

Basic understanding of posterior probability

Vittorio Girotto* and Stefania Pighin

Center for Experimental Research on Management and Economics, Department of Culture Project, University IUAV of Venice, Venice, Italy

Keywords: posterior probability, updating, number of chances, natural frequencies, intuitive judgments

Consider the following task

[Task A]

A prenatal test determines whether an unborn child has a chromosomal anomaly. A priori, namely, before undergoing the test, a pregnant woman has a 4% chance of having a child with the anomaly. If a woman has a child with the anomaly, there is a 75% chance that she has a positive test result. If she does not have a child with the anomaly, there is still a 12.5% chance that she has a positive test result. Emma, a pregnant woman, undergoes a prenatal test. The result is positive. What is the probability that she has a child with the anomaly?

To answer correctly, one has to integrate the prior probability that a woman has a child with the anomaly (i.e., the prevalence rate: 4%) with information about the test's statistical properties. On the basis of this information and the evidence that Emma tested positive, one can produce a correct posterior evaluation by computing the ratio:

$$\text{Probability (Anomaly|Positive Test Result)} = \frac{\text{Probability ("Positive Test Result and Anomaly")}}{\text{Probability ("Positive Test Result")}}$$

To obtain the numerator, one has to combine the prevalence rate and the test's sensitivity rate (i.e., $4\% \times 75\% = 3\%$). To obtain the denominator, one has to combine the complement of the prevalence rate and the false positive rate (i.e., $96\% \times 12.5\% = 12\%$), and then add it to the initially obtained value (i.e., $3\% + 12\% = 15\%$). Very few respondents, including health-care professionals, produce the correct probability ratio (i.e., $3\%/15\% = 20\%$). Failures to solve tasks of this sort lead to pessimistic conclusions about naive probabilistic reasoning (e.g., Casscells et al., 1978). Subsequent studies, however, licensed more optimistic conclusions, showing that some versions of these tasks led to better performances. About half of the respondents succeed when reasoning with natural frequencies (e.g., "Three out of the 4 women who had a child with the anomaly had a positive test result") or numbers of chances (e.g., "In 3 out of the 4 chances of having a child with the anomaly the test result is positive"; see, respectively, Hoffrage and Gigerenzer, 1998; Girotto and Gonzalez, 2001). On the basis of these results, the current, common account is that posterior probability reasoning improves in versions that allow respondents to both rely on an appropriate representation of subsets of countable elements (e.g., observations, tokens), and to easily associate posterior evidence with one of these subsets (Barbey and Sloman, 2007).

A generally unnoticed aspect of the results mentioned above is that they concern educated respondents, like undergraduates and physicians, and that only about half of these respondents benefit from the simplified versions of the tasks. Even more unnoticed is the fact that respondents sampled from the general public do not benefit at all from these versions. Indeed, in samples of pregnant women, many of whom had a high school level of education or less, almost all respondents failed to compute the correct probability ratio, even if they had to reason about natural frequencies (Bramwell et al., 2006) or numbers of cases (Pighin et al., 2015). In other words, they failed tasks that, in principle, should have activated the appropriate set representation. Their failure is striking because, unlike the participants of previous studies who had to reason about hypothetical

OPEN ACCESS

Edited by:

David R. Mandel,
Defence Research and Development
Canada, Toronto Research Centre,
Canada

Reviewed by:

Ulrich Hoffrage,
University of Lausanne, Switzerland
Stephanie Denison,
University of Waterloo, Canada

*Correspondence:

Vittorio Girotto,
vgirotto@iuav.it

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 10 February 2015

Accepted: 09 May 2015

Published: 22 May 2015

Citation:

Girotto V and Pighin S (2015) Basic
understanding of posterior probability.
Front. Psychol. 6:680.
doi: 10.3389/fpsyg.2015.00680

scenarios, these women reasoned about realistic prenatal test results, and were personally interested in understanding them correctly.

In sum, contrary to the common account, naive respondents do not perform well on tasks devised to improve their understanding of posterior probability. These tasks mimic everyday problems, like calculating the post-test probability of diseases. However, they are unlikely to be the best tools to investigate whether naive respondents possess a basic intuition of posterior probability, and whether they are able to update their evaluations in the light of new evidence (Giroto and Gonzalez, 2007). Indeed, these tasks do not require respondents to revise any initial judgment (Giroto and Gonzalez, 2008; Mandel, 2014). Rather, they simply ask for only one judgment on the basis of various pieces of evidence (e.g., the prevalence rate, the result of the test and its statistical properties). Moreover, these verbal tasks convey numerical information by means of symbols and require an explicit numerical evaluation. Therefore, they can be employed only with literate respondents who have acquired a numerical symbolic system. Producing an explicit numerical estimation in numbers or words, however, is not the only way in which individuals may assess chance. Consider the following task:

[Task B]

Respondents are presented with a box containing five red chips (four round and one square) and three green chips (all square). The experimenter says, “I will take one chip out of the box without looking inside. Do you think that I will get a red or a green chip?”

Unlike Task A, and other verbal tasks used in adult Western literature, Task B does not convey probabilities by means of numerical symbols, and does not require respondents to produce an explicit numerical evaluation. Rather, it presents a set of tokens, and asks for a qualitative judgment or choice between two outcomes that may occur by taking one token out of the set at random (i.e., drawing a red vs. a green chip). To produce a suitable answer, respondents can reason extensionally, by considering and comparing the ways in which the outcomes may occur (Johnson-Laird et al., 1999). Accordingly, respondents will predict the occurrence of the outcome that may be produced in more ways (i.e., drawing a red chip). Numerate respondents could make a precise enumeration of the chances favoring each outcome (e.g., “There are 5 chances of drawing a red chip vs. 3 chances of drawing a green chip”). On this basis, they could even produce an explicit and correct absolute evaluation (e.g., “There are 5 chances out of 8 of drawing a red chip”). Of course, non-numerate respondents could not do so. However, the ability to make approximate comparisons of quantities emerges before (e.g., Barth et al., 2005) and without schooling (e.g., Pica et al., 2004). Therefore, even individuals who lack any formal numerical knowledge should produce suitable predictions in simple tasks like Task B. Indeed, both Western 5-year-olds (e.g., Davies, 1965; Giroto and Gonzalez, 2008) and preliterate Mayan adults (Fontanari et al., 2014) answer “red,” that is, they choose the more likely outcome, and they do so even when they have to consider large sets of tokens. In sum, non-numerate individuals are able to compare the chances of two competing

outcomes, without being able to express them numerically, and without necessarily making an explicit and precise counting of the number of chances favoring each of them.

Notably, these individuals also revise their evaluations on the basis of a new piece of evidence:

[Task B']

Upon the completion of Task B, the experimenter says, “I have taken one chip out of the box. I have it in my hand and I feel that it is square. Do you think that I got a red or a green chip?”

To choose the more likely outcome (“green”), respondents should focus on the subset of possibilities compatible with the evidence (the four squares). Five-year-olds do so, updating their initial judgments and choices suitably (Giroto and Gonzalez, 2008/Studies 1 and 2). They succeed even in tasks that imply more complex combinations of prior and posterior information (Bonawitz et al., 2013), or reasoning about a single, non-repeatable event produced by an intentional agent (Giroto and Gonzalez, 2008/Study 3). Fontanari et al. (2014) have extended these results by presenting preliterate Mayan adults with the same sort of tasks. Despite their lack of any sort of formal education, these respondents performed like Western controls, revising their initial choices in the light of new evidence. Finally, measures of looking times suggest that even preverbal infants form rational expectations about uncertain events by integrating different sources of information in a coherent way (Teglas et al., 2011). Together, these findings corroborate the view that, along with the application of non-extensional heuristics (Tversky and Kahneman, 1974), naive reasoning about probabilities often relies on extensional procedures: respondents infer the probability of an event from the various ways in which it could occur (Johnson-Laird et al., 1999).

Two notes are in order about the tasks that have documented the existence of an early understanding of prior and posterior probability (e.g., Task B and B'). First, these tasks are not natural frequency tasks. Indeed, they do not convey natural frequency information and do not ask for a frequency prediction. The following one is an example of a proper natural frequency task:

[Task C]

The experimenter says, “This box contains some chips. You do not know their colors. You observe me drawing a chip at random from the box, and replacing it in the box 8 times. My sample shows 5 red and 3 green chips. I'll draw a chip at random 8 more times. Do you think that the new sample will show more red or more green chips?”

Task C is apparently similar to Task B. In both cases, one can answer by considering sets of countable elements (i.e., prior possibilities and actual frequencies, respectively), and by making a similar comparison (i.e., 5 red chips vs. 3 green chips, and 5 draws of a red chip vs. 3 draws of a green chip, respectively). The two answers, however, cannot be assimilated. In Task B, one reasons about a set of prior possibilities before making any actual experience. In Task C, one reasons about a set of observations gathered through a “natural sampling,” which is “the process of encountering instances in a population sequentially. The outcome of natural sampling is natural frequencies” (Gigerenzer and Hoffrage, 1999, p. 425).

Second, tasks that do not ask for an explicit numerical evaluation, including those that imply reasoning about few possibilities, do not guarantee correct performance neither in children nor in adults (Nickerson, 1996; Johnson-Laird et al., 1999). Consider, for example, Task B. Young children succeed in it, basing their answer on prior possibilities (e.g., “You will get a red chip because there are more red than green chips”). However, if one transforms Task B into a frequency-like task, they fail. In other words, if one makes a series of random draws from the same box, and asks young children to make a prediction for each of them, they tend to use erroneous strategies like “Predict the color that was not predicted in the previous trial” (Brainerd, 1981; Teglas et al., 2007/Studies 3 and 4). It should be noted that even literate adults make erroneous predictions in situations in which they have to extract frequencies from actual observations rather than to process numerical symbols. For example, they fail versions of Task A in which they are presented with a series of medical records, each representing a patient, his/her health condition and the presence/absence of a given symptom (e.g., Gluck and Bower, 1988). Along with the finding that young children can reason correctly about events before experiencing their actual frequency, the finding that literate adults err in experience-based reasoning tasks is difficult to explain following

the hypothesis that the human mind is “developmentally and evolutionary prepared to handle natural frequencies” (Gigerenzer and Hoffrage, 1999, p. 430).

In conclusion, even literate adults have difficulties in producing correct posterior evaluations. They appear to be unable to combine prior information and new evidence in a normative way in tasks whose solution depends on the combination of numerical values, including tasks that have been devised to improve posterior probability reasoning. However, recent studies have shown that even young children and preliterate adults can succeed in tasks whose solution depend on a simple comparison of possibilities. In sum, naive individuals possess correct intuitions of prior and posterior probabilities, and such intuitions emerge early in the course of development and regardless of culture and education.

Acknowledgments

The authors thank Ulrich Hoffrage, David Mandel and Stephanie Denison for their comments on a previous version of this paper, which was supported by grants from Swiss&Global—Ca’ Foscari Foundation and the Italian Ministry of Scientific Research (PRIN grant 2010-RP5RNM).

References

- Barbey, A. K., and Sloman, S. (2007). Base-rate respect: from ecological rationality to dual processes. *Behav. Brain Sci.* 30, 241–297. doi: 10.1017/S0140525X07001653
- Barth, H., Le Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Bonawitz, E., Denison, S., Griffith, T. L., and Gopnick, A. (2013). Rational variability in children’s causal inferences: the sampling hypothesis. *Cognition* 126, 285–300. doi: 10.1016/j.cognition.2012.10.010
- Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychol. Rev.* 88, 463–502. doi: 10.1037/0033-295X.88.6.463
- Bramwell, R., West, H., and Salmon, P. (2006). Health professionals’ and service users’ interpretation of screening test results: experimental study. *Br. Med. J.* 333, 284–289. doi: 10.1136/bmj.38884.663102.AE
- Casscells, W., Schoenberger, A., and Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *N. Engl. J. Med.* 299, 999–1000. doi: 10.1056/NEJM197811022991808
- Davies, C. M. (1965). Development of the probability concept in children. *Child Dev.* 36, 779–788. doi: 10.2307/1126923
- Fontanari, L., Gonzalez, M., Vallortigara, G., and Giroto, V. (2014). Probabilistic cognition in two indigenous maya groups. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17075–17080. doi: 10.1073/pnas.1410583111
- Gigerenzer, G., and Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: a reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychol. Rev.* 106, 425–430. doi: 10.1037/0033-295X.106.2.425
- Giroto, V., and Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition* 78, 247–276. doi: 10.1016/S0010-0277(00)00133-5
- Giroto, V., and Gonzalez, M. (2007). How to elicit sound probabilistic reasoning: beyond word problems. *Behav. Brain Sci.* 30, 268. doi: 10.1017/S0140525X07001768
- Giroto, V., and Gonzalez, M. (2008). Children’s understanding of posterior probability. *Cognition* 106, 325–344. doi: 10.1016/j.cognition.2007.02.005
- Gluck, M. A., and Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *J. Exp. Psychol. Gen.* 117, 227–247. doi: 10.1037/0096-3445.117.3.227
- Hoffrage, U., and Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Acad. Med.* 73, 538–540. doi: 10.1097/00001888-199805000-00024
- Johnson-Laird, P. N., Legrenzi, P., Giroto, V., Legrenzi, M., and Caverni, J. P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychol. Rev.* 106, 62–88.
- Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Front. Psychol.* 5:1144. doi: 10.3389/fpsyg.2014.01144
- Nickerson, R. S. (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychol. Bull.* 120, 410–430. doi: 10.1037/0033-2909.120.3.410
- Pica, P., Lerner, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Pighin, S., Gonzalez, M., Savadori, L., and Giroto, V. (2015). Improving public interpretation of probabilistic test results: distributive evaluations. *Med. Decis. Making* 35, 12–15. doi: 10.1177/0272989X14536268
- Teglas, E., Giroto, V., Gonzalez, M., and Bonatti, L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19156–19159. doi: 10.1073/pnas.0700271104
- Teglas, E., Vul, E., Giroto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. (2011). Pure reasoning in 12-months-old infants as probabilistic inference. *Science* 332, 1054–1059. doi: 10.1126/science.1196404
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Giroto and Pighin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.