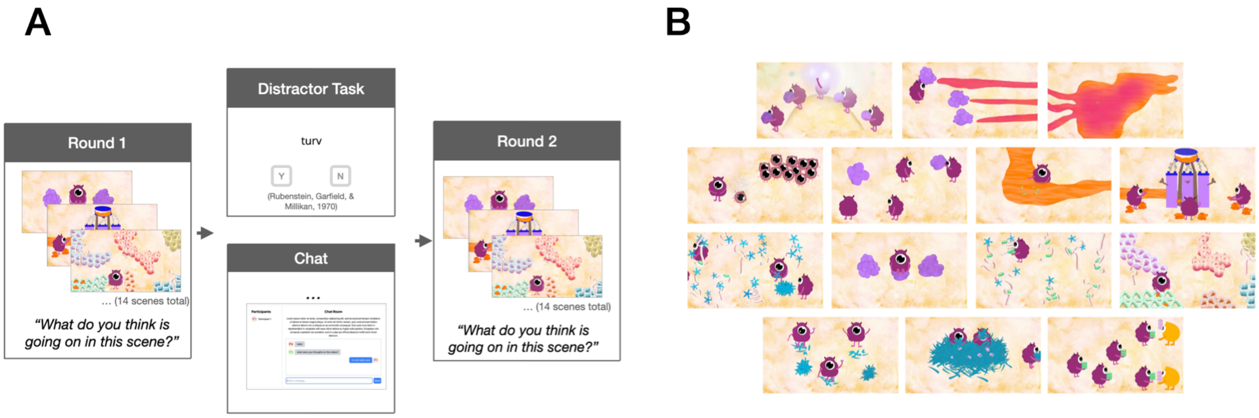

Figure 2: (A) Experimental design for Experiments 1, 2, and 3 (B) Stimuli

Social interaction However, our primary interest here is in the effect of social interaction on sense-making. Many researchers have noted the distinctive aspects of reasoning that emerge in social contexts. Thus for instance, Lombrozo and Carey (2006) have suggested that the function of explanation is to transmit useful information ("Explanation for export") and Mercier and Sperber (2017) have proposed that one of the primary purposes of reasoning is to persuade others ("Reasoning for argument") (Lombrozo, 2006; Lombrozo & Liquin, 2023; Mercier & Sperber, 2011, 2017).

Social interaction has a number of features that might change the structure of knowledge. In any sufficiently rich environment, people are likely to have only partial knowledge and different people will have access to different information. Interacting with others allows access to information that can only be learned from others' testimony (Harris, 2002). The attempt to integrate new information with one's prior beliefs will often lead to belief updating (Anderson, 1990; Oaksford & Chater, 1994; Shepard, 1987; Tenenbaum & Griffiths, 2001).

Moreover, even when people observe the same events, they are likely to interpret it in different ways. Interacting with others will also expose discrepancies in how people interpret shared observations. These inconsistencies are likely to lead people to search for explanations (Legare, 2012; Legare, Gelman, & Wellman, 2010).

As noted, compression is likely to be an important feature of sense-making (Chater & Loewenstein, 2016). Many researchers have proposed that language is optimized for communicative efficiency; thus as people interact and share information, they are likely to develop more integrated, compressed, efficient, and re-usable representations (Gibson et al., 2019; Piantadosi, Tily, & Gibson, 2011).

Finally, in the course of communication and social interaction, people are likely to disagree with each other about aspects of events, especially those that are potentially matters of opinion. This kind of interaction may lead people to increasingly impute values or valence to the contexts.

In short, we believe that, even more than mere repeated exposure, social interaction is likely to shift the structure of human knowledge from observation to explanation, specific to the generic, associative to causal, and neutral to value-laden. While previous work has examined how social interaction fosters these shifts in isolation, here we want to explore how these structures emerge collectively and the role that they play in sense-making. We investigate this effect of social interaction on people's tendency to make sense of events in Experiments 2 and 3.

To assess people's ability to develop ad hoc theories that make sense of novel information, we needed stimuli that fulfilled a number of desiderata: 1) They had to be sufficiently novel and complex that people could not account for the events simply by importing existing knowledge; 2) they had to be sufficiently ambiguous that people had the opportunity to make sense of their observations in different ways; 3) They have to be rich with respect to 'surface content' so observers who chose to focus only on description might do so; 4) There had to be at least a possibility of making sense of the data. That is, there had to be at least one way in which the content could be integrated and explained.

To achieve all these goals, we created a novel alien world, consisting of fourteen looping animated movies, each depicting elements of the aliens' activities. We ensured that the world was, in principle, interpretable by basing the alien activities on a real-world activity of Earthlings. We will not here reveal what that activity actually was for two reasons. First, it would deprive the reader of a chance to try to make sense of the events on their own. Second, and more importantly, our goal here was not to see if observers would discover the events these were based on; nor was it to see if observers would converge on the ground truth. To our minds, one of the most salient features of sense-making, and one of the respects that most clearly distinguishes it from intuitive theories, is that the ways each of us first makes sense of novel observations may be quite diverse. Although we believe sense-making as a whole is characterized by common explanatory, generalizable, causal, and evaluative claims, people may make sense of the same observations in very different ways.

In all three experiments, participants first got a chance to observe all (Experiments 1 and 2) or some (Experiment 3) scenes in the novel world and give an account of what they saw. Then, they either participated in a brief distractor task (so that people could not simply remember verbatim their initial reports) and they got a second chance to observe the scenes and make a fresh report (Experiment 1) or they participated in a text-based chat with three other participants (all strangers to one another) about the events and then returned to the scenes and were given a chance to report on them again. We compared people's first and second accounts within-participants in each experiment and between participants across experiments.

Experiment 1: Repeated Exposure

In this experiment, we tested the hypothesis that merely having a chance to revisit their initial accounts and see the scenes again would lead adults to make more sense of the events, in the form of exhibiting more sense-making attributes in their accounts on their second exposure.

Methods

All experiments were approved under an existing IRB. All participants provided informed consent.

Participants For this study, we recruited 40 participants. In all experiments, participants were recruited online (Peer, Brandimarte, Samat, & Acquisti, 2017) via the online participant recruitment service Prolific (`https://www.prolific.co`). Participants were compensated at the rate of 15 USD per hour. The inclusion criteria were as follows: participants had to be 18 years old or older, residing in the United States, with a Prolific approval rate at or above 95%. Participants who took part in relevant pilot studies or other studies in this series of experiments could not participate in a given study.

Materials The stimuli for this experiment were carefully designed to challenge participants' ability to develop ad hoc theories by presenting them with novel and complex scenarios that they would have to make sense of. A series of 14 cartoon-like videos served as the primary materials, depicting a unique creature world where characters engaged in various activities with ambiguous objects. These objects transformed across scenes, indirectly linking the activities and making the integration of scene-by-scene information possible but not necessary or obvious.

Procedure In the first part of the study, participants saw a total of 14 scenes and wrote their accounts of what they think is going on in each scene. At a time, a participant saw one scene, the question "What do you think is going on in this scene?", and a text entry box. There was no time limit. Part 2 consisted of a distractor task–the word-non-word task. In Part 3, participants went through the same scenes as in Part 1 and gave their accounts of each scene.

Coding We collected participant explanations in written natural language format. For each of the three experiments, we recruited 80 coders on Prolific who rated the participant-generated scene accounts on a range of sense-making characteristics: how descriptive/explanatory they were, whether they involved more generic or more specific language, whether they invoked psychological or physical causes, and whether they included evaluative language. Coders were shown one account at a time (in text form) and selected from two options in a 2AFC format. We followed this coding procedure in all of our experiments.

Results

For each coded question, we specified a Generalized Linear Mixed Model (GLMM) with a logit link function to model the probability of a sense-making characteristic being present in a participant's response as a function of exposure round (first vs. second). The model included random intercepts for both coders and stimuli to control for individual differences in rating tendencies and variability in the sense-making demand of different stimuli.

The analysis of the second coding question on whether the account is more generic or specific yielded significant results. Specifically, the effect of repeated exposure was significant, with an estimate of 0.48 (SE = 0.11), $z = 4.35$, $p < .0001$. The proportion of responses classified as generic increased from 36% in Round 1 to 46% in Round 2. This result suggests that at least one aspect of sense-making—a tendency to shift accounts from more specific to more generic—can be influenced merely by repeated exposure and a second chance to respond to observations.

Analyses of the first, third, and fourth coding questions did not yield significant results, suggesting that repeated exposure in itself has a relatively small effect on people's tendency to make sense of their observations. These results also suggest however, that sense-making is not a rapid or automatic process. Most of the characteristics of sense-making are not triggered by mere repeated exposure.

Discussion

We found some minimal effects within-participants of repeated exposure, suggesting that indeed that repeated exposure and perhaps the increased thinking time does shift people towards more sense-making.

Experiment 2: Live Chat

In this experiment, we look into the effects of social interaction. Specifically, we are interested in whether having an opportunity to discuss what you saw with others who saw the same thing might contribute to sense-making.

Methods

The aim of this experiment was to investigate the effect of social interaction on sense-making. Our hypothesis here was two-fold: Firstly, we hypothesized that participants in the chat condition will, on average, exhibit more sense-making characteristics in their post-chat explanations, compared to pre-chat. Secondly, we hypothesized that participants in the chat condition will, in aggregate, exhibit more sense-making characteristics in their second exposure explanations than participants in the repeated exposure condition.

Participants We recruited 40 participants on Prolific, following the same recruitment process as in Experiment 1.

Materials and Procedure Participants saw the same stimuli as in Experiment 1. Part 1 and Part 3 of this study were identical to Part 1 and 3 of Experiment 1. In Part 2, participants in the 'Chat' condition formed groups of 4 and joined a live chat room with their randomly assigned group. Participants were given these instructions: "Now you will have a chance to message and discuss with 3 other people who saw the same scenes you did", along with technical instructions about the chat. The minimum chat time was 5 minutes, the maximum was 20 minutes.

Results

We used a GLMM with a logit link function to examine the within-subject influence of social interaction through live chat on participants' engagement in sense-making. The response variable in each of our model was the presence of either one of the four sense-making characteristics in participants' explanations. The main predictor in our analysis was round number, contrasting responses before (Round 1) and after (Round 2) the live chat intervention. We included random intercepts for both coders and stimuli.

Our analyses showed that the live chat intervention significantly encouraged participants to provide more explanatory rather than descriptive responses on their second exposure ($z = 3.26$, $p = .001$). The proportion of explanatory responses increased from 36% in Round 1 to 44% in Round 2. Additionally, we observed a statistically significant trend towards the use of more general language ($z = 3.10$, $p = .002$) after the live chat, with general language usage rising from 37% to 44% between rounds.

Comparing the effects of the live chat condition to the repeated exposure condition from Experiment 1 in a between-subjects analysis, we found no significant differences in the development of sense-making characteristics.

Discussion

These results suggest that while live chat may influence sense-making processes within subjects, its aggregate impact compared to repeated exposure alone is not significantly different in enhancing sense-making characteristics as measured in this study.

Experiment 3: Live Chat with Partial Information

We would expect the most powerful effects of social interaction come from the actual complexity of the real world, where you have incomplete evidence and you are sharing it with others who do not see the same things you do. It is this effort to try to integrate things that were not directly observed that would be the most likely to contribute to sense-making. Indeed, if you have first-hand experience with all the evidence yourself, there isn't a great deal of need to consider a result of any conflict with the interpretation of others.

Methods

Participants We recruited 40 participants on Prolific, following the same recruitment process as in Experiments 1 and 2.

Materials and Procedure Participants saw the same stimuli as in Experiment 1 and 2, however, in this study, instead of seeing 14 scenes in Part 1 and 3, participants saw 4 scenes. We had four sets of scenes, with one scene in each set overlapping with another condition. In all other respects, Part 1 and 3 were identical to Experiment 1 and 2

In Part 2, participants were randomly assigned into live chat groups of four. They were given the following instructions: "You are about to enter a chat with three other participants. Each of you has observed scenes from the same world, but you've each seen mostly different parts of it. In this chat, your task is to share your interpretations and ideas about this world. Minutes into the chat, a question will appear. At that point, your group's task is to arrive to a single account on what is happening in this world. Each participant must write this agreed-upon account in the chat before you can proceed. After the chat, you will see and write about the scenes again."

5 minutes into the chat, this system message was displayed: "In general, what do you think is going on in this world? You do not need to respond to this immediately. You have up to 15 minutes more to discuss your interpretation of what is going on in this world with your peers. When you all agree on a single account, each of you should write it down in this text box. Then each of you individually will have a chance to look at your scenes again and say what you now think is going on."

Results

We applied a GLMM with a logit link function to assess the impact of partial information exchange through live chat on sense-making attributes. Our models treated the round of exposure as a fixed effect and included random intercepts for both coders and individual stimuli.

The intervention led to a statistically significant increase in explanatory over descriptive responses in the second exposure, with an effect size of 0.59 ($SE = 0.12$, $z = 4.98$, $p < .001$). The proportion of explanatory responses increased from 40% in Round 1 to 53% in Round 2. Additionally, there was a notable shift towards the usage of more general language ($z = 2.00$, $p = .046$), with the proportion of general language use rising from 43% to 48% between the rounds. Contrasts in participants' inclination to reference physical or psychological causes, as well as their use of evaluative language, did not exhibit significant changes across rounds, ($z = -0.57$, $p = .570$ for causes; $z = -1.27$, $p = .203$ for evaluative language).

A between-subjects comparison to Experiment 1's repeated exposure condition highlighted the contribution of live chat with partial information to more explanatory accounts ($z = 2.64$, $p = .008$). The use of evaluative language was also significantly different in the between-subject comparison, however, this effect was explained away by a significant difference also found when comparing round 1 (pre-intervention or distractor task) accounts between the repeated exposure and the live chat with partial information conditions. Generic language or reference to causes did not differ in this between-subject comparison.

Discussion

The results of this experiment suggest that live chat where participants had partial, mostly non-overlapping partial information about the same complex fictional world, had the most prominent effect on sense-making, compared to repeated exposure and live chat with full information conditions. Specifically, the effect was expressed in more general and more explanatory accounts after this manipulation.

General Discussion

Across three experiments, we investigated the effect of repeated exposure and social interaction on people's tendency to form ad hoc theories to make sense of novel events. Repeated exposure had a slight effect on people's sense-making, shifting them from more specific observations to more general ones. The effect of social interaction was more powerful. After people communicated with others who had seen the same observations, their responses became not only more general but also more explanatory. However, the most influential manipulation was social interaction when the people participating in the interaction had partial and distinct knowledge of the target world. People's responses were not only more general and explanatory after this manipulation, they were also more general and explanatory than in any other context.

Why was the partial knowledge manipulation the most effective? As we noted in the Introduction, in a rich, complex world, each of us only has limited first-hand knowledge of events and we rely heavily on others' testimony for learning (Harris, 2002). The third experiment was perhaps the best analogue to the context in which social interaction and community usually takes place. When everyone has seen exactly the same evidence, we may have less reason to learn from others and update their beliefs. However, when we have only limited knowledge, the act of trying to integrate our information with others may lead to more generalizable and explanatory representations. Moving forward, the next step in this line of research will be to better isolate these effects to determine whether the observed changes in sense-making were primarily driven by the back-and-forth process of social interaction itself, or by the mere addition of new information. This distinction is crucial for unpacking the specific mechanisms through which social interaction facilitates sense-making.

Conversely, why were none of the manipulations successful at causing people to impute more causes or more evaluative judgments to the scene? One possibility is that this is because people were already attributing causes relatively frequently, even from the first exposure. For the sake of brevity, we focused here on the effect of interventions rather than the distribution of responses but in fact, roughly half of all participants invoked causes in all conditions on all rounds. Thus there may have been relatively little room to impact these responses. The same was not the case for evaluative judgments; these occurred very rarely in any condition on either round (10-15% of all responses). One possibility is that different interventions, including pedagogical interactions, might be more likely to impact this measure. Another possibility is that our stimulus set, while engaging for some aspects of sense-making, did not encourage value or valenced judgments of any kind. In particular, we did not individuate any aliens in our alien world. Insofar as people have more of a stake in individuals than nameless entities, they may have been particularly unlikely to trigger evaluative claims.

This work suggests many future avenues for research. As noted, the current account is a first attempt to discuss humans' capacity for building proto-theories on the fly. We believe these initial representations play a crucial role in human cognition; they have sufficient structure to constrain future search and hypothesis testing but they are sufficiently loose and flexible to allow divergent accounts and a rich space of possibilities. However, as noted, our suggestion of both the features of sense-making and the factors that influence it is pragmatic rather than principled. Future investigations may allow us to more formally consider the progression from observation to understanding by way of trying to make sense of a complex world.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of science*, *63*(4), 515–533.

Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization*, *126*, 137–154.

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), 461–480.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences*,

14(3), 119–130.
Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford Cognitive Development.
Gelman, S. A., Cimpian, A., & Roberts, S. O. (2018). How deep do we dig? formal explanations as placeholders for inherent explanations. *Cognitive psychology*, *106*, 43–59.
Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic bulletin & review*, *22*, 1320–1327.
Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019, May). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. doi: 10.1016/j.tics.2019.02.003
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological review*, *111*(1), 3.
Gopnik, A., & Wellman, H. M. (2012a). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. doi: 10.1037/a0028044
Gopnik, A., & Wellman, H. M. (2012b). Reconstructing constructivism: causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085.
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.
Harris, P. L. (2002). Checking our sources: The origins of trust in testimony. *Studies in History and Philosophy of Science Part A*, *33*(2), 315–333.
Hirstein, W. (2005). *Brain fiction: Self-deception and the riddle of confabulation*. Mit Press.
Legare, C. H. (2012, January). Exploring Explanation: Explaining Inconsistent Evidence Informs Exploratory, Hypothesis-Testing Behavior in Young Children. *Child Development*, *83*(1), 173–185. doi: 10.1111/j.1467-8624.2011.01691.x
Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child development*, *81*(3), 929–944.
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*, e1.
Lombrozo, T. (2006, October). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. doi: 10.1016/j.tics.2006.08.004
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204.
Lombrozo, T., & Liquin, E. G. (2023). Explanation is effective because it is selective. *Current Directions in Psychological Science*, *32*(3), 212–219.
Mercier, H., & Sperber, D. (2011, April). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. doi: 10.1017/S0140525X10000968
Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological review*, *101*(4), 608.
Pearl, J. (2009). *Causality*. Cambridge university press.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of experimental social psychology*, *70*, 153–163.
Piantadosi, S. T., Tily, H., & Gibson, E. (2011, March). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. doi: 10.1073/pnas.1012551108
Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674.
Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in cognitive sciences*, *16*(7), 382–389.
Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child development*, *77*(2), 427–442.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.
Spelke, E. S., & Kinzler, K. D. (2009). Innateness, learning, and rationality. *Child development perspectives*, *3*(2), 96–98.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*(4), 629–640.
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.
Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual review of psychology*, *43*(1), 337–375.
Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive science*, *34*(5), 776–806.
Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, *22*(1), 28–32.