Published in final edited form as:

Epidemiology. 2017 March ; 28(2): e10–e11. doi:10.1097/EDE.0000000000000587.

Dogmatists cannot learn

Stephen R. Cole^a, Haitao Chu^b, M. Alan Brookhart^a, and Jess K. Edwards^a

^aDepartment of Epidemiology, UNC Gillings School of Global Public Health

^bDepartment of Biostatistics, University of Minnesota School of Public Health

To the Editor

We wish to provide some rational in defense of the title of this comment. If we agree that a dogmatist is one whose beliefs cannot be influenced by observations (i.e., data), and we define learning as having your beliefs influenced by observations, then it follows that dogmatists cannot learn. While this statement has been made previously ((1), page 47)(2), we believe it is useful to expand on this point with a simple example. Indeed, some dangers of dogmatism in epidemiology have been documented (e.g. (3)), and one does not have to look far for trivial examples of dogmatic statements (4).

First, recall that the standard rules of probability theory are convexity, $0 \leq P(a) \leq 1$; multiplication, $P(a \text{ and } w) = P(a|w)P(w)$, where $P(a|w)$ is the probability of arbitrary statement a presuming arbitrary statement w holds; and addition, $P(a \text{ or } w) = P(a) + P(w) - P(a \text{ and } w)$. Next, we will need to review a simple application of probability theory which follows from the above rules (5), and is adapted from an example given by Lindley ((6), page 131). Here we are using probability to denote belief, where probability 1 is certainty and probability 1/2 is equally likely as not. Though we also ascribe to probability as chance, and strive to follow Lewis' principal principle (i.e., to bring your belief-probabilities in line with known chance-probabilities) (5).

Consider the following question: XDSWQGFCVDSUDZQVOQDQUYLLEM? To make things interesting, this question is encrypted in a language we do not understand. For the purposes of this exercise, say we know that the possible answers to this question are: 1/3 or 2/3. We will let θ denote this discrete parameter space, such that $\theta = 1/3$, or $\theta = 2/3$. We can extend to a continuous parameter space, but we wish to keep this example simple enough to tabulate.

Before we see any data, $n = 0$, we might have no knowledge about the relative probability of the two possible values of θ . To translate this ignorance into a prior probability, we employ Laplace's rule of succession ((7), page 19), $P = (y + 1)/(n + 2)$, and assign probability 1/2 to each possible value of the statement. This means we have prior probability $P(\theta = 1/3) = P(\theta = 2/3) = 1/2$. To learn, we might gather observations or information. As an aside, information seems necessary but is not sufficient for learning. In addition to information, we

Correspondence to Dr. Cole, cole@unc.edu, UNC Campus Box 7435, Chapel Hill, NC, 27599-7435.

The authors declare no conflicts.

require a system or engine to translate the information into knowledge. Here, in a setting where we wish to learn about the factual natural course (i.e., what is), we use probability logic as our engine (8). In more complex settings where we wish to learn about what might be we use a counterfactual probability logic as our engine (9). Information coupled with our engine and identification conditions appears sufficient for learning, but is not necessary (as other engines or conditions exist).

Say we find 12 people who can decrypt or translate this question and report to us whether they believe that $\theta = 1/3$, or $\theta = 2/3$. For simplicity, we will consider these 12 people independent and exchangeable (10) (or permutable (11)) with respect to their decryption abilities. Also, we will ignore the certainty with which their reports are provided. Say the reports are 7 of 12 in favor of $\theta = 2/3$. Recall that Bayes' rule is $P(w|a) = P(a|w)P(w)/P(a)$. In our setting this is $P(\theta|y) = P(y|\theta)P(\theta)/P(y)$, where y is shorthand for the observed data y/n and $P(y|\theta)$ is the likelihood (12). In our setting the likelihood is binomial, or

$$P(y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

. Details of calculations are shown in the upper panel of the Table. Our posterior probability that $\theta = 2/3$ is 0.8, much increased from our prior probability that $\theta = 2/3$, which was 0.5.

Now we are prepared to demonstrate that dogmatists cannot learn. Say our prior probabilities are as given in the lower panel of the Table. This dogmatic prior probability completely rules out the option $\theta = 2/3$. The data and resultant likelihood are unchanged from the upper panel of the Table. Regardless of the data, here the dogmatists' posterior equals their prior. Therefore, dogmatists cannot learn. Of course, this prior statement (and the title of this comment) is itself dogmatic, though we encourage dissent (2, 13). Finally, we are usually dogmatic regarding options we do not foresee. A point of this comment is to prepare ourselves to be open to alternatives as they present themselves. As Wittgenstein said: "What is thinkable is also possible." ((14), comment 3.02).

Acknowledgments

Dr. Cole and Dr. Edwards were funded in part by NIH grant R01AI100654.

References

1. Savage, LJ. The foundations of statistics. 2nd. New York: Dover; 1972.
2. Cole SR. Nondogmatism. *Ann Epidemiol*. 2016 In press.
3. Stolley PD. When genius errs: R.A. Fisher and the lung cancer controversy. *Am J Epidemiol*. 1991; 133(5):416–425. discussion 426–418. [PubMed: 2000852]
4. Cole SR. Atheoretical science: a response to the poverty of popperian epidemiology (comment). *Int J Epidemiol*. 1996; 25(4):899–900. [PubMed: 8921473]
5. Greenland S. Probability logic and probabilistic induction (with comment by Maclure). *Epidemiology*. 1998; 9(3):322–332. [PubMed: 9583426]
6. Lindley, DV. Understanding uncertainty. New York: John Wiley & Sons; 2014. Revised ed
7. Laplace, PS. A philosophical essay on probabilities. New York: Wiley; 1814. Translated in 1902 by Truscott, F.W., Emory, F. W
8. Jaynes, ET. Probability theory: the logic of science. Cambridge University Press; 2003.

9. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: Application to control of the healthy worker survivor effect. *Math Modelling*. 1986; 7:1393–1512.
10. de Finetti B. Foresight: Its logical laws, its subjective sources. *Annales de l'Institut Henri Poincaré*. 1937; 7:95–157.
11. Good, IJ. *Good Thinking*. Minneapolis: University of Minneapolis Press; 1983.
12. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *Am J Epidemiol*. 2014; 179(2):252–260. [PubMed: 24173548]
13. Watkins JWN. Comprehensively critical rationalism. *Philosophy*. 1969; 44:57–62.
14. Wittgenstein, L. *Tractatus Logico-Philosophicus*. New York: Harcourt, Brace & Company; 1921.

Table 1

Posterior probability distributions under two different priors.

Laplace prior				
θ	Prior Probability $P(\theta)$	Likelihood $P(y \theta)^a$	$P(y \theta)P(\theta)$	Posterior Probability $P(\theta y)^c$
1/3	0.5	0.04769	0.02385	0.2
2/3	0.5	0.19076	0.09538	0.8
Total	1.0	–	$P(y)=0.11922^b$	1.0

Dogmatic prior				
θ	Prior Probability $P(\theta)$	Likelihood $P(y \theta)^a$	$P(y \theta)P(\theta)$	Posterior Probability $P(\theta y)^c$
1/3	1.0	0.04769	0.04769	1.0
2/3	0.0	0.19076	0.0	0.0
Total	1.0	–	$P(y)=0.04769^b$	1.0

$$^a P(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

$$^b P(y) = \sum_{\theta} P(y|\theta) P(\theta)$$

$$^c P(\theta|y) = P(y|\theta) P(\theta) / P(y)$$