COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–13

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12356

When Absence of Evidence Is Evidence of Absence: Rational Inferences From Absent Data

Anne S. Hsu,[a] Andy Horng,[b] Thomas L. Griffiths,[b] Nick Chater[c]

[a]*School of Electronic Engineering and Computer Science, Queen Mary, University of London*
[b]*Department of Psychology, University of California*
[c]*Warwick Business School, University of Warwick*

Received 8 March 2015; received in revised form 30 October 2015; accepted 7 December 2015

Abstract

Identifying patterns in the world requires noticing not only unusual occurrences, but also unusual absences. We examined how people learn from absences, manipulating the extent to which an absence is expected. People can make two types of inferences from the absence of an event: either the event is possible but has not yet occurred, or the event never occurs. A rational analysis using Bayesian inference predicts that inferences from absent data should depend on how much the absence is expected to occur, with less probable absences being more salient. We tested this prediction in two experiments in which we elicited people's judgments about patterns in the data as a function of absence salience. We found that people were able to decide that absences either were mere coincidences or were indicative of a significant pattern in the data in a manner that was consistent with predictions of a simple Bayesian model.

Keywords: Absent data; Category learning; Rational analysis; Bayesian modeling

1. Introduction

Bayesian inference prescribes how ideal rational agents make inferences. Under the Bayesian framework, inferences update probabilities to hypotheses in light of observed data. If a single hypothesis must be chosen, then a natural criterion is to choose the hypothesis that is most probable in light of the observed data and the prior probability assigned to the hypotheses. Bayesian models have been used to describe how humans make inferences about categories and rules from observed data (Anderson, 1991; Chater & Vitányi, 2002; Feldman, 2000; Goodman, Tenenbaum, Feldman, & Griffiths, 2008;

Correspondence should be sent to Anne S. Hsu, School of Electronic Engineering and Computer Science, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. E-mail: anne.hsu@eecs.qmul.ac.uk

Griffiths, Canini, Sanborn, & Navarro, 2007; Tenenbaum, 2000; Tenenbaum & Griffiths, 2001). In addition to the data that are present, inferences are also made from observations about data that are absent. Absent data could be explained in two different ways: We can assume the absent event will never occur, or we can assume the event is possible but has not yet been observed. A substantial body of previous work has suggested that the Bayesian approach captures the degree to which people are convinced by arguments based on absent evidence, for example, "This drug is safe because no side effects have been observed or Ghosts exist because it has not been proven otherwise" (Corner & Hahn, 2009; Hahn & Oaksford, 2007; Hahn, Oaksford, & Bayindir, 2005; Harris, Corner, & Hahn, 2013; Oaksford & Hahn, 2004).

Our current work seeks to examine how people learn from their own experience of absent data. For example, consider trying to infer whether men in a club wear ties or not for a weekly dinner. A new member, Bob, has attended two dinners so far and did not wear a tie both times. One may infer that "Bob never wears ties." Here, if Bob does sometimes wear ties, one is at risk of what we will refer to as an *undergeneralization.* Undergeneralizations are easily corrected because a single counterexample, that is, seeing Bob without a tie, forces an expansion of the generalization. On the other hand, one could infer, "Bob, like other men, sometimes wears ties and sometimes doesn't." If Bob actually always wore ties, this would be what we refer to as an *overgeneralization.*[1] When working with positive examples only, recovery from overgeneralization requires the prolonged absence of evidence, that is, "Bob is never seen without a tie," to become evidence of total absence.

In the context of learning exceptions from absent data, Bayesian inference supports the intuition that absences of events that are considered more probable should be spotted more easily and be more likely interpreted as exceptions to a general rule. To return to our previous example, suppose ties are worn frequently and attire without ties is only occasionally observed in club members. With finite observations of a man who has only been seen in a tie, it is hard to be certain that he will never appear without a tie (because perhaps it can happen, but is just rare). However, several sightings of Bob without a tie, when ties are very common, may lead one to suspect more strongly that Bob actually never wears ties. Conversely, if ties are not worn most of the time, but are occasionally worn, it would require a very long time (if ever) before one suspected that Bob *never* (rather than merely *rarely*) wears ties. So, in the same vein, noting that you have never seen a particular friend wearing a purple tie, or a diagonally striped tie, does not provide strong evidence that this friend will never wear such ties, because such ties are fairly rarely seen in any case.

The learning of such over- or undergeneralizations from absent evidence has been widely studied in the domain of language learning (Ambridge, 2013; Ambridge, Pine, & Rowland, 2011; Ambridge, Pine, Rowland, & Young, 2008; Baker & McCarthy, 1981; Bowerman, 1988). This is because mastering language requires making generalizations while also learning absence exception to general rules. For example, "The rabbit hid" and "The rabbit disappeared" are both grammatical sentences. However, "I hid the rabbit" is considered grammatical, whereas "I disappeared the rabbit" is not. Much research suggests children learn their first language from positive examples alone (Bowerman, 1988), with little regard to the small amount of corrective negative feedback they receive from carers when they produce ungrammatical sentences. Because children *do* eventually learn these absence exceptions, the question arises of how such absence exceptions are learned without explicit negative feedback. Bayesian models offer a perspective on this problem and have been used to show how cognition-general principles of statistical learning can be used to identify absence exceptions (Dowman, 2000; Hsu & Chater, 2010; Hsu, Chater, & Vitányi, 2011; Kemp, Perfors, & Tenenbaum, 2007; Langley & Stromsten, 2000; Onnis, Roberts, & Chater, 2002; Perfors, Regier, & Tenenbaum, 2006; Perfors, Tenenbaum, & Regier, 2011; Stolcke, 1994). Further experimental work has suggested that absence exceptions in first languages may indeed be learned in a manner corresponding to the predictions of Bayesian rational models (Hsu & Chater, 2010; Hsu et al., 2011; Perfors, Tenenbaum, & Wonnacott, 2010; Wonnacott, Newport, & Tanenhaus, 2008).

Outside of the language domain, there has been little work using learning tasks to examine how people generalize from absences in cognition-general contexts. Here, we present two experiments examining how people learn from absence in a simple cognitive task. By manipulating the probability of an absence, we test whether people's sensitivity to absences is consistent with the predictions of a Bayesian model.

2. Experiment 1: Salience of absence

In Experiment 1, we examined whether participants' confidence judgments concerning conclusions from absent data varied in a way consistent with Bayesian model predictions in a game-like task. People had to sweep "mines" from areas of land—in the critical "absence of evidence" trials, where no mines are encountered, how confident are they that this area of land is completely mine-free? Crucially, how does this confidence depend on the "base rate" at which mines are generally found? According to the Bayesian account, the higher the "base rate" in the surrounding area, the greater confidence people should have that absence of mines implies that an area is "mine-free."

2.1. Method

2.1.1. Participants

Fifty participants over the age of 18 were recruited through Amazon's Mechanical Turk. Only Mechanical Turk users with an approval rating of at least 96% were allowed to participate, and participants were paid \$0.35 to complete the 12-minute experiment.

2.1.2. Stimuli

We created an experimental paradigm akin to the game Minesweeper (Johnson, 1990). A central plot of land (10×10 grid) was fully surrounded by eight equally sized plots of land, which were already "dug up" to reveal the presence of mines of a particular density (see Fig. 1).

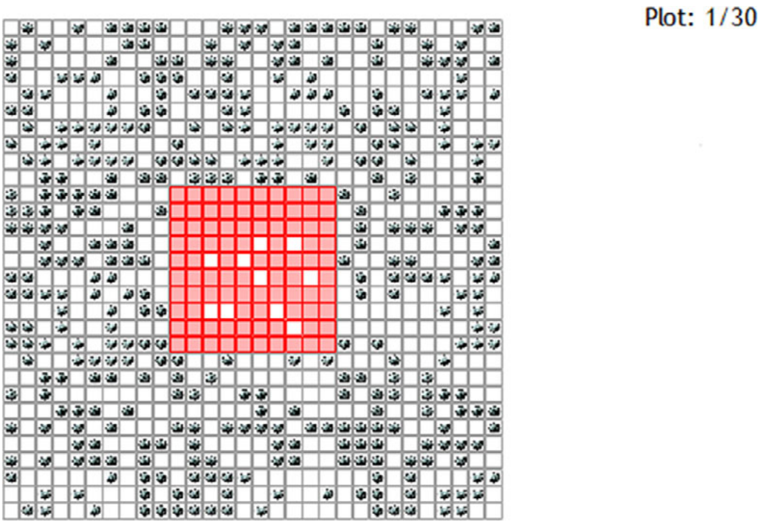

Fig. 1. An individual trial of the kind encountered by participants in Experiment 1. The participant was asked to make 10 “digs” by clicking the mouse over one of the red squares in the central plot. The density of mines in the surrounding area varied with trial number.

2.1.3. Procedure

We conducted a within-participants experiment with 30 trials each. Participants were told they were acting as land inspectors and asked to click on covered plots to make digs. In each trial, participants could make up to 10 “digs” by clicking on one of the 10 × 10 squares in the center plot of land. Participants were told that there were two types of plots, Type 1, where the central plot contained the same density of mines as the surround, and Type 2, where the central plot has been cleared to be completely mine-free. Every click would reveal that square to contain either a mine or no mine. Mine density in the surrounding plots randomly varied across trials, evenly spread between density values of 0.01, 0.1, 0.2, 0.4, and 0.7. Participants were told that in about half the trials, the uncovered plots would contain mines occurring at the same frequency as the surround (Type 1), and in the other half the uncovered plots will be mine-free (Type 2). Participants encountered a new plot in each trial. Once participants encountered a mine, or the limit of 10 digs was reached, we asked participants to indicate on a slider how likely they believed the plot to be Type 1 versus 2, corresponding to left- and right-hand sides of the sliding scale, respectively. Example Type 1 and 2 plots were placed on either side of the scale to ensure it was clear which was Type 1 versus 2. For trials where a mine is encountered, the rational response would be complete certainty of it being a Type 2 plot, which was indeed the response of all participants on these trials.

Three training trials were given prior to the main experiment. In the main experiment, participants encountered 15 “Type 1 trials,” where they encountered a mine before the 10-click limit, randomly mixed with 15 “Type 2 trials,” where 10 clicks were made without encountering a mine. For Type 1 trials, the number of clicks participants made before encountering a mine was sampled from a specialized capped geometric distribution (see Appendix).

2.2. Bayesian model predictions

To determine how far participants responded in line with the Bayesian model in their confidence judgments, participants' confidence in the appearance of a Type 2 plot was compared to the posterior probability of a Type 2 plot[2] given as follows (see Supplementary Materials for full derivation):

$$p(Z = 2 | X_i = k, \rho_i) = \frac{1}{1 + (1 - \rho_i)^k} \tag{1}$$

where X_i is the number of mine-free digs made in the ith plot, ρ_i is the density of mines surrounding the ith plot, Z is the plot type, and the prior probability of $Z = 1$ and $Z = 2$ are assumed to be equal.

2.3. Results and discussion

Responses for relevant trials were grouped by the surrounding mine density and averaged within groups. Participant completion rate was 83% (dropout rate 17%). Degree of confidence in a Type 2 plot was taken to be the proportion to which the slider was placed toward the right-hand side of the slider. A one-way ANOVA on confidence responses, grouped as above, yielded $F(4,740) = 26.8$, MSE = 0.049, $p < .0001$, leading to the conclusion that altering the surrounding mine density had a significant effect on the salience of absences. There was a positive correlation between participants' confidence in a Type 2 plot and the posterior probability given by our Bayesian model, $r(3) = .897$, $p < .05$ (see Fig. 2). Except for the plots where the surrounding density was $\rho_i = 0.01$, participants' judgments tended to fall below the predictions of the Bayesian model, showing a kind of conservatism in judgment that has been found in previous work (Fischhoff & Beyth-Marom, 1983; Philliip & Edwards, 1966).

The fact that participants' average judgments fell above the Bayesian model predictions when the surrounding plot density was very low ($\rho_i = 0.01$) may have been due to a slight ambiguity in our question: Is a mine-free plot interpreted as one with a density of 0 (i.e., mines *cannot* occur) or one in which no mines *happen* to occur by chance (analogous to Bob never actually wearing one of his few ties to the club, despite having no rule against it)? This can plausibly arise in our experiment only for the lowest mine density in the surrounding sarea ρ_i =0.01. If we include within the definition of "Type 2" plots, those plots who have the same density as the surrounding plots but, due to chance, happen to not have any mines in the middle, the model prediction for the probability of being empty in the middle becomes .71 (see Equation 2 below). The average participant response lies in between the model predictions under these two different assumptions.

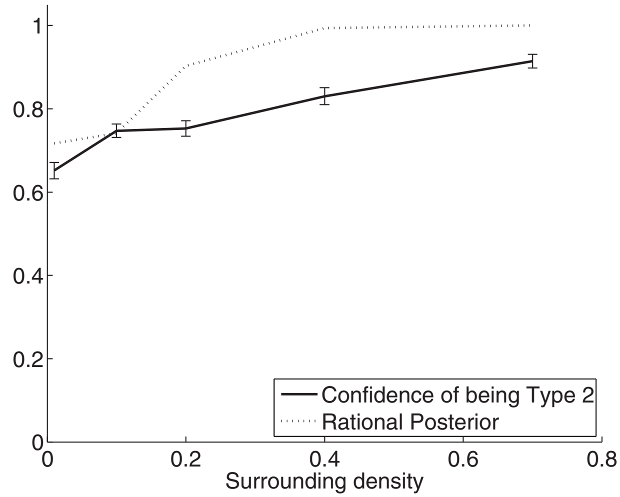

Fig. 2. Participants' report of confidence of being a Type 2 versus Type 1 mine, as a function of surrounding mine density. Error bars show 1 SE. The predictions of a rational model for the probability of a Type 2 plot is also shown.

3. Experiment 2: Inferences from absent data

Experiment 1 suggested that people were sensitive to the salience of an absence when making judgments about the source of their observations, as predicted by a simple Bayesian model. However, there was a slight ambiguity in how participants might have interpreted the phrasing of our question. Thus, in Experiment 2, we asked participants to make the unambiguous judgment of whether there are mines in the central plot. Also, in Experiment 2 we wanted to investigate absence judgments in a more controlled situation compared to Experiment 1, where mine density varied across trials. Hence, we used a design where participants first gained experience making judgments from absence with fixed surrounding mine densities and were only later exposed to plots with a different density. This gave participants more experience in making these types of judgments before the surrounding density changed. Furthermore, to create a stronger test of the Bayesian model, we chose the parameters in Experiment 2 such that the model has a distinctive U-shape prediction—the plot is assumed to be clear when surrounding mine density is either very low (because it is likely that no mine will be present, by chance) or high (so that finding no mines in the central area is highly surprising, unless it is cleared, unlike the surround).

3.1. Method

3.1.1. Participants

In all, 149 participants over the age of 18 were recruited through Amazon's Mechanical Turk. Only Mechanical Turk users with an approval rating of at least 96% were allowed to participate, participants were paid $0.30 to complete the 10-minute experiment.

3.1.2. Stimuli

The animated experimental stimuli were similar to those in Experiment 1.

3.1.3. Procedure

This was a between-participants experiment. As stated earlier, participants, acting as land inspectors, were asked to click on the covered plots to make digs on covered plots of land where there were the two types of plots, Types 1 and 2. They were again told these two types of plots were about equally likely.

Participants were told, for each trial, there would be funds that they could potentially invest in building on the central plot of land and that mine-free center plots would generate returns in proportion to the investment, whereas a center plot with even a single mine present would result in a loss of all the money invested. After each trial, participants were told to indicate what percentage (between 0% and 100%) of total possible funds allowed for that trial would they be willing to spend investing on this center plot for building. Available funds were renewed on each trial, so that up to 100% investment could be made on each trial. The actual amount of funds was left unspecified. Thus, percentage of investment on the trial was now a proxy for a participant's confidence that the central plot was free of mines. Note that there was no longer the ambiguity of whether a Type 2 plot was one that did not have mines in the center due to having a "mine density of 0," in contrast to the surrounding land, or to having the same density as the surround but, by chance, having no mines. Under the new instructions, both of these are valid causes for a mine-free central plot, and the only task is to judge whether the central plot if free of mines.

We employed a between-participants design with five conditions. Participants underwent 22 trials total. So that all participants had a similar initial experience of learning to make judgments from absence, all participants in all conditions experienced 18 trials of reporting investments for land plots with surrounding mine density of $\rho_1, \rho_2, \ldots \rho_{18} = 0.1$, where ρ_i is the density of the surround on the i^{th} trial. We did this to get participants into a mode of steady response for the amount they were willing to invest, and to familiarize them with the concept of investing between 0% and 100% of funds on each trial for building on a plot. Participants received no feedback on the outcome of their investments. To minimize between-participant noise, all participants experienced the same sequence of Type 1 versus 2 plots in the following fixed order: 2,1,1,1,2,2,1,2,1,2,2,2,1,2,1,1,2,2. Also, the number of digs before a mine appeared in trials with Type 1 plots was also determined by a fixed sequence: 6,7,8,9,7,8,6,9. Because Type 1 plots revealed a mine in one of the 10 digs, the sensible investment here would be 0, which was indeed the choice of all participants in these trials. The final two trials 17 and 18 before shifting to the new density, ρ_{19-22}, were Type 2 (with no mines revealed in 10 clicks) because we wanted participants to have their recent betting amounts for $\rho_{1-18} = 0.1$ fresh in their memory. After the first 18 trials, participants answered four final trials, for which the surrounding mine densities, ρ_{19-22}, were fixed at 0.01, 0.03, 0.1, 0.3, and .9, respectively, for the five conditions. For all of these final four trials, participants made 10 digs without encountering a mine and were asked what proportion out of the maximum money allowed for each plot they would invest from 0% to 100%. The outcome measures were the proportion of investments made in these last four trials.

3.2. Bayesian model predictions

The rational Bayesian solution for the probability of a plot with surrounding density ρ_i having No Mines given number of digs, X_i (= 10), were made without turning up a mine is as follows (see Supplementary Materials for full derivation):

$$P_\rho(\text{No Mines}|X_i = 10) = (1-\rho_i)^{90}\frac{(1-\rho_i)^{10}}{1+(1-\rho_i)^{10}} + \frac{1}{1+(1-\rho_i)^{10}} \tag{2}$$

3.3. Results and discussion

Random assignments resulted in the following numbers of participants $N = 31, 30, 27, 30, 31$ assigned to conditions 1–5, respectively. The completion rate was 85% (dropout rate 15%). There was no difference in dropout rates between conditions: A chi-square test for dropouts versus completers between conditions was insignificant, $X^2(4, N = 149) = 0.296$, $p = .99$. Each individual's investments in the last four trials were averaged into a mean investment. A one-way between-subjects ANOVA found significant differences in these average investments as a function of condition, $F(4,144) = 4.6$, MSE = 897.15, $p < .005$. Investment proportion was strongly correlated with the posterior probability of no mines, with $r(3) = 0.95$, $p < .05$ (see Fig. 3). To test for the significance of the U-shaped pattern over the first three low-density conditions ($\rho_i = 0.01$, 0.03, and 0.1), we ran two tests: First, we ran a regression model, regressing the mean scores in these first three conditions for each participant on predictors corresponding to quadratic, linear, and constant components, and found all three coefficients were significant: $t(85) = 2.3$, $p < .05$ for quadratic, $t(85) = -2.3$, $p < .05$ for linear, $t(85) = 4.4$, $p < .0001$ for constant. Second, we also performed two t tests, one contrasting $\rho_i = 0.01$ and $\rho_i = 0.03$, and the other contrasting $\rho_i = 0.03$ and $\rho_i = 0.1$. Significant differences in average investments were found between $\rho_i = 0.01$ ($M = 60.7$, $SD = 38.0$) and $\rho_i = 0.03$ ($M = 39.6$, $SD = 32.9$) conditions, $t(59) = 2.3$, $p < .05$ as well as between $\rho_i = 0.03$ ($M = 39.6$, $SD = 32.9$) and $\rho_i = .1$ ($M = 57.7$, $SD = 25.6$) conditions, $t(55) = 2.3$, $p < .05$. Thus, there was a strong correspondence between people's willingness to reason from absence in order to bet that there were no mines in the plot and the posterior probability of there actually being no mines; moreover, participants' responses followed the distinctive U-shaped pattern of the Bayesian model. While results follow the predictions of the Bayesian model quite tightly, as with Experiment 1, we again found that people were less willing to be confident in absence than a perfect rational model would predict, displaying conservatism.

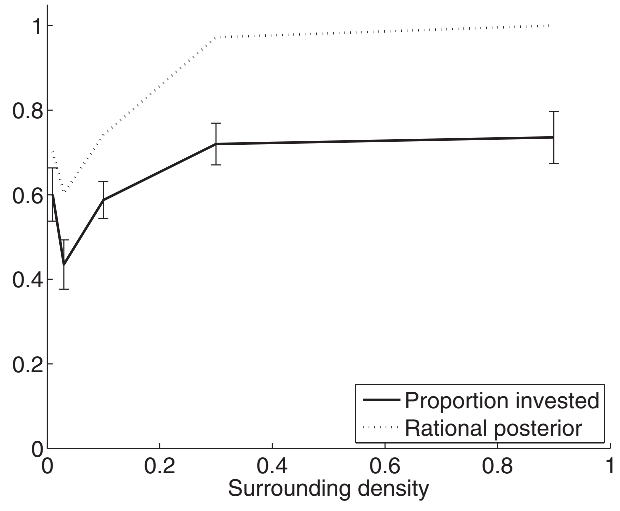

Fig. 3. The proportion of total money participants were willing to invest in building in a central plot of land as a function of surrounding mine density. Error bars show standard errors. The predictions of a rational model for the probability that the center plot has no mines is also shown.

4. General discussion

Being able to make appropriate inferences from the absence of an event is a key part of forming accurate generalizations. Our results suggest that people are able to reason about absent events to form generalizations about the structure of patterns in data according to the predictions of Bayesian inference. In Experiments 1 and 2, participants were more likely to assess the absence of mines as an indication that no mines were present when the base rate for finding mines was higher—the most fundamental prediction of a Bayesian analysis of this problem.

While our results follow the trend predicted by Bayesian models, we did find people's responses, if interpreted as subjective probabilities, to be less confident in absence than a perfect rational model would predict. However, it is unclear to what extent the elicited judgments of confidence (Experiment 1) and proportion of investment (Experiment 2) actually can be interpreted as subjective probabilities defined on a probability scale, and it is possible we cannot expect anything more than ordinal relationships between the judgments we elicited and the subjective probabilities predicted by Bayesian models.

There are a number of ways in which future work could build on this result. First, our simple, controlled experiments inevitably do not capture the reality of exception-learning situations in everyday life. These include demand characteristics, such as the prior probabilities (visually displayed as surrounding mine density), were extremely salient and fixed, and highlighted as the parameter that was changing. The training phase in Experiment 2 focused exclusively on a density of 0.1, which could produce anchoring effects. The impact of all of these factors can be addressed in future work that explores absence learning in a wider range of settings. Also, we provided a clear explanation for why there might be zero land mines, whereas in real life it is rarely so clear that "precisely zero" is a reasonable hypothesis. The question of how people may arrive at the "precisely zero" hypothesis, purely from patterns of observed data, is an interesting question that remains to be explored in future research. Our current work also limited the number of digs to 10, while in reality people may have more choice in how often can they look for absence evidence. Further work could use paradigms with a more flexible amount of allowed enquiries and examine the "stopping criteria" of when people become confident enough in their hypotheses to stop looking for absence. Finally, future work might also examine whether our results regarding learning of absence patterns extend to temporal data and the extent to which this may be similar to such processes in language learning.

Acknowledgments

N.C. was supported by ERC grant 295917-RATIONALITY, the ESRC Network for Integrated Behavioural Science (grant number ES/K002201/1), the Leverhulme Trust (grant number RP2012-V-022), and the Research Councils UK Grant EP/K039830/1. A.H. was supported by ESRC grant for the project "The learnability of linguistic irregularities: A simplicity-based approach" (grant number RES-000-22-2768).

Notes

1. Here, overgeneralization refers to inferring that more possibilities are possible than are actually possible; undergeneralization is the reverse. This usage should not be confused with the use of the terms "generalization" and "overgeneralization" in colloquial language, which refers to all types of incorrect inferences (both over- and undergeneralizations) on too little data. Thus, our term "undergeneralization," for example, "Bob never wears ties" may be commonly referred to, colloquially, as overgeneralization. However, what we call overgeneralization, for example, "Bob sometimes wears ties and sometimes doesn't" may not, in everyday terms, even be considered a generalization at all.
2. Only Type 2 trials are analyzed. In Type 1 trials, where a mine is encountered within 10 clicks, it is immediately clear that the current plot was Type 1.

References

Ambridge, B. (2013). How do children restrict their linguistic generalizations? An (Un-) grammaticality judgment study. *Cognitive Science*, *37*, 508–543.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics*, *22*, 303–323.

Ambridge, B., Pine, J., Rowland, C., & Young, C. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*, 87–129.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Baker, C. L., & McCarthy, J. J. (1981). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford, UK: Blackwell.

Chater, N., & Vitányi, P. (2002). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.

Corner, A., & Hahn, U. (2009). Evaluating science arguments: Evidence, uncertainty, and argument strength. *Journal of Experimental Psychology: Applied*, *15*, 199–212.

Dowman, M. (2000). Addressing the learnability of verb subcategorizations with Bayesian inference. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society* pp. 107–112.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *403*, 630–633.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*, 239–260.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule based concept learning. *Cognitive Science*, *32*, 108–154.

Griffiths, T., Canini, K., Sanborn, A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*, 323–328.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Hahn, U., Oaksford, M., & Bayindir, H. (2005). How convinced should we be by negative evidence. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 887–892.

Harris, A. J., Corner, A., & Hahn, U. (2013). James is polite and punctual (and useless): A Bayesian formalisation of faint praise. *Thinking & Reasoning*, *19*, 414–429.

Hsu, A., & Chater, N. (2010). The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science*, *34*, 972–1016.

Hsu, A., Chater, N., & Vitányi, P. (2011). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition*, *120*, 380–390.

Johnson, C. (1990). Microsoft Minesweeper [Computer software].

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypothesis with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.

Langley, P., & Stromsten, S. (2000). Learning context-free grammars with a simplicity bias. *Proceedings of the Eleventh European Conference on Machine Learning*, 220–228.

Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology*, *58*, 75.

Onnis, L., Roberts, M., & Chater, N. (2002). Simplicity: A cure for overgeneralizations in language acquisition? *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 720–725.

Perfors, A., Regier, T., & Tenenbaum, J. B. (2006). Poverty of the Stimulus? A rational approach. *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, 663–668.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*, 306–338.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*, 607–642.

Philliip, L., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, *72*, 346–354.

Stolcke, A. (1994). *Bayesian learning of probabilistic language models*. Department of Electrical Engineering and Computer Science, University of California Berkeley.

Tenenbaum, J. B. (2000). Rules and similarity in concept learning. *Advances in Neural Information Processing Systems*, *12*, 59–65.

Tenenbaum, J. B., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640.

Wonnacott, E., Newport, E., & Tanenhaus, M. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165–209.

Appendix: Model details

Experiment 1: Geometric distribution

For Type 1 trials, the number of clicks the participant makes before encountering a mine was sampled from a specialized capped geometric distribution, with the probability of success (ρ) equal to the mine frequency of surround plots.

$$\text{Geometric}_{\text{capped}}(\rho, x) = \begin{cases} P(X = x | x < 10) = \text{Geometric}(\rho, x) \\ P(X = 10) = \sum_{x=10}^{\infty} \text{Geometric}(\rho, x) \\ P(X > 10) = 0 \end{cases} \tag{A1}$$

Experiment 1: Model

In order to determine whether participants responded rationally in their confidence judgments in response to the perceptual salience of absences in this minesweeping framework, participants' confidence in the appearance of a Type 2 plot was compared to the posterior probability of a Type 2 plot. This probability was computed via Bayesian inference given the trial-specific density of mines in the surrounding plots that half the plots were mine-free and that participants were only allowed 10 digs per trial. The posterior probability of a Type 2 plot is given by:

$$\begin{aligned} p(Z = 2 | X_{\mathrm{i}} = k, \rho_{\mathrm{i}}) &= \frac{p(X_{\mathrm{i}} = k | Z = 2, \rho_{\mathrm{i}}) p(Z = 2 | \rho_{\mathrm{i}})}{\sum_{z=1,2} p(X_{\mathrm{i}} = k | Z, \rho_{\mathrm{i}}) p(Z | \rho_{\mathrm{i}})} \\ &= \frac{1(0.5)}{0.5 + (1.\rho_{\mathrm{i}})^k (0.5)} \\ &= \frac{1}{1 + (1 + \rho_{\mathrm{i}})^k} \end{aligned} \tag{A2}$$

where X_i is the number of mine-free digs made in the i^{th} plot, ρ_i is the density of mines surrounding the i^{th} plot where Z is the plot type, and the prior probability of $Z = 1$ and $Z = 2$ are assumed to be equal.

Experiment 2: Model

In order to determine whether participants responded rationally, investment proportions were compared to the posterior probability of a plot having no mines, given the density of mines in the surrounding plots that half the plots were mine-free and that participants were only allowed 10 digs per trial. However, calculating this posterior probability is a little more involved than in the previous experiment.

First, we need the probability that it was a Type 1 plot, given that 10 digs were made without turning up a mine and that the prior probability of Type 1 and Type 2 plots was equal. This is the quantity calculated in Experiment 1, and is given in Equation 2 above.

We will abbreviate this probability as $P_\rho(Z = 1|X_i = 10)$.

Next, we have to calculate the probability of no mines at all in a Type 1 plot given that 10 digs have already been made without any mines. This is equivalent to the probability that 90 digs in a row will turn up to have no mines (because there have already been 10 digs without mines, we assume each dig is independent, and there are 100 dig locations total per plot). This is $P_\rho(\textit{No Mine}|Z = 1, X_i = 10) = (1-\rho_i)^{90}$. Finally, we have to calculate the probability of no mines at all in a Type 2 plot. This is equal to one because by definition Type 2 plots have no mines: $P_\rho(\textit{No Mine}|Z = 2, X_i = 10) = 1$.

The final probability of no mines in the center is the combination of the probability of there being a Type 1 plot with the same density as surrounding lands, but by chance has no mines in the center, combined with the probability of it being a Type 2 plot, giving

$$\begin{aligned} P_\rho(\text{No Mines}|X_i = 10) &= P_\rho(\text{No Mines}|Z = 1, X_i = 10)P_\rho(Z = 1|X_i = 10) \\ &\quad + P_\rho(\text{No Mines}|Z = 2, X_i = 10)P_\rho(Z = 2|X_i = 10) \\ &= (1\rho_i)^{90}\frac{(1\rho_i)^{10}}{1 + (1\rho_i)^{10}} + \frac{1}{1 + (1\rho_i)^{10}} \end{aligned}$$