Cognitive Psychology 81 (2015) 26–47

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Within-subject consistency and between-subject variability in Bayesian reasoning strategies

Andrew L. Cohen *, Adrian Staub

University of Massachusetts, Amherst, United States

A R T I C L E I N F O

Article history:
Accepted 9 August 2015
Available online 29 August 2015

Keywords:
Bayesian reasoning
Individual differences
Modeling

A B S T R A C T

It is well known that people tend to perform poorly when asked to determine a posterior probability on the basis of a base rate, true positive rate, and false positive rate. The present experiments assessed the extent to which individual participants nevertheless adopt consistent strategies in these Bayesian reasoning problems, and investigated the nature of these strategies. In two experiments, one laboratory-based and one internet-based, each participant completed 36 problems with factorially manipulated probabilities. Many participants applied consistent strategies involving use of only one of the three probabilities provided in the problem, or additive combination of two of the probabilities. There was, however, substantial variability across participants in which probabilities were taken into account. In the laboratory experiment, participants' eye movements were tracked as they read the problems. There was evidence of a relationship between information use and attention to a source of information. Participants' self-assessments of their performance, however, revealed little confidence that the strategies they applied were actually correct. These results suggest that the hypothesis of base rate neglect actually underestimates people's difficulty with Bayesian reasoning, but also suggest that participants are aware of their ignorance.

* Corresponding author at: 427 Tobin Hall, 135 Hicks Way, Department of Psychology, University of Massachusetts, Amherst, MA 01003, United States. Fax: +1 (413) 545 0996.

E-mail address: acohen@psych.umass.edu (A.L. Cohen).

http://dx.doi.org/10.1016/j.cogpsych.2015.08.001
0010-0285/

1. Introduction

It is well known that people perform poorly on problems that require them to reason about probabilities in a Bayesian manner (Bar-Hillel, 1980; Kahneman & Tversky, 1972). The following is a classic example:

> The probability of breast cancer is 1% for a woman at age forty who participates in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?
>
> [Gigerenzer & Hoffrage, 1995, p. 685; adapted from Eddy, 1982]

In this example, and in the many similar examples in the literature, the participant is provided with three probabilities, and must use these to compute a fourth. The participant is given the probability that some hypothesis obtains, in the absence of specific data. This probability is known as the *prior* or *base rate*, $p(H)$, e.g., the 1% probability that a forty-year-old woman has breast cancer. The participant is also given the *likelihood* or *true positive rate*, $p(D|H)$, the conditional probability that if the hypothesis is true, the datum in question would be observed, e.g., the 80% probability that if a woman has breast cancer, her mammogram will be positive. Finally, the participant is given the *false positive rate*, $p(D|{\sim}H)$, the conditional probability that if the hypothesis in question is not true, the datum in question would be observed, e.g., the 9.6% probability that a woman who does not have breast cancer will get a positive mammogram.

According to Bayes' Theorem, the correct answer to this problem is .078, or 7.8%. Bayes' Theorem provides the *posterior probability* of a hypothesis, $p(H|D)$, on the basis of the three probabilities given in the problem:

$$p(H|D) = \frac{p(D|H)p(H)}{p(D|H)p(H) + p(D|\neg H)p(\neg H)}. \tag{1}$$

Although there is some evidence that people – even young children – can reason probabilistically (e.g., Girotto & Gonzalez, 2008), typically few of the answers provided by experimental participants are identical to, or even very near, the normatively correct posterior probability. Often, and especially when the base rate is low, participants' answers are much too high. This finding is true not only of the undergraduate students, who are the typical participants in such experiments. For example, Eddy (1982) found that even physicians' answers to the medical diagnosis problem above were on average too high by a factor of about ten. Although cognitive and demographic factors appear to play a role in predicting participants' performance on these problems (e.g., Brase, Fiddick, & Harries, 2006; Chapman & Liu, 2009; Sirota, Juanchich, & Hagmeyer, 2014), it is clear that the vast majority of participants in experimental studies do not produce the normatively correct answers.

A large literature has investigated the role of problem format in Bayesian reasoning problems, typically focusing on the potential benefits of presentation of natural frequencies (e.g., 10 out of 1000 forty-year-old women have breast cancer) as opposed to the standard presentation of probabilities (e.g., Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Zhu & Gigerenzer, 2006). While it appears that presentation in the form of natural frequencies can indeed improve performance (e.g., Hill & Brase, 2012; Zhu & Gigerenzer, 2006), the mechanism underlying this facilitation is unclear, and even the natural frequency format does not result in correct posterior estimates in most cases (Barbey & Sloman, 2007).

A standard interpretation of these findings holds that participants either ignore the base rate altogether, or weight it less than they should (Bar-Hillel, 1980; see Koehler, 1996, for discussion and review). This interpretation, which is known as *base rate neglect*, holds that answers to problems such as the medical diagnosis problem are too high because reasoners fail to appropriately utilize the information that only 1% of women have breast cancer. In this example, the true positive rate is high (80%), and the false positive rate is low (9.6%), so in the absence of base rate information (or assuming uniform priors, i.e., that having cancer and not having cancer are equally likely) one would conclude that a