

Replies to Kristin Andrews's, Gordon Belot's, and Patrick Forber's reviews

**Elliott Sober: Ockham's razors: A user's manual. Cambridge:
Cambridge University Press, 2016, 322pp, \$29.99 PB, \$99.99 HB**

Elliott Sober¹

Published online: 21 September 2016
© Springer Science+Business Media Dordrecht 2016

I am grateful to Kristin Andrews, Gordon Belot, and Patrick Forber for their incisive comments on *Ockham's Razors* (henceforth OR). I do not have space to reply to all their ideas, so I will have to be selective.

Reply to Patrick Forber

Forber's discussion of my "modest pluralism" provides me with a welcome opportunity to clarify how the various razors that float around in my book are related to each other. Forber thinks that my pluralism consists in the fact that I discuss both the razor of silence and the razor of denial, but in fact my pluralism is mainly to be found elsewhere. Both these razors concern whether you should believe a proposition. The first says you should remain silent (suspend judgment) about whether a hypothesis is true if the hypothesis is not needed to explain anything. The second says you should deny (disbelieve) that a hypothesis is true if it is not needed to explain anything. I note in OR that the razor of silence (or something like it) has a Bayesian rationale (p. 71), but I never defend the razor of denial and I criticize the way it is sometimes used (e.g., p. 253, p. 265, and p. 276). I talk about these two razors from time to time in OR because of their historical importance and also because philosophers and scientists sometimes invoke Ockham's razor to deny that a hypothesis is true without explaining why they choose denial rather than silence.

The pluralism about parsimony that I embrace in OR is mainly to be found in my thesis that there are three "parsimony paradigms"; each shows why parsimony is epistemically relevant. None of these paradigms provides a sufficient condition for believing a hypothesis or for disbelieving it, and none makes mention of whether a

✉ Elliott Sober
ersober@wisc.edu

¹ Department of Philosophy, University of Wisconsin, Madison, WI, USA

hypothesis is needed for explanation. This is why the three paradigms cannot be said to justify either the razor of silence or the razor of denial. Nonetheless, it is true that the paradigms do address questions that are *relevant* to rational belief (pp. 150–151), but their importance goes considerably beyond that relevance. Each paradigm shows why parsimony is epistemically relevant even if the concept of dichotomous belief is set to one side because it is too coarse-grained to be of much use in epistemology.

The first parsimony paradigm connects parsimony with likelihood. The second connects parsimony with predictive accuracy. And the third connects parsimony with probability. For me, this third paradigm plays third fiddle (p. 141). My distinguishing the first paradigm from the third may sound odd, given that “likely” and “probably” are synonyms in ordinary English. However, these two paradigms really are different because I use “likelihood” in the technical sense that is standard in statistics. Consider hypothesis H and observation O. The probability of H given O is a different quantity from the probability of O given H. The latter is confusingly called H’s likelihood. Parsimony is used in these three paradigms to answer different epistemic questions, and so the paradigms do not conflict.

With respect to the razor of silence and the razor of denial, Forber thinks that “in principle, this distinction is ironclad,” though “in practice ... the distinction may be of limited use, for the way we use models to represent phenomena may not respect the difference. It may be that silence in our representations effectively is denial of the phenomena.” I agree with Forber that it is sometimes unclear whether a hypothesis is denying some proposition or is remaining silent about whether it is true. However, once this is clarified, the distinction between silence and denial is not just okay in principle; it is often important in practice. There is often an important practical difference between suspending judgment on a hypothesis because of insufficient evidence and deciding that the evidence licenses rejection.

Forber defends his claim that the silence/denial distinction may be of limited use by discussing the causal modeling framework developed by Spirtes et al. (2001). In the SGS context, Forber thinks that the error terms in a model should not be interpreted as summarizing the influence of causal variables that are not represented explicitly in the arrow diagram or in the structural equations. He worries that if error terms are allowed to take on that role, they will “hide real causal influence.” I think there is nothing wrong with using error terms in this way, and my impression is that this is a standard interpretation of what they mean. There is no disreputable hiding going on here, only the modest admission that the model may be incomplete.

Here is a better example that illustrates how interpreting a model’s “silences” can be a delicate matter. Consider a causal model that says that X and Y each cause Z. The model is represented by an arrow diagram and a set of structural equations. The diagram and the equations do not say whether X and Y have a common cause; indeed, they do not say whether X and Y have any causes at all. However, the SGS framework interprets these two “silences” differently. The framework has no quarrel with the model’s being agnostic on whether X and Y have causes, but it is part of the framework that if a model fails to explicitly say that X and Y have a common cause, then the model is committed to there being none. Here “silence” seems to mean denial, but, as noted, I do not think that shows that distinguishing the

razor of silence from the razor of denial has limited utility. The point is just that the framework's interpretation of a model goes beyond what is stated in the arrow diagram and the structural equations.

Forber makes an additional point about the razor of silence in connection with my discussion of whether chimpanzees are mind readers. He says that “one thread of the argument in De Waal's (2016) for inferring rich mental lives of animals turns on the importance of considering a range of possible hypotheses for animal cognition. In effect, the razor of silence limits this range of possibilities.” Forber's point, I think, is that the razor of silence wrongly forecloses legitimate hypotheses. I disagree; the razor of silence does not tell you to ignore a hypothesis, but rather advises you to be agnostic about it if you think it is not needed to explain anything. Furthermore, the razor of silence can be reformulated so that it says that you should be agnostic about a hypothesis if there is no evidence that discriminates between it and relevant alternatives. De Waal (1991) plays by this rule when he argues that the close genealogical relatedness of human beings and chimpanzees means that our being mind readers shows that they probably are too, an argument I discuss in Chapter 3 of OR.

I now turn to Forber's comments on the historical material in Chapter 1 of OR. This chapter is a series of brief snapshots that show how different thinkers, from Aristotle to C. Lloyd Morgan, used Ockham's razor and what they said about the razor's justification. In the section about Copernican and Ptolemaic astronomy, I describe how Copernicus and his student Rheticus both invoke the principle of parsimony as a reason to prefer heliocentrism over geocentrism. It was not until later that astronomers obtained observational evidence that discriminated between these two hypotheses. As other scholars have done, I note that Copernicus and Ptolemy both used epicycles in their theories. This means that if Copernicus's theory is more parsimonious, this is not because he does without epicycles. Drawing on the analyses of Lange (1995) and Myrvold (2003), I argue that the Copernican system predicts observational regularities that the Ptolemaic system can only accommodate. In Chapter 2, I clarify the distinction between prediction and accommodation and explain how the distinction connects with Ockham's razor.

Reacting to my claim that parsimony was taken to be relevant to the competition between geocentrism and heliocentrism, Forber says that “the razor had little traction until Kepler improved the Copernican system, doing away with epicycles in favor of elliptical orbits. Sober explains the impact of Kepler's innovation as providing a way for the heliocentric model to *predict* certain generalities about retrograde motion of planets whereas the geocentric system merely *accommodates* them.” My reply is that it was Copernicus and Rheticus who put Ockham's razor on the table, and this happened well before Kepler entered the fray and got rid of epicycles. Also, my appeal to the distinction between prediction and accommodation is not offered as an explanation of why Kepler's theory had an impact, but of the logic underlying Copernicus and Rheticus's thought that heliocentrism is better than geocentrism. Kepler and Galileo later agreed with that general line of argument.

Forber notes two silences in Chapter 1. I do not talk about Tycho Brahe's geocentric theory when I discuss Copernicus versus Ptolemy, and when I talk about

Newton's Rules of Reasoning, I do not discuss their role in regulating idealizations. As interesting as both topics are, they were not relevant to the task I set myself in Chapter 1. Forber says that the omission of Brahe means that my "analysis does not accurately represent the controversy" concerning geocentrism and heliocentrism. I agree if he means that my account is *incomplete*. In any event, I am pleased that Forber thinks that the "contrast [between prediction and accommodation] may favor the Copernican system over the Ptolemaic system." As for Newton, my goal was to show that his Rules of Reasoning invoke Ockham's razor, to note the Aristotelian provenance of one of his formulations, and to describe Newton's theological justification of the razor, which he presents in an unpublished manuscript on interpreting Biblical prophecies. I explore the relationship between idealization and parsimony in Chapter 2, but not in connection with Newton.

Reply to Kristin Andrews

To respond to Andrews's interesting discussion of the ongoing controversy in comparative psychology about whether chimpanzees are mind readers, I first want to provide some background about the analysis I develop in Chapter 4 of OR.

Suppose you run an experiment and find that chimpanzees produce response R more often when they are in stimulus situation S than they do when they are not in S. Consider two hypotheses about what is going on here:

$$(H_1) S \rightarrow I_1 \rightarrow R$$

$$(H_2) S \rightarrow I_1 \rightarrow I_2 \rightarrow R$$

The first says that S causes intervening variable I_1 , and I_1 causes R. The second says that S causes I_1 , I_1 causes I_2 , and I_2 causes R. The arrows here should be understood probabilistically. I_1 says that chimpanzees form some specified beliefs about the behaviors of others; I_2 says that chimpanzees form some particular beliefs about the mental states of others. If what you observe in your experiment is just the frequency data I mentioned, then your data do not discriminate between the two hypotheses. I interpret Povinelli and his coauthors as making this point. And just as you can snip I_2 from H_2 , you also can snip I_1 from H_1 . The result is:

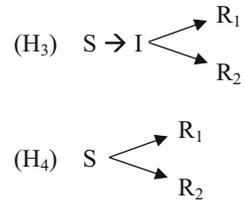
$$(H_0) S \rightarrow R$$

H_0 conforms to the strictures of methodological behaviorism; it postulates no intervening variable at all. It too is compatible with your frequency data.

This example may suggest something more general—that, for *any* model that postulates intervening variables, you can create a new model that deletes those variables, and the old and new models will be empirically indistinguishable. It is important to see that this thesis is wrong. To see why, consider the following two models (Fig. 1).

Given the framework that is now widely used in causal modeling (Spirtes, Glymour, and Scheines 2001; Woodward 2005; Pearl 2009), H_3 and H_4 make different predictions. H_4 says that S screens off R_1 and R_2 from each other, while H_3

Fig. 1 Two models that disagree about screening-off



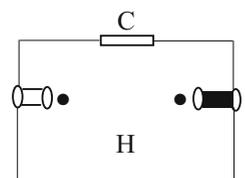
denies this. That is, H₄ says that the R's are probabilistically independent, conditional on S, while H₃ says that they are probabilistically dependent, conditional on S. Introducing or deleting an intervening variable *sometimes* changes a model's predictions, even though this is not true in the case of H₀, H₁, and H₂.

In Chapter 4 of OR, I design an experiment in which chimpanzees confront a pair of tasks. My experiment combines two experiments that were carried out separately by Melis et al. (2006). In both tasks, the chimpanzee (C) can look through a window to see that there is a human being (H) in a chamber. C can see that there are two items of food in there as well. In the first task, C chooses between using an opaque tunnel or a transparent tunnel to reach into the chamber, as shown in Fig. 2. If the opaque tunnel is used, H allows C to remove the food; if the transparent tunnel is used, H prevents C from doing so. In the second task, the chamber has noisy and silent trapdoors instead of transparent and opaque tunnels. C chooses a trapdoor to use in reaching into the chamber. If the silent trapdoor is used, H allows C to remove the food; if the noisy trapdoor is used, H prevents C from doing so. In the experiment I propose, C faces the tunnel problem and then the trapdoor problem. Sometime later, C faces the same pair of problems. Then, after a while, C again faces the pair of tasks. And so on. The result is a two-winged *n*-trial experiment.

In Fig. 3, you see a mind-reading hypothesis (MRH) about what is going on in the experiment. MRH says that the tunnels + trapdoor stimulus fails to screen-off the two behavioral responses from each other; it says that there should be a correlation between using the opaque tunnel and using the silent trapdoor. According to MRH, C is able to understand that there is a connection between the first task and the second; this common thread is represented by the box in the middle—the belief that in both tasks, success depends on whether H notices C. MRH does not say whether this ability is strong or weak; it just says that it is not zero.

I argue in OR that this connection between the two tasks should be invisible to chimpanzees if they are (purely) behavior readers. This idea is represented in Fig. 4

Fig. 2 The chimpanzee outside the chamber can reach inside to get at two food items, either by using a transparent tunnel or an opaque tunnel. There is a human experimenter inside (Melis et al. 2006)



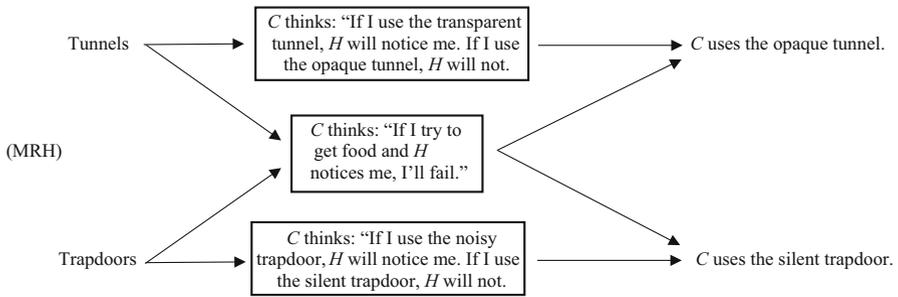


Fig. 3 A mind-reading hypothesis about the 2-wing n-trial experiment

by the behavior-reading hypothesis BRH. BRH entails that the stimulus conditions screen-off the two responses from each other, which is what MRH denies. Data from the experiment thus can discriminate between MRH and BRH. I make no prediction in OR about which hypothesis the data will favor.

Andrews agrees that these two hypotheses make different predictions, but contends that the general idea of behavior reading is not committed to BRH’s screening-off claim. I said a bit about this possibility in OR (pp. 240–241), pointing out that a new behavior-reading model, BRH*, can easily be constructed by inserting two vertical arrows (one pointing up, the other down) between the two boxes in BRH. MRH and BRH* both predict a failure of screening-off.

I made two comments in OR on whether BRH* properly represents the hypothesis of behavior reading. First, what would friends of behavior reading say if the data from this experiment turned out to be just what one would expect according to BRH? I doubt that these friends would regard that as bad news for their position. This is one reason to think that the behavior-reading hypothesis should not be reconfigured so that it is committed to there being a failure of screening-off. But there is a further reason why BRH* is not a good representation of what a behavior-reading hypothesis should say about this experiment. If chimpanzees are not mind readers, why should their thinking “I will get food if I use the opaque tunnel, but not if I use the transparent tunnel” raise the probability of their thinking “I will get food if I use the silent trapdoor, but not if I use the noisy trapdoor”? It is not enough for defenders of behavior reading to simply *postulate* that chimpanzees are built that way. As Heyes (2015) points out, behavior-reading hypotheses should not be

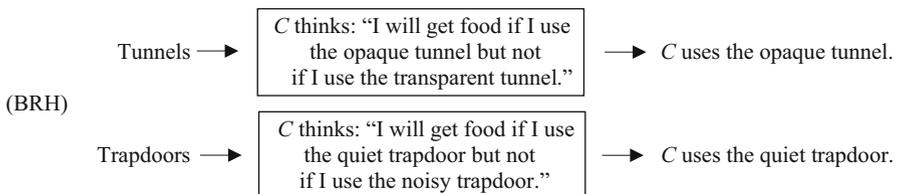


Fig. 4 A behavior-reading hypothesis about the 2-wing n-trial experiment

concocted out of thin air; they should have some antecedent and independent plausibility in cognitive science.

Andrews suggests a different way of constructing a behavior-reading hypothesis that predicts a failure of screening-off. Rather than adding arrows to BRH as I just described, she modifies MRH. Her idea is that “perhaps we can unify the behavior-reading hypotheses in terms of *detection*. Detection can be glossed as responsiveness to the presence or absence of a certain feature of the world. A piece of litmus paper is a detector, and not minded.” This new version of BRH (call it BRH**) can be constructed from MRH by placing “If I try to get food and H detects me, I’ll fail” in MRH’s middle box. Heyes’s point applies to this suggestion as well. Litmus paper is a mindless detector, but can a chimpanzee do the detecting postulated by BRH** without being a mind reader?

Andrews suggests that the mind-reading hypothesis and the behavior-reading hypothesis may be empirically indistinguishable. It is not just that MRH, BRH*, and BRH** are; she floats the idea that, for any mind-reading hypothesis, a behavior-reading hypothesis can be concocted that makes the same predictions. According to Andrews, Povinelli and Vonk think this is so; she says that “according to [their] reinterpretation hypothesis the added value of mindreading is explanation, not enhanced prediction.” Andrews also says that “the theory of behavior that Povinelli and colleagues build into their model is so rich and robust that it is going to be behaviorally indiscernible from a mind-reading model.” Even if this is the right interpretation of Povinelli and Vonk (2004), Penn and Povinelli (2007) tell a different story. They say that “the ability to cognize the world from the cognitive perspective of another agent would provide the animal with enormous advantages over and above the ability to reason in terms of observable first-person relations alone (p. 741).” These advantages, I take it, include an enhanced ability to make predictions about the behaviors of others. Indeed, Andrews quotes their saying that “being able to recode perceptually disparate behavioral patterns resulting from the same underlying cognitive state as instances of the same abstract equivalence class is a bona fide example of postulating [a mind-reading] variable.” The implication is that this recoding cannot be done by pure behavior reading. At the end of her review, Andrews remarks that “the behavior-reading hypothesis is a bit of shifting sand that does not manage to stay still for very long.” She adds that “it is not exactly clear what they [Povinelli and Vonk] do think.” The puzzle that Andrews raises is important. It is clear enough which representational contents the behavior-reading hypothesis prohibits; what requires clarification is the causal connections between belief states that behavior-reading hypotheses are permitted to postulate.

In designing the two-winged experiment, I wanted to find two tasks that would “look different” to chimpanzees if they were behavior readers, but might “look the same” if they were mind readers. It may be that the tasks I chose were not different enough. Opaque and transparent tunnels seem rather different from silent and noisy trapdoors, but both tasks involve obtaining food by reaching into a chamber where a human being is present. This may mean that the details of the experiment I described need to be tweaked. I nonetheless hope that the two-task *n*-trial format will be useful in cracking what is in fact a very hard nut. Povinelli demands that a mind-reading hypothesis identify the “unique causal work” that mind reading does.

This problem cannot be solved if we consider models like H_0 , H_1 , and H_2 , but the situation changes when we consider models like H_3 and H_4 . H_3 and H_4 differ in their predictions about screening-off just as MRH and BRH do.

Reply to Gordon Belot

I mentioned in my reply to Forber that the parsimony paradigms I describe in OR are not in conflict since they relate to different questions. The likelihood paradigm has to do with cases in which a simple hypothesis fits the observations you already have better than a more complex hypothesis does. The model selection paradigm has to do with using the simplicity of a model to help judge how accurately a model will predict new data when fitted to old.

Although these two paradigms are different, each leads me to be a reductionist about parsimony (p. 149). Parsimony is not an epistemic end in itself, but is justified only insofar as it promotes some more fundamental epistemic goal. Within the first paradigm, I show in Chapter 2 that when common cause and separate cause explanations are formulated in a certain standard way (inspired by Hans Reichenbach's work on the principle of the common cause), the common cause hypothesis will have the higher likelihood. That is, there are Reichenbachian assumptions that entail that the data are more probable under the common cause hypothesis than they are under the hypothesis of separate causes. The law of likelihood (see Hacking 1965) interprets this inequality to mean that the data *favor* the first hypothesis over the second. I then point out that if you modify the Reichenbachian assumptions, you can get the favoring to flip—the same data will favor separate causes over common cause. With these altered assumptions, the likelihood justification of preferring common cause over separate causes evaporates. I am undismayed. When likelihood and parsimony conflict I say: *so much the worse for parsimony*. When they agree, I take that agreement to provide a likelihood justification of parsimony.

Although I am now a reductionist about parsimony, I have sometimes thought that I should be an eliminativist. I toyed with this position when I wrote a paper called “Let's Razor Ockham's Razor” (Sober 1990). If the common cause hypothesis has the higher likelihood, why cite the fact that the common cause hypothesis is more parsimonious? If likelihood is the relevant consideration, maybe parsimony is an epistemic epiphenomenon.

I resisted eliminativism in OR when I discussed common cause versus separate causes because the common cause and separate cause explanations I considered are mostly the same, save for the fact that the former postulates one cause and the latter postulates two. Comparing the common cause and separate cause hypotheses approximates what empirical scientists do when they carry out a controlled experiment (p. 150). The common cause hypothesis has the higher likelihood in part *because* it is more parsimonious. It is true that you can compare the common cause and separate cause explanations without ever using the p-word. But that does not show that parsimony is epistemically irrelevant.

A similar set of issues arises in connection with the second parsimony paradigm. Belot points out that the Akaike information criterion (AIC) is asymptotically equivalent to a method called “leave-one-out cross validation” (CV); I mention in OR (p. 133) that this result is due to Stone (1977). AIC explicitly takes account of a model’s parsimony (its number of adjustable parameters) in estimating the model’s predictive accuracy, but CV does not. Given this, Belot poses the following neat challenge: If AIC’s reference to parsimony shows that parsimony is epistemically relevant, then CV’s failure to refer to parsimony shows that parsimony is epistemically irrelevant.

I disagree. The fact that CV does not use the *p*-word does not show that parsimony is irrelevant in model selection, just as the likelihood evaluation of common cause and separate cause explanations does not show that parsimony is irrelevant to that comparison. We know from Akaike’s theorem that AIC is an asymptotically unbiased estimator of predictive accuracy, given the assumptions that Akaike makes. In estimating a model’s predictive accuracy, AIC imposes a penalty for complexity by subtracting k , the number of adjustable parameters in the model, from the log-likelihood of the fitted model. The mathematics shows why parsimony matters.

Having said this, I want to agree with Belot that more work is needed on the relationship of AIC and CV. As mentioned, there is Stone’s (1977) result about their asymptotic equivalence; however, when two or more models are evaluated in the light of a single finite data set, AIC and CV can fail to be ordinally equivalent in their ranking of candidate models. When this happens, which evaluation of comparative predictive accuracy is more trustworthy? One answer would be that you should postpone judgment as to which model is better and simply gather more data, hoping that the disagreement between AIC and CV will disappear. This is a practical suggestion, but it fails to address the theoretical question of which method is on firmer footing. My impression is that the theoretical foundations of CV and its relation to AIC have not been much explored (Elisseeff and Pontil 2003). Notice how my question differs from Belot’s. He focuses on the asymptotic agreement of AIC and CV and asks what this reveals about the relevance of parsimony. I am thinking about cases where AIC and CV disagree.

Belot extracts a very circumspect conclusion from his discussion of AIC and CV. He says that we are

able only to assert very weak conditional conclusions: if future discoveries determine that all good methods of model selection are like AIC rather than CV, then parsimony is epistemically relevant to model selection for predictive accuracy. But if all good methods are like CV rather than AIC then parsimony is epistemically irrelevant.

In contrast, my view is that you do not need to make a synoptic survey of all possible methods of model selection to assess whether parsimony matters. The same point holds for the question of whether some empirical claim is true (e.g., that smoking cigarettes causes lung cancer)—you can give a reasonable answer to this question without having to survey all possible theories that describe how smoking

and cancer are related. In both cases, your assessment is made given the limited evidence and options at hand, and might well be overturned as you learn more.

There is another idea from model selection theory that Belot might have taken up in posing his challenge. Rather than focus on AIC and CV, consider the relationship of AIC and TIC, the Takeuchi information criterion. TIC does not consider a model's parsimony and it is an asymptotically unbiased estimator of predictive accuracy. Takeuchi (1976) proved this without assuming that one of the candidate models considered is true or close to the truth. Akaike, on the other hand, was able to prove his result about AIC only because he made that assumption. Takeuchi's result is more general, and you can derive AIC from TIC if you add the assumption that one of the candidate models is true (p. 134). So TIC is more fundamental than AIC, and TIC says nothing about parsimony. Does this count against the claim that parsimony is relevant to model selection? I sense that Belot would say *yes*, but I do not hesitate to say *no*. AIC is a special case of TIC, but reductionism is not the same as eliminativism. The derivation of AIC from TIC vindicates AIC's claim that parsimony is epistemically relevant and shows that its relevance depends on an assumption. In similar fashion, the Reichenbachian derivation vindicates parsimony's epistemic relevance in the comparison of common cause and separate cause explanations and shows that its relevance depends on assumptions. In each case, the derivation, far from showing that parsimony is irrelevant, explains *when* and *why* it is relevant.

Acknowledgments I thank Dylan Beschner, Hayley Clatterbuck, Daniel Hausman, Michael Schon, Mike Steel, Olav Vassend, Naftali Weinberger, and Aaron Yarmel for useful discussion.

References

- De Waal, F. 1991. Complementary Methods and Convergent Evidence in the Study of Primate Social Cognition. *Behaviour* 118: 297–320.
- De Waal, F. 2016. *Are We Smart Enough to Know How Smart Animals Are?* New York: W. W. Norton and Company.
- Elisseeff, A., and M. Pontil. 2003. Leave-one-out Error and Stability of Learning Algorithms with Applications. In *NATO Science Series Sub Series III: Computer and Systems Sciences*, vol. 190, ed. J. Suykens, G. Horvath, S. Basu, C. Micchelli, and J. Vandewalle, 111–130. Amsterdam: IOS Press.
- Hacking, I. 1965. *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.
- Heyes, C. 2015. Animal Mindreading—What's the Problem? *Psychonomic Bulletin & Review* 22: 313–327.
- Lange, M. 1995. Spearman's Principle. *British Journal for the Philosophy of Science* 46: 503–521.
- Melis, A., J. Call, and M. Tomasello. 2006. Chimpanzees (*Pan troglodytes*) Conceal Visual and Auditory Information from Others. *Journal of Comparative Psychology* 120: 154–162.
- Myrvold, W. 2003. A Bayesian Account of the Virtue of Unification. *Philosophy of Science* 70: 399–423.
- Pearl, J. 2009. *Causality—Models, Reasoning, and Inference*, 2nd ed. Cambridge: Cambridge University Press.
- Penn, D.C., and D.J. Povinelli. 2007. On the Lack of Evidence that Non-human Animals Possess Anything Remotely Resembling a 'Theory of Mind'. *Philosophical Transactions of the Royal Society B* 362: 731–744.
- Povinelli, D.J., and J. Vonk. 2004. We don't Need a Microscope to Explore the Chimpanzee's Mind. *Mind and Language* 19: 1–28.
- Sober, E. 1990. Let's Razor Ockham's Razor. In *Explanation and Its Limits*, ed. D. Knowles, 73–94. Cambridge: Cambridge University Press.

- Spirtes, P., Glymour, C., and Scheines, R. 2001. *Causation, prediction, and search*, 2nd edition. Cambridge, MA: MIT Press.
- Stone, M. 1977. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society B* 39: 44–47.
- Takeuchi, K. 1976. Distribution of Information Statistics and a Criterion of Model Fitting. *Suri-Kagaku (Mathematical Sciences)* 153: 12–18 (**in Japanese**).
- Woodward, J. 2005. *Making Things Happen*. Oxford: Oxford University Press.