Inferring the truth from deception: What can people learn from helpful and unhelpful information providers?

Manikya Alister (alisterm@student.unimelb.edu.au)
Keith Ransom (keith.ransom@unimelb.edu.au)
Andrew Perfors (andrew.perfors@unimelb.edu.au)
School of Psychological Sciences, University of Melbourne

Abstract

Sampling assumptions — the assumptions people make about how an example of a category or concept has been chosen — help us learn from examples efficiently. One context where sampling assumptions are particularly important is in social contexts, where a learner needs to infer the knowledge and intentions of the information provider and vice-versa. The pedagogical sampling assumptions model describes a Bayesian account of how learners and providers should behave given different assumptions they have about the other (e.g., is the provider trying to deceive or help me? Does the learner trust me?). In this study, we tested how well this model could describe learning behaviour in the rectangle game, where a fictional information provider revealed clues about the structure of a rectangle that the learner (a participant) needed to guess. Participants received clues from either a helpful information provider, a provider who was randomly sampling clues, or one of two kinds of unhelpful providers (who could mislead but could not lie). We found that people learned efficiently and in line with model predictions when the provider was helpful and that this was the case even when no cover story was provided. However, although participants could identify that unhelpful providers were not being helpful, they struggled to learn the strategy those providers were using.

Keywords: sampling assumptions; pedagogical reasoning, generalisation; learning; Bayesian modelling

Introduction

Inductive learning (learning from examples) is an important capacity that allows us to efficiently adapt to and make sense of the world as we experience it. Humans are naturally efficient inductive learners: both children and adults need only a few examples to distinguish between different concepts like *cats* vs *dogs*, which is especially impressive given that much of the data we encounter every day is sparse and ambiguous (Tenenbaum et al., 2006).

One possible reason people are so good at inductive learning is that we make assumptions about the generative processes underlying the examples we see. These assumptions are commonly referred to as *sampling assumptions*. Much of the research investigating sampling assumptions has focused on how people reason from *weak* or *strong* samples. Under weak sampling, examples have been randomly sampled from the world and then labelled, while under strong sampling, they are randomly sampled from the category being learned (Anderson, 1981; Hendrickson et al., 2019; Navarro et al., 2012; Ransom et al., 2021; Tenenbaum & Griffiths, 2001). While this research is useful for beginning to think about what assumptions underlie people's ability to make rich inferences from sparse data, realistic scenarios encompass a range of assumptions that go beyond strong and weak alone.

One important limitation is that strong and weak sampling does not explicitly consider the full social context through which most inductive learning occurs. Although strong sampling can sometimes be approximated as a helpful provider selecting examples from the category, this is only a rough approximation: a *real* helpful provider would sometimes choose examples from outside a category (a *whale* is not a *fish*), or would identify some examples of a category as more useful than others (a *robin* is a better example of a bird than an *emu* is). Moreover, sometimes the person providing examples is wrong: someone might incorrectly say a particular mushroom is not poisonous if they are a murderer or just misinformed.

The reason this social element matters is that almost everything we learn is from other people. As infants and children, we interact with parents and teachers; as adults, we talk to each other. In order to reason in social contexts such as these, people need to make assumptions about the intentions and capabilities of the information providers they are learning from; in turn, the information providers need to make assumptions about the learner. The pedagogical framework of Shafto et al. (2014) captures this sort of recursive reasoning through the formalisation below, where a learner's belief in a particular hypothesis $P_{Learner}(h|x)$ is determined by their priors and assumptions about how the provider generated the data:

$$P_{Learner}(h|x) = \frac{P_{Provider}(x|h) \cdot p(h)}{\sum_{h'} P_{provider}(x|h) \cdot p(h)} \tag{1}$$

The provider is formalised based on their assumptions about the learner, in combination with a parameter α which captures the extent to which they are trying to be helpful:

$$P_{Provider}(x|h) = \frac{P_{Learner}(h|x)^{\alpha}}{\sum_{x} P_{provider}(h|x)^{\alpha}} \tag{2}$$

As this equation makes clear, α captures the extent to which the provider aims to increase the probability that the learner believes the correct hypothesis. If $\alpha > 0$ that means the provider is trying to be helpful, whereas $\alpha < 0$ is deceptive: the provider is trying to *decrease* the probability that the learner arrives at the correct hypothesis (while being unable to lie overtly). When $\alpha = 0$ this is equivalent to weak sampling, as the provider is choosing data completely at random.

The predictions of the pedagogical model for both helpful provider ($\alpha > 0$) and learner were tested by Shafto et al. (2014). They used a task called the "Rectangle Game" similar to that shown in Figure 1. In it, a learner tries to guess the location and shape of a rectangle using clues given by

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Figure 1: **Experiment screenshot.** In the experiment, participants were given clues consisting of green points (within the rectangle) and red points (outside the rectangle). After each additional point, people were asked to indicate what rectangle they thought was correct by clicking and dragging over the grid to change the cell colours to light blue. Selections were required to contain the green points and exclude the red point, but no other constraints were imposed. In this example, the participant is in the HELPFUL COVER STORY condition ("Ellie is trying to help you guess the correct answer"), meaning the points were generated by a model set to $\alpha = 1$. The points on target trials 2 and 8, which were the focus of our analysis, were identical across conditions. The conditions with no cover story were identical, but instead of a reminder of the cover story in the side panel, there was a reminder of the general instructions.

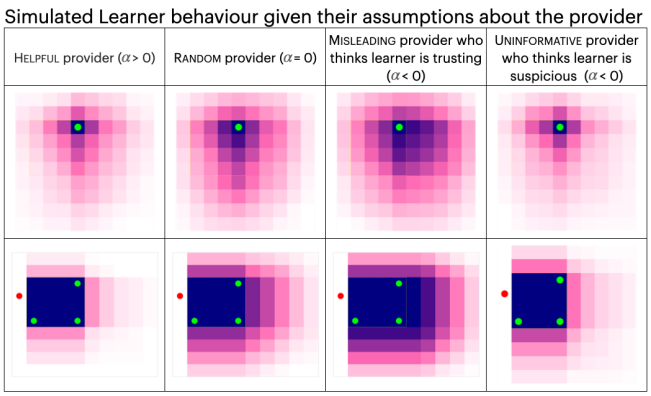

Figure 2: **Model predictions in the rectangle game**. Heatmap of 100 rectangles simulated by the pedagogical sampling model for what a learner would infer given three different assumptions about the provider α (columns) and either a single positive data point (top row) or three positive and one negative data points (bottom row). A learner assuming a *helpful* provider will assume that positive points indicate the corners of the rectangle and will thus infer a tight rectangle around them. A learner who assumes that the provider is *unhelpful*, but knows they cannot explicitly lie, will consider all of the rectangles that are consistent with the data without necessarily favouring tighter ones. An important additional consideration for the learner, however, pertains to what the learner thinks the provider assumes about them. If the provider thinks that the learner is trusting, the learner can infer that the provider believes the learner will draw rectangles around the points tightly, so the learner should actually generalise more loosely. Alternatively, if the provider thinks the learner is suspicious, the teacher might choose points that are closer to the true rectangle, since they think the learner won't generalise tightly anyway. A learner who assumes that clues have been sampled *randomly* won't be more likely to draw any particular rectangle, so long as it is consistent with the points. The model assumes a flat prior over rectangle size.

the provider, who knows the true rectangle and can provide either positive evidence as green dots (points inside the true rectangle) or negative evidence as red dots (points outside it).

Shafto et al. (2014) demonstrated that the pedagogical model accounted for provider and learner behaviour when providers were encouraged to be helpful and the learners knew this, and it has productively been extended to other situations (like Gricean reasoning and pragmatics) where this is a safe assumption (Degen, 2023). However, there are many situations in the real world where providers could be deceptive or simply not put much thought into choosing their examples. And in *those*, different levels of recursive reasoning yield different predictions. For instance, a provider will be unhelpful in different ways depending on what they believe the learner assumes. If they think the learner knows they are unhelpful, the provider might share points that give as little information as possible. Conversely, if they think the learner thinks they are being helpful, the provider might share points that encourage the learner to infer the wrong rectangle.

Are human learners capable of this level of recursive reasoning? Will they, like the model, make looser inferences if they believe the provider is unhelpful? Will that change if they make different assumptions about whether the provider knows that they know? Are they capable of learning from the pattern of data presented if a provider is helpful or not?

Relatively little is known about the answers to these questions. There is some work looking at how people make inferences from unhelpful or deceptive providers, but it is not linked to model predictions and/or does not explore how this changes with additional data, which is where the largest differences are visible (Montague et al., 2011; Ransom et al., 2019). Most other studies have focused on how a provider would choose unhelpful examples (Franke et al., 2020; Ransom et al., 2017), rather than the learner's response. There is no work we are aware of that explores whether people can infer how helpful a provider is based only on the data they provide, in the absence of a cover story making that clear.

The current study addresses these gaps. We asked whether learners made different inferences given providers with different intentions, as predicted by the pedagogical model. Crucially, we tested this by comparing people's performance on target trials in which the same clues were provided across all conditions; this is important given our goal of direct comparison, and contrasts with other work with different goals (Montague et al., 2011; Shafto et al., 2014). We also varied across conditions whether participants were given a cover story or not, which allows us to investigate to what extent they can infer the sampling process on the basis of the data alone.

Experiment 1

Method

Participants 798 participants were recruited from Amazon Mechanical Turk from a pool who had previously passed a qualification task measuring facility with English and ensuring that they were not bots. Ages ranged from 19 to 75 (M = 41, SD = 11), with 45% identifying as female and 83% from the US. They were paid $1.50USD for this 5-10 minute task.

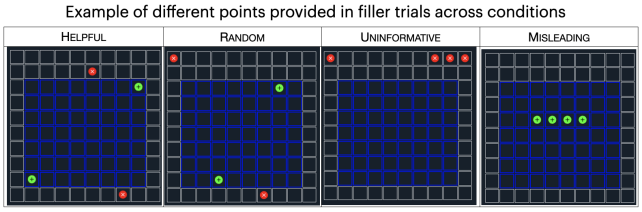

Figure 3: **Example filler trial.** Illustration of the different points provided by providers of different helpfulness in one of the filler trials. The true rectangle is blue, but participants were not shown it, although it was revealed after some of the blocks in Experiment 2.

Design Our experiment followed a 4×4×2 mixed factorial design. The factors were provider HELPFULNESS (Helpful, Random, Uninformative, Misleading), NUMBER OF DATA POINTS (1, 2, 3, 4), and COVER STORY (present, absent). The number of data points was the only within-subjects factor.[1]

Procedure After providing consent and passing a short quiz about the instructions of the task, participants were randomly assigned to one of the eight between-subject conditions.

Participants in the COVER STORY condition read a short cover story about the person who was providing the clues they would use to guess the rectangle. In every cover story, the information provider was a student in our research group called Ellie. For participants in the HELPFUL condition, Ellie was on the same team as they were and would receive a reward that would increase the closer the participant's guess was to the true rectangle. People in the RANDOM condition were told that Ellie was labelling clues (as inside the rectangle or out) that were randomly generated by a computer. Those in the UNINFORMATIVE and MISLEADING conditions were told that Ellie was not on the same team and *could not lie* but would receive a reward that would increase the further their guess was from the true rectangle. In the UNINFORMATIVE condition, people were told that Ellie knew that the learner was aware of her intentions, whereas in the MISLEADING condition, they were told that Ellie thought they were unaware. After the cover story, people answered two questions to check that they understood it. People in the NO COVER STORY condition skipped straight to the main phase.

In the main phase of the experiment, participants completed eight blocks of the rectangle game. Each block contained a different true rectangle that the participants had to guess out of a 10×10 grid. People were shown four clues and after each were asked to indicate what they thought the rectangle was. Clues indicating a point inside the rectangle (positive evidence) were a green plus, while those indicating a point outside it (negative evidence) were a red minus.

Two of the eight blocks were the target blocks, which were the same across all conditions; they appeared on the second and eighth (last) block. For both target blocks the true rectangle was the same size and shape (corresponding to the one shown in Figure 2) but inverted and translated so this was not

[1]The method and analyses for Experiment 1 were preregistered at https://aspredicted.org/i6ci6.pdf.

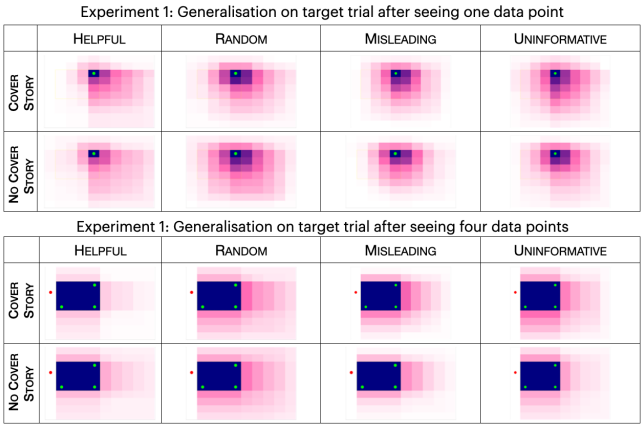

Figure 4: **Experiment 1: Generalisation by condition**. Heat map showing which rectangles participants inferred on the target trial. Probability of each cell in the grid reflects the proportion of people whose rectangle contained that cell (darker = greater probability). People drew slightly tighter rectangles in the HELPFUL conditions, but not as strongly as the pedagogical model predicted.

obvious to participants. The other six blocks were unanalysed filler blocks where the clues were different in different HELPFULNESS conditions (see Figure 3). They were presented to either support the cover story or (when there was no cover story) to explore whether people could make inferences about the provider on the basis of the data alone.

After the fourth block and the final block, participants were given a score reflecting how close their guess was to the true rectangle. This score was intended to motivate them but was deliberately presented only twice so that people could not use it as feedback about any specific guess. After completing the rectangle game, people were asked to indicate how they thought the clues were sampled by selecting either HELPFUL, RANDOM, MISLEADING, or UNINFORMATIVE.

Results and Discussion

The heat map in Figure 4 shows the rectangles that participants inferred[2] in the different conditions of Experiment 1. Although people were somewhat more likely to draw smaller rectangles that fit tightly around the points in the HELPFUL conditions, the differences were not nearly as pronounced as predicted by the pedagogical model in Figure 2.

One explanation for this might be that participants didn't pay attention to the cover stories and drew rectangles that fit tightly around the points because it seemed most obvious. However, this is unlikely for several reasons. Firstly, participants were required to pass a quiz before beginning the experiment that confirmed that they understood the cover story. Also, at the end of the experiment, they were asked to describe how they thought the clues were *actually* sampled, and most people in the cover story conditions correctly inferred whether their provider was being helpful, misleading, or randomly sampling clues (Table 1). Furthermore, when we re-

[2]Because space is limited, all analyses here consider only the second target block; results were qualitatively similar for the first, but noisier because people were still figuring out the task.

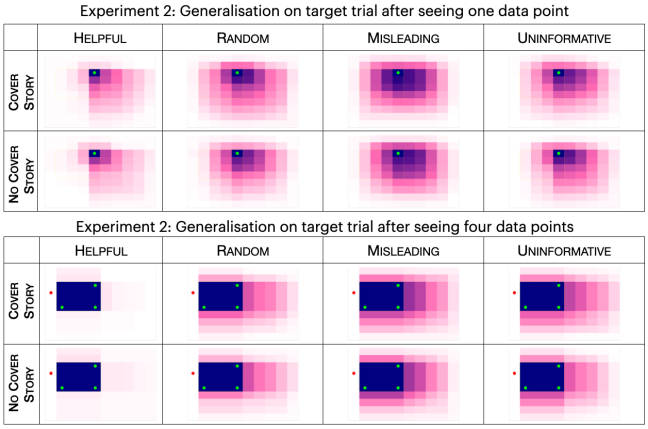

Figure 5: **Experiment 2: Generalisation by condition**. Heat map showing which rectangles participants inferred on the target trial. The probability of each cell in the grid reflects the proportion of people whose rectangle contained that cell (darker = greater probability). Compared to Experiment 1, the rectangles were closer to the predictions of the pedagogical model: more tightly around the points in the HELPFUL conditions and wider in the other conditions.

strict our analyses to include only those people whose answer to the final question matched their condition, the pattern of inference is qualitatively identical to that shown in Figure 4.

Another explanation is that the participants understood the cover stories, but the task was too unconstrained for them to figure out what to do. Since there is only one true rectangle but many different ways to provide misleading or uninformative clues, it makes sense that it would be easier to make inferences based on data from a helpful provider than that of an unhelpful provider. If participants found the task too difficult, they may have just drawn their rectangles tightly around the positive points because they didn't know what else to do.

In the second experiment, we investigated whether people made more appropriate inferences when we didn't just tell them *what* the provider was trying to do (as in the cover story) but also gave them feedback containing information about *how* they were trying to do it (by showing them the true rectangle at the end of some of the filler blocks).

Experiment 2

Method

Participants The recruitment process was identical to Experiment 1, with the additional constraint that people who were in it could not also be in Experiment 2. There were 804 participants of ages from 18 to 79 (M = 41, SD = 12), 48% female, and 81% from the US.

Design and Procedure The design and procedure were identical to Experiment 1, but after completing every odd numbered block, participants were able to see the true rectangle for that block.[3] (Note that this does not include the target blocks, which people got no feedback on). We also removed the scoring function that appeared in Experiment 1.

[3]The method and analyses for Experiment 2 were preregistered at `https://aspredicted.org/xs8c4.pdf`.

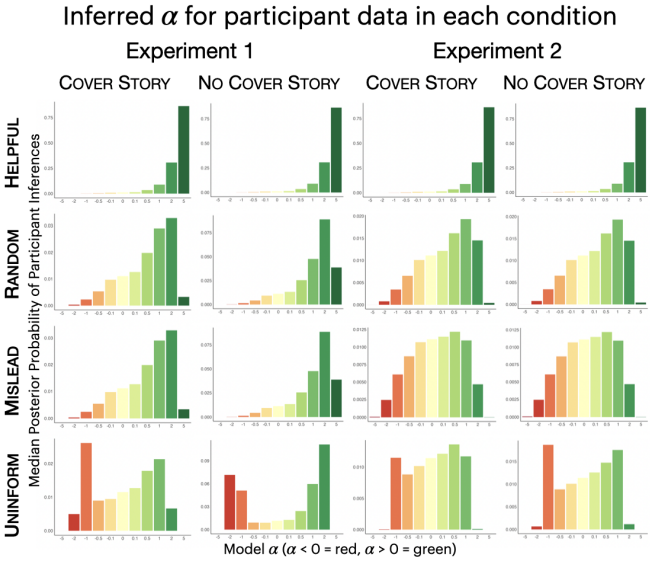

Figure 6: **Probability that participants' inferred rectangles in each condition would have been generated by different α values.** Bars represent the median posterior probability that the pedagogical model set at that α (x axis) would choose the rectangle drawn by participants after all four data points in the second target block. The plot shows that in Experiment 1, even participants not allocated to the helpful conditions behaved in line with a learner who assumes the provider is being helpful ($\alpha > 0$). However, in Experiment 2, participants outside of the helpful conditions most closely resembled a learner who believes clues have been randomly sampled ($\alpha = 0$).

Results and Discussion

As Figure 5 shows, people in all conditions of made inferences much more in line with the predictions of the pedagogical model. Not only did those in the HELPFUL conditions infer even tighter rectangles, but people in the other conditions made wider inferences than they previously had. Further, these differences in the tightness of the rectangles between the HELPFUL and other conditions was maintained even when there was no COVER STORY. That said, even with the benefit of explicit information about the true rectangle, people showed fewer differences between the RANDOM and two unhelpful conditions than the pedagogical model predicts.

This similarity between the non-helpful conditions is especially apparent if we examine the probability that each α parameter would have generated the participants' data in each trial of the second target block (see Figure 6). In this analysis, for each participant response we calculate the α value that best captures the rectangle they drew. What Figure 6 shows is that in Experiment 2, while participants in the HELPFUL conditions resembled the model predictions of learners with a helpful assumption (i.e., $\alpha > 0$), rectangles in the MISLEADING and UNINFORMATIVE conditions (as well as the RANDOM one) did not. Instead, they most closely resembled the kinds of responses that would be expected by a participant who assumed the cues were generated randomly, with inferred α parameters centred around zero.

	Experiment 1		Experiment 2	
Provider	Story	No Story	Story	No Story
HELPFUL	82	46	94	76
RANDOM	64	30	51	30
MISLEADING	83	5	81	3
UNINFORMATIVE	89	47	82	47

Table 1: **Accuracy of inferences about provider.** Numbers show percent of people who correctly assessed the intentions of the provider on the follow-up question at the end of the experiment. We considered people in the MISLEADING and UNINFORMATIVE conditions correct if they chose either option since they correctly identified the provider as being unhelpful, and the definitions of misleading and uninformative in this context are rather ambiguous.

Although this result[4] is a departure from Experiment 1 (where people regardless of condition acted more in line with having a HELPFUL provider), it still suggests that the feedback was only partially helpful. Because of it, participants were able to recognise that the true rectangle was not the tightest one, but they were unable to go beyond that. Indeed, when given the opportunity to tell us the strategy they used, many people in the non-helpful conditions noted that they were unsure how to interpret the clues they were given, so just guessed randomly.

What would an appropriate response look like? One way we can answer this question is by looking at the size of the rectangles participants should (and do) infer in each condition (Figure 7). The model predictions in the HELPFUL condition (green line) show that a Bayesian reasoner should infer the smallest rectangles that are consistent with all of the data; indeed, for the most part participants (green bars) do this. In the RANDOM condition, the model predicts that all rectangles should be equally likely; and although people showed less of a bias towards small ones than they did in the HELPFUL condition, one was still evident. Model predictions in the two unhelpful conditions – especially the UNINFORMATIVE one – were even more divergent from people's performance. The model predicted that a Bayesian reasoner should generally be more likely to select the *largest* rectangles that were consistent with the data since if the provider is trying to be unhelpful, the learner shouldn't believe that the points are informing the outline of the true rectangle. Despite this, most people in the unhelpful conditions did not draw large rectangles.

Interestingly, Table 1 suggests the addition of the feedback in Experiment 2 did not markedly improve the proportion of people whose perception of how the clues were sampled matched the condition they were in.

[4]In Experiment 1, looking at Figure 6, it appears that participants in the UNINFORMATIVE condition may have behaved in line with the model predictions for that condition. However, this is likely due to the fact that in some of the trials within the target block, at some values of α, the behaviour expected in the helpful and uninformative conditions resemble each other (i.e., tighter inferences around the points; see Figure 2). As a result, there is is a reasonably high probability for both some of the negative α values and the positive α values.

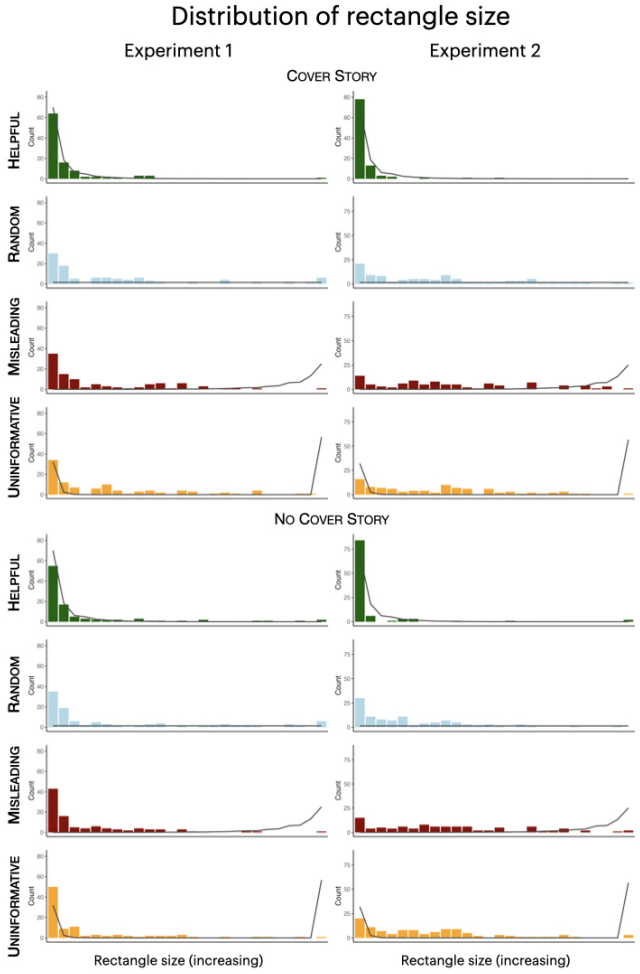

Figure 7: **Distribution of rectangle size after the third clue and model predictions**. We chose the third clue as this had the clearest distinctions between conditions, but note that these predictions vary depending on the clue. The *x* axis shows the size of rectangles drawn ranging from the smallest consistent with the data (leftmost) to the largest (rightmost). The *y* axis reflects the frequency that rectangles at that size were drawn by participants (bars) as well as model predictions (line), normalised to fit on the same scale as the frequencies. They illustrate a different qualitative pattern than the participants show, with the model predicting that in the misleading condition the rectangles should be large, in the RANDOM condition they should be uniform, and in the HELPFUL condition should they be small. In the uninformative condition the model predicts a split between large and small rectangles, with a relatively higher proportion of larger rectangles. Participants preferred small rectangles most of the time, particularly in Experiment 1, and avoided large rectangles, even in the UNINFORMATIVE and MISLEADING conditions.

General Discussion

We found that people tended to be good at inferring the strategies used by the helpful provider, and could even do so without a cover story. These results are consistent with previous research investigating pedagogical sampling assumptions (e.g., Shafto et al., 2014). However, as far as we are aware, our study is the first study that looks at sampling assumptions directly to find that people can infer a provider's sampling strategy without a cover story. This result suggests that people may have been thinking about the generative pro-

cess underlying how the clues were sampled without being prompted, and the helpful clues in the filler trials aligned with their perception of what helpful clues should look like, so they reasoned as if they were learning from a helpful provider. That said, they were clearly better at inferring the strategy when it was helpful, suggesting that in our study, people had a strong prior to expect that an information provider will not be random or unhelpful (although in other contexts, there is evidence to suggest people might have a strong prior for unhelpful sampling, e.g., Alister et al., 2022).

Interestingly, although participants seemed to be able to tell when clues were *not* being sampled helpfully, it was difficult for them to figure out exactly what to do with that information (a similar finding was also found in Ransom et al., 2019). They did not realise, for instance, that larger rectangles were more probable in some of the unhelpful conditions. This is an interesting finding because previous studies have shown that when playing the role of the provider, people tend to mislead others in line with how the pedagogical sampling assumptions model predicts (Ransom et al., 2017). However, our results suggest that it may be much more difficult to infer this strategy when an extra layer of recursion is involved. That is, successful learning requires learners to put themselves in the shoes of the unhelpful provider *while also learning the concept* and that may be what makes this so difficult.

These considerations suggest that learners might be better at inferring the strategies of unhelpful information providers if they were given the opportunity to be the provider before being the learner. This would help them consider the intentions of unhelpful providers, since they would have recent experience trying to do the same thing (i.e., trying to mislead people). If even this experience did not improve their performance, that would suggest that their difficulties arose from something other than a failure to simulate what an unhelpful provider might be doing.

One important consideration is that the rectangle game is quite an abstract task that may not map particularly well onto the kind of inductive learning people are used to in the real world. If the same inductive problems were tested with participants using a more familiar task, it is possible that learners might be better at guessing the strategies used by unhelpful providers. That being said, the rectangle game is quite constrained, more so than the kinds of reasoning scenarios people often encounter. The strategy the learners needed to use in order to align with the pedagogical model in the rectangle game was a fairly straightforward size rule, whereby they should have drawn the smallest rectangle around positive points if they thought the provider was helpful, and the largest possible rectangle around positive points if they thought the provider was being unhelpful. Given the relative simplicity of this strategy, it is interesting that people had so much trouble in the unhelpful conditions, particularly after being able to see some of the true rectangles in Experiment 2.

Our study has implications for understanding inductive generalisation more broadly. In the rectangle game, the rectangles drawn by participants are analogous to a category boundary, and the clues are examples of that category. Although this game is quite different to a typical inductive generalisation task (e.g., it is much more constrained and forces people to conceptualise a category boundary), the size-based generalisation strategy predicted here can easily be extended to more typical inductive problems. A trusting learner should generalise tightly around positive examples of a category, and assume that negative examples are indicative of the category boundary. Conversely, a suspicious learner should generalise broadly from positive points and infer that negative examples are not indicative of the category boundary.

Although there is robust evidence that people follow something like the size principle when generalising from helpful providers (Navarro & Perfors, 2010), much less is known about how people handle unhelpful ones. Our results suggest that people might not be very well-adapted to this situation, particularly in novel contexts where they cannot rely on any domain expertise. This is interesting because typical inductive generalisation tasks either haven't elicited pedagogical assumptions explicitly and only looked at sampling assumptions that do not include negative evidence (e.g., Navarro et al., 2012; Tenenbaum & Griffiths, 2001), or have only looked at helpful providers (e.g., Shafto et al., 2014). Future studies should specifically test whether these findings extend to more typical inductive generalisation tasks.

An important caveat to our conclusions is that in all of our model simulations we assumed flat prior distributions. Instead, it is possible that people people were more likely to believe that certain rectangles were more probable than others before seeing any points. For example, participants may have had a prior belief that small or medium sized rectangles were more likely than rectangles that were the size of only one cell, or the size of the whole grid. At the very least, this likely exaggerated the differences that we could have reasonably expected between different experimental conditions. Because people only saw a small number of data points, any non-uniform prior distribution that was consistent across participants would would have pulled participants' behaviour closer to that distribution. Future research should make an effort to measure participants' prior distributions, for example by asking them do guess the true rectangle before seeing any points (e.g., see Howe et al., 2021).

Acknowledgements

This work was supported by the Australian Defence Science & Technology Group. Alister was supported by an Australian Government Research Training Program Scholarship.

References

Alister, M., Ransom, K. J., & Perfors, A. (2022). Source independence affects argument persuasiveness when the relevance is clear. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). Retrieved February 1, 2023, from https://escholarship.org/uc/item/5hg4p8cm

Anderson, J. R. (Ed.). (1981). Mechanisms of Skill Acquisition and the Law of Practice: Allen Newell and Paul S. Rosenbloom. In *Cognitive Skills and Their Acquisition*. Psychology Press.

Degen, J. (2023). The Rational Speech Act Framework. *Annual Review of Linguistics*, *9*(1), 519–540. https://doi.org/10.1146/annurev-linguistics-031220-010811

Franke, M., Dulcinati, G., & Pouscoulous, N. (2020). Strategies of Deception: Under-Informativity, Uninformativity, and Lies—Misleading With Different Kinds of Implicature. *Topics in Cognitive Science*, *12*(2), 583–607. https://doi.org/10.1111/tops.12456

Hendrickson, A. T., Perfors, A., Navarro, D. J., & Ransom, K. (2019). Sample size, number of categories and sampling assumptions: Exploring some differences between categorization and generalization. *Cognitive Psychology*, *111*, 80–102. https://doi.org/10.1016/j.cogpsych.2019.03.001

Howe, P., Perfors, A., Walker, B., Kashima, Y., & Fay, N. (2021). *Base rate neglect and conservatism in probabilistic reasoning: Insights from eliciting full distributions* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/q7znk

Montague, R., Navarro, D. J., Prefors, A., Warner, R., & Shafto, P. (2011). To catch a liar: The effects of truthful and deceptive testimony on inferential learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33). https://escholarship.org/uc/item/7t67z4kh

Navarro, D. J., Dry, M. J., & Lee, M. D. (2012). Sampling Assumptions in Inductive Generalization. *Cognitive Science*, *36*(2), 187–223. https://doi.org/10.1111/j.1551-6709.2011.01212.x

Navarro, D. J., & Perfors, A. F. (2010). Similarity, feature discovery, and the size principle. *Acta Psychologica*, *133*(3), 256–268. https://doi.org/10.1016/j.actpsy.2009.10.008

Ransom, K., Perfors, A., Hayes, B., & Connor Desai, S. (2021). *What do our sampling assumptions affect: How we encode data or how we reason from it?* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/yqtd9

Ransom, K., Voorspoels, W., Navarro, D., & Perfors, A. (2019). *Where the truth lies: How sampling implications drive deception without lying* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/pv5tk

Ransom, K., Voorspoels, W., Perfors, A., & Navarro, D. J. (2017). A cognitive analysis of deception without lying. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 7.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640. https://doi.org/10.1017/S0140525X01000061

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009