Evolution and the Problem of Other Minds
Author(s): Elliott Sober
Source: *The Journal of Philosophy*, Vol. 97, No. 7 (Jul., 2000), pp. 365–386
Published by: Journal of Philosophy, Inc.
Stable URL: http://www.jstor.org/stable/2678410
Accessed: 14/01/2015 07:17

EVOLUTION AND THE PROBLEM OF OTHER MINDS*

The following diagram illustrates two inference problems. First, there is the strictly third-person behavior-to-mind problem[1] in which I observe your behavior and infer that you occupy some mental state. Second, there is the Self-to-Other problem, in which I notice that I always or usually occupy some mental state when I behave in a particular way; then, when I observe you produce the same behavior, I infer that you occupy the same mental state. This second inference is the subject of the traditional philosophical problem of other minds.

How are these inferences related? Notice that the inputs to behavior-to-mind inference are a subset of the inputs to Self-to-Other inference. In the first, I consider the behavior of the other individual; in the second, I consider that behavior, as well as my own behavior and mental state. This suggests that, if Self-to-Other inferences cannot be drawn because the evidence available is too meager, the same will be true of behavior-to-mind inferences as well.

I see no reason to think that a purely third-person scientific psychology is impossible. Of course, if one is a skeptic about *all* nondeductive inference, that skepticism will infect the subject matter of psychology. And if one thinks that scientific inference can *never*

    [1] This terminology is from Larry Shapiro, "Presence of Mind," in Valerie Gray Hardcastle, ed., *Biology Meets Psychology: Constraints, Connections, and Conjectures* (Cambridge: MIT, 1999), pp. 133-50.

                                    © 2000 The Journal of Philosophy, Inc.
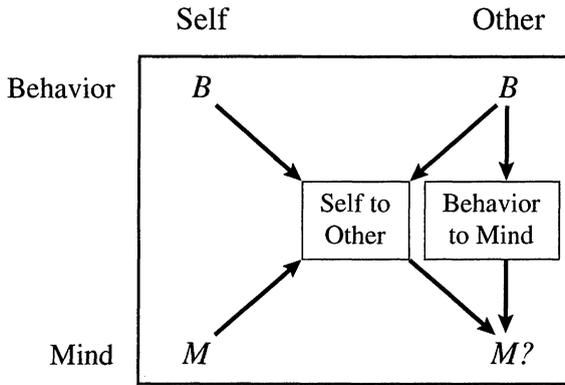
Figure 1

discriminate between empirically equivalent hypotheses, one also will hold that science is incapable of discriminating between such hypotheses when their subject matter is psychological. These are not special problems about psychology, however. What, then, becomes of the problem of other minds? If that problem concerns the tenability of Self-to-Other inference, it appears to be no problem, if science is able to draw behavior-to-mind inferences.

To see how the Self-to-Other problem can be detached from the problem of strictly third-person behavior-to-mind inference, we need to distinguish *absolute* from *incremental* versions of the Self-to-Other problem. The absolute problem concerns whether certain input information permits me to infer that the other person occupies mental state *M* rather than some alternative state *A*. As I have said, if information about the behavior of others permits me to infer that they are in mental state *M*, then it is hard to see why this inference should be *undermined* by adding the premise that I myself am in mental state *M* when I produce behavior *B*. The question remains, however, of whether first-person information makes a difference. To whatever degree third-person information provides an indication of whether the other person has *M* or *A*, does the addition of first-person information modify this assessment? This incremental version of the Self-to-Other problem is neutral on the question of whether third-person behavior-to-mind inference is possible. It is this incre-

mental problem which I think forms the core of the problem of other minds, and this is the problem which I want to address here.[2]

I

Although the problem of other minds usually begins with an introspective grasp of one's own mental state, it can be detached from that setting and formulated more generally as a problem about "extrapolation." Thus, we might begin with the assumption that human beings produce certain behaviors because they occupy particular mental states and ask whether this licenses the conclusion that members of other species that exhibit the behavior do so for the same reason. None of us knows just by introspection, however, that *all* human beings who produce a given behavior do so because they occupy some particular mental state. In fact, this formulation of the problem of other minds, in which it is detached from the concept of introspection, is usually what leads philosophers to conclude that inferences about other minds from one's own case are weak. The fact that I own a purple bow tie should not lead me to conclude that you do, too. This point about bow ties is supposed to carry over to the fact that I have a mind and occupy various mental states. I know that I own a purple bow tie, but not by introspection.[3]

Discussion of the problem of other minds in philosophy seems to have died down (if not out) around thirty years ago.[4] Before then, it was discussed as an instance of "analogical" or "inductive" reasoning and the standard objection was that an inference about others based on your own situation is an extrapolation from too small a sample. This problem does not disappear merely by thinking of introspective experience as

---

[2] There is a further reason for focusing on the incremental problem; it permits us to investigate the problem of other minds without formulating it as a problem about *acceptance.* The lottery paradox shows how difficult it is to say how much evidence in favor of a hypothesis is needed for one to believe the hypothesis. On this, see Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown, CT: Wesleyan, 1961).

[3] The fact that the problem of other minds can be detached from the idea of introspection does not entail that introspection has no special epistemic standing. Even if I can know things about my mind by pathways not open to you, the question remains of how I should extrapolate this self-knowledge to others. I should mention that I see no reason to think that the semantics of mentalistic terms entails that this epistemological problem cannot be posed intelligibly. Even if I fix the meaning of the word 'pain' by applying it to various experiences that I have had to date, it remains to be said whether the term also applies to others (or to myself at later dates).

[4] Perhaps one reason the problem largely lapsed from philosophical discussion is that it had become so closely connected with questions about logical behaviorism. When logical behaviorism went out of fashion, so did the problem of other minds. Behaviorism was replaced by a version of mentalism that emphasized the idea that third-person mentalistic hypotheses are to be judged, like other scientific hypotheses, on the basis of their ability to explain and predict. In consequence, the Self-to-Other problem was replaced by the behavior-to-mind problem.

furnishing you with thousands of data points. The fact remains that they all were drawn from the same urn—your own mind. How can sampling from one urn help you infer the composition of another?

When the problem is formulated in this way, it becomes pretty clear that what is needed is some basic guidance about inductive inference. The *mental content* of Self-to-Other inference is not what makes it problematic, but the fact that it involves *extrapolation*. Some extrapolations make sense while others do not. It seems sensible to say that thirst makes other people drink water, based on the fact that this is usually why I drink. Yet it seems silly to say that other folks walk down State Street at lunchtime because they crave spicy food, based just on the fact that this is what sets me strolling. By the same token, it seems sensible to attribute belly buttons to others, based on my own navel gazing. Yet it seems silly to universalize the fact that I happen to own a purple bow tie. What gives?

The first step toward answering this question started to emerge in the 1960s, not in philosophy of mind, but in philosophy of science. There is a general point about confirmation that we need to take to heart: observations provide evidence for or against a hypothesis only in the context of a set of background assumptions. If the observations do not deductively entail that the hypothesis of interest is true, or that it is false, then there is no saying whether the observations confirm or disconfirm, until further assumptions are put on the table. This may sound like the Duhem/Quine thesis, but that way of thinking about the present point is somewhat misleading, since Pierre Duhem[5] and W. V. Quine[6] discussed deductive, not probabilistic, connections of hypotheses to observations. If we want a person to pin this thesis to, it should be I. J. Good.[7] Good made this point forcefully in connection with Carl Hempel's[8] formulation of the ravens paradox. Hempel thought it was clear that black ravens and white shoes both confirm the generalization that all ravens are black. The question that interested him was why one should think that black ravens provide stronger confirmation than white shoes. Good responded by showing that empirical background knowledge can have the consequence that

---

[5] In *The Aim and Structure of Physical Theory* (Princeton: University Press, 1914/1954).

[6] In "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge: Harvard, 1953), pp. 20-46

[7] In "The White Shoe Is a Red Herring," *British Journal for the Philosophy of Science,* XVII (1967): 322, and in "The White Shoe Qua Herring Is Pink," *British Journal for the Philosophy of Science,* XIX (1968): 156-57.

[8] "Studies in the Logic of Confirmation," in *Aspects of Scientific Explanation and Other Essays* (New York: Free Press, 1965), and "The White Shoe: No Red Herring," *British Journal for the Philosophy of Science,* XVIII (1967): 239-40.

black ravens actually *dis*confirm the generalization. Hempel replied by granting that one could have special information that would undercut the assumption that black ravens and white shoes both confirm. He thought that in the absence of such information, however, it was a matter of logic that black ravens and white shoes are confirmatory. For this reason, Hempel asked the reader to indulge in a "methodological fiction." We are to imagine that we know nothing at all about the world, but are able to ascertain of individual objects what colors they have and whether they are ravens. We then are presented with a black raven and a white shoe; logic alone is supposed to tell us that these objects are confirming instances. Good's response was that a person in the circumstance described would not be able to say anything about the evidential meaning of the observations. I think that subsequent work on the concept of evidence, both in philosophy and in statistics, has made it abundantly clear that Good was right and Hempel was wrong. Confirmation is not a two-place relationship between observations and a hypothesis; it is a three-place relation between observations, a hypothesis, and a background theory.[9] Black ravens confirm the generalization that all ravens are black, given some background assumptions, but fail to do so, given others. And if no background assumptions can be brought to bear, the only thing one can say is: *out of nothing, nothing comes.*

If we apply this lesson to the problem of other minds, we obtain the following result: the fact that I usually or always have mental property *M* when I perform behavior *B* is, by itself, *no reason at all* to think that you usually or always have *M* when you perform *B*. If that sounds like skepticism, so be it. The inference from Self to Other can make sense, however, once additional background assumptions are stated. A nonskeptical solution to the problem of other minds, therefore, must identify plausible further assumptions that bridge the inferential gap between Self and Other.[10]

[9] In fact, there are many important circumstances in which the evidence relation must be four-placed—a set of observations favors one hypothesis over another, relative to a set of background assumptions. This is the proper format for likelihood inference; see Richard M. Royall, *Statistical Evidence: A Likelihood Paradigm* (Boca Raton: Chapman and Hall/CRC, 1997), and my "Testability," *Proceedings and Addresses of the American Philosophical Association*, LXXIII (1999): 47-76; the latter is also available at the following URL—http://philosophy.wisc.edu/sober.

[10] I am not going to discuss in this paper what it means for two individuals to exhibit "the same behavior," but I shall make two comments. First, there is no requirement that they exhibit the same muscular movements; on this point, see Berent Enç, "Units of Behavior," *Philosophy of Science*, LXII (1995): 523-42. Second, however behaviors are individuated, it is important that one be able to say that two individuals share a behavioral trait without already knowing what the proximate

II

The problem of other minds has been and continues to be important in psychology (actually, in comparative psychology), except that there it is formulated in the first-person plural. Suppose that when human beings perform behavior B, we usually or always do so because we have mental property M. When we observe behavior B in another species, should we take the human case to count as evidence that this species also has mental property M? For this problem to be nontrivial, we assume that there is at least one alternative internal mechanism, A, which also could lead organisms to produce the behavior. Is the fact that humans have M evidence that this other species has M rather than A?

Discussion of this problem in comparative psychology has long been dominated by the fear of naive anthropomorphism.[11] This attitude was crystalized by C. Lloyd Morgan,[12] who suggested that, if we can explain a nonhuman organism's behavior by attributing to it a "higher" mental faculty, or by assigning it a "lower" mental faculty, then we should prefer the latter explanation. Morgan's successors embraced this "canon" because of its prophylactic qualities—it reduces the chance of a certain type of error. It is important to recognize, however, that there are *two* types of error that might occur in this situation:

	O lacks M	O has M
Deny that O has M		type-2 error
Affirm that O has M	type-1 error	

Morgan's canon does reduce the chance of type-1 error, but that is not enough to justify the canon. By the same token, a principle that encouraged anthropomorphism would reduce the chance of type-2 error, but that would not justify this liberal principle, either.

---

mechanism is (cognitive or otherwise) that leads them to produce the behavior; otherwise the inference problem here addressed could not get off the ground; see my "Black Box Inference: When Should an Intervening Variable Be Postulated?" *British Journal for the Philosophy of Science*, XLIX (1998): 469-98.

[11] See Daniel C. Dennett, "Intentional Systems in Cognitive Psychology: The 'Panglossian Paradigm' Defended," in *The Intentional Stance* (Cambridge: MIT, 1989), pp. 237-68, and Colin Allen and Marc Bekoff, *Species of Mind: The Philosophy and Biology of Cognitive Ethology* (Cambridge: MIT, 1997).

[12] In *An Introduction to Comparative Psychology* (London: Walter Scott, 1903).

Morgan tried to give a deeper defense of his canon; he thought he could justify it on the basis of Darwin's theory of evolution. Although Morgan's interesting argument does not work,[13] it is noteworthy that Morgan explicitly rejects what has become a fairly standard view of his principle—that it is a version of the principle of parsimony. Morgan thought that the simplest hypothesis about nonhuman organisms is that they are just like us. Parsimony favors anthropomorphism; the point of the canon is to counteract this tendency of thought. As we shall see, Morgan's conception of the relationship of his canon to the principle of parsimony was prescient.

I now want to leave Morgan's late-nineteenth century ideas about comparative psychology behind, and fast forward to the cladistic revolution in evolutionary biology that occurred in the 1970s and after. The point of interest here is the use of a principle of phylogenetic parsimony to infer phylogenetic relationships among species, based on data concerning their similarities and differences.[14] Although philosophers often say that parsimony is an ill-defined concept, its meaning in the context of the problem of phylogenetic inference is pretty clear. The hypotheses under consideration specify phylogenetic trees. The most parsimonious tree is the one that requires the smallest number of changes in character state in its interior to produce the observed distribution of characteristics across species at the tree's tips.

Consider the problem of inferring how sparrows, robins, and crocodiles are related to each other. Two hypotheses that might be considered are depicted in Figure 2. The (SR)C hypothesis says that sparrows and robins have a common ancestor that is not an ancestor of crocs; the S(RC) hypothesis says that it is robins and crocs that are more closely related to each other than either is to sparrows. Now consider an observation—sparrows and robins have wings, while crocodiles do not. Which phylogenetic hypothesis is better supported by this observation?

If winglessness is the ancestral condition, then the (SR)C hypothesis is more parsimonious. This hypothesis can explain the data about tip taxa by postulating a single change in character state on the branch with a slash through it. The S(RC) hypothesis, on the other hand, must postulate at least two changes in character state to explain the data. The principle of cladistic parsimony says that the observations favor (SR)C over S(RC). Notice that the parsimoniousness of a

[13] For discussion, see my "Morgan's Canon," in Colin Allen and Denise Dellarosa Cummins, eds., *The Evolution of Mind* (New York: Oxford, 1998), pp. 224-42.

[14] See, for example, Niles Eldredge and Joel Cracraft, *Phylogenetic Patterns and the Evolutionary Process* (New York: Columbia, 1980).
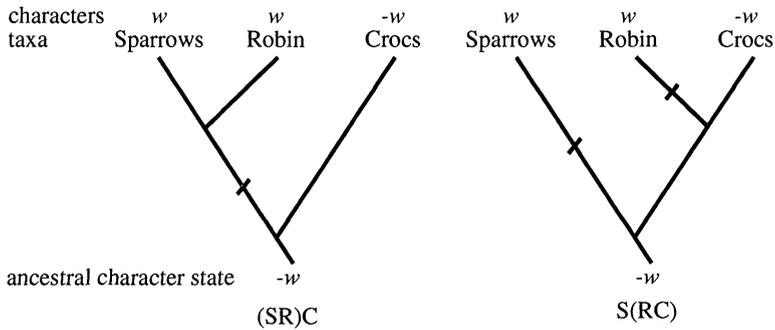
Figure 2

hypothesis is assessed not by seeing how many changes it says *actually* occurred, but by seeing what the *minimum* number of changes is that the hypothesis requires. (SR)C is more parsimonious because it entails a lower minimum.[15]

Although cladistic parsimony first attracted the attention of biologists because it helps one infer genealogical relationships, there is a second type of problem that parsimony allows one to address. If you use a set of traits to reconstruct the genealogy that connects several contemporaneous species (as in Figure 2), you can place a *new* set of traits on the tip species in that inferred tree and use parsimony to infer the character states of their ancestors. This inference is illustrated in Figure 3. Given the character states (1s and 0s) of tip species, the most parsimonious assignment of character states to interior nodes is the one shown.[16]

Parsimony is now regarded in evolutionary biology as a reasonable way to infer phylogenetic relationships. Whether it is the absolutely best method to use, in all circumstances, is rather more controversial. And the foundational assumptions that need to be in place for

[15] Cladistic parsimony does not entail that all similarities are evidence of common ancestry. For example, the two hypotheses depicted in Figure 2 would be equally parsimonious if wings were ancestral rather than derived. For discussion, see my *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge: MIT, 1988).

[16] Figure 3 illustrates a pattern of inference that is important in the study of human evolution. Why should features that are shared among current hunter-gatherer societies be thought to provide an indication of the ancestral human condition? After all, we cannot assume that hunter gatherers are "living fossils." The reason the inference makes sense is that current hunter gatherers are very distantly related to each other; this is what makes any similarities they may exhibit relevant to estimating the character state of the most recent common ancestor shared by all human lineages.
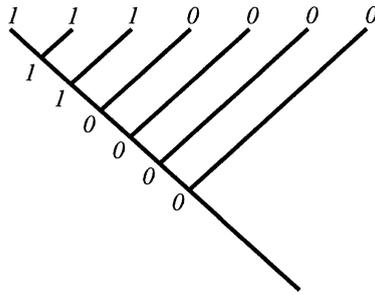
Figure 3

parsimony to make sense as an inferential criterion are also a matter of continuing investigation.

The reason I have explained the basic idea behind cladistic parsimony is that it applies to the problem of other minds, when Self and Other are genealogically related. Suppose both Self and Other are known to have behavioral characteristic B, and that Self is known to have mental characteristic M. The question is whether Other should be assigned M as well. As before, I shall assume that M is sufficient for B, but not necessary (an alternative internal mechanism, A, also could produce B). The two hypotheses we need to consider are depicted in Figure 4. If the root of the tree has the characteristic not-B (and so has neither-M-nor-A),[17] then the (Same) hypothesis is more parsimonious than the (Diff) hypothesis. It is consistent with (Same) that the postulated similarity linking Self and Other is a homology; it is possible that the most recent common ancestor of Self and Other had M, and that M was transmitted unchanged from this ancestor to the two descendants. The (Same) hypothesis, therefore, requires only a single change in character state, from neither-M-nor-A to M. In contrast, the (Diff) hypothesis requires at least two changes in character state.[18]

Frans De Waal[19] presents this cladistic argument in defense of the idea that parsimony favors anthropomorphism—we should prefer the

[17] We might have evidence that not-B is the ancestral character state by looking at a number of further individuals, besides Self and Other, who provide relevant "out-groups," as in Figure 3.

[18] Notice that, if no assumption is made about the character state at the root of the tree, (SAME) is more parsimonious. But if A is the ancestral character state, (SAME) and (DIFF) are equally parsimonious.

[19] In "Complementary Methods and Convergent Evidence in the Study of Primate Social Cognition," *Behaviour*, CXVIII (1991): 297-320, and in "Anthropomorphism and Anthropodenial: Consistency in Our Thinking about Humans and Other Animals," *Philosophical Topics* (forthcoming).
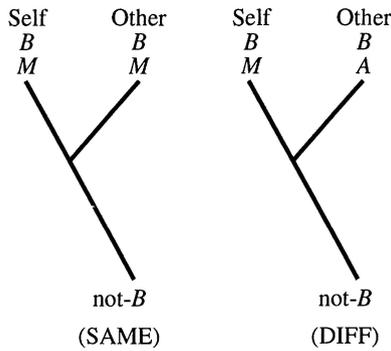
Figure 4

hypothesis that other species have the same mental characteristics that we have when they exhibit the same behavior.[20] This parsimony argument goes against Morgan's canon, just as Morgan foresaw. De Waal adds the reasonable proviso that the parsimony inference is strongest when Self and Other are closely related.

There is another proviso that needs to be added to this analysis, which is illustrated in Figure 5. As before, Self and Other are observed to have behavioral trait B. We know by assumption that Self has the mental trait M. The question, as before, is what we should infer about Other: Does it have M or A? The new wrinkle is that there are additional species depicted in the tree, ones that are known to lack B. The genealogical relationships that connect these further species to Self and Other entail that the most parsimonious hypothesis is that B evolved twice. It now makes no difference in parsimony whether one thinks that Other has M or A. What this shows is that parsimony favors anthropomorphism about mentalistic properties only when the behaviors in question are thought to be homologous.

This point has implications about a kind of question that often arises in connection with sociobiology. Parsimony does not oblige us to think that "slave-making" in social insects has the same psychosocial causes as slave-making in humans (or that rape in human beings has the same proximate mechanism as "rape" in ducks). The *behaviors* are not homologous, so there is no argument from parsimony for thinking that the same *proximate mechanisms* are at work. This is a point in favor of the parsimony analysis—cladistic parsimony explains why certain types of implausible inference really are implausi-

─────────

[20] See also my "Morgan's Canon"; and Shapiro, "Adapted Minds" (unpublished).
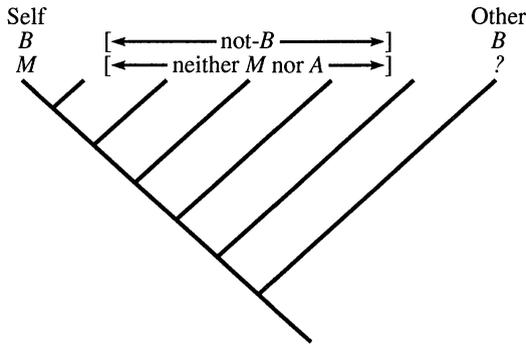
Figure 5

ble. *Parsimonious* anthropomorphism is not the same as *naive* anthropomorphism.

Just as extrapolation from Self to Other can be undermined by behavioral information about other species, extrapolation also can be undermined by neurophysiological information about Self and Other themselves. If Self has mental state M by virtue of being in physical state P1, what are we to make of the discovery that Other has physical state P2 (where P1 and P2 are mutually exclusive)? If we accept functionalism's assurance that mental state M is multiply realizable, this information does not entail that Other cannot be in mental state M; after all, P2 might just be a second supervenience base for M. On the other hand, P2 might be a supervenience base for alternative state A, and not for M at all. This inference problem is illustrated in Figure 6. For simplicity, let us suppose that P1 and P2 each suffice for B. P1 produces B by way of mental state M; it is not known at the outset whether P2 produces B by way of mental state M, or by way of alternative state A. We assume that the ancestor at the root of the tree did not exhibit the behavior in question, and that this ancestor had trait P0.

The conclusion we must reach is that the two hypotheses in Figure 6 are equally parsimonious. (Same) says that Self and Other both have M and that P1 and P2 are two of M's supervenience bases. (Same) requires two changes to account for the characteristics at the tips; either P1 or P2 evolved long ago from an ancestor who had P0, and then either P1 changed to P2, or P2 changed to P1, in a subsequent branch of the tree. Of course, not-B changed to B in the same branch in which P0 changed to either P1 or P2, but this does not count as *two* changes, since the first entails the second. The (Diff) hypothesis also requires two changes if the root of the tree's having
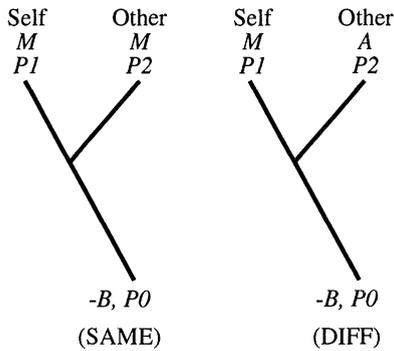
Figure 6

P0 is to be transformed into Self's having P1 and Other's having P2. Of course, (Diff) also will require changes from neither-*M*-nor-*A* at the root to *M* and *A* at the tips, but these do not count as changes *additional* to the ones that occur among P0, P1, and P2. Thus, the discovery of relevant neurophysiological differences can provide a context in which anthropomorphism is not sanctioned by parsimony considerations.

As a final exercise in understanding how the machinery of phylogenetic parsimony applies to the problem of extrapolating from Self to Other, let us consider the current controversy in cognitive ethology concerning whether chimps have a theory of mind.[21] It is assumed at the outset that chimps have what Daniel C. Dennett (*op. cit.*) terms first-order intentionality; they are able to formulate beliefs and desires about the extra-mental objects in their environment. The question under investigation is whether they, in addition, have second-order intentionality. Do they have the ability to formulate beliefs and desires about the mental states of Self and Other? Adult human beings have both first- and second-order intentionality. Do chimps have both, or do they have first-order intentionality only?

Figure 7 provides a cladistic representation of this question. I assume that the ancestral condition is the absence of both types of intentionality, and that second-order intentionality can evolve in a lineage only if first-order intentionality is already in place. The point to notice is that parsimony considerations do not discriminate between the two hypotheses. The (SAME) hypothesis and the (DIFF)

---

[21] See the essays collected in Peter Carruthers and Peter K. Smith, eds., *Theories of Theories of Mind* (New York: Cambridge, 1995), and C. M. Heyes, "Theory of Mind in Nonhuman Primates," *Behavioral and Brain Sciences,* XXI (1998): 101-48.

species	humans	chimps	humans	chimps
first-order intentionality	yes	yes	yes	yes
second-order intentionality	yes	yes	yes	no

(SAME)   (DIFF)
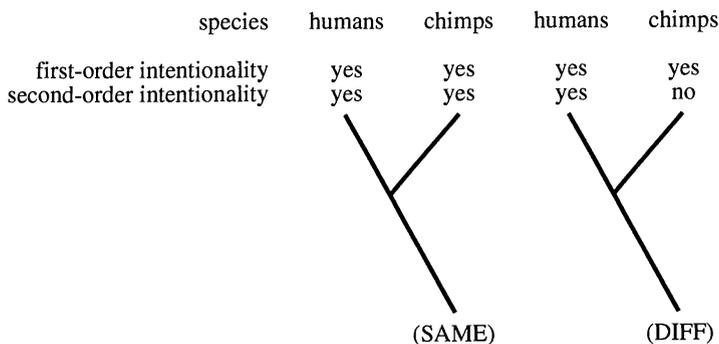
Figure 7

hypothesis both require two changes in the tree's interior—first- and second-order intentionality each must evolve at least once. The principle of parsimony in this instance tells us to be agnostic, and so disagrees with Morgan's canon, which tells us to prefer (DIFF).

Why does the problem formulated in Figure 7 lead to a different conclusion from the problem represented in Figure 4 in terms of the two states M and A? In Figure 4, we find that parsimony considerations favor extrapolation; in Figure 7, we find that extrapolating is no more and no less parsimonious than not extrapolating. Notice that Figure 4 mentions the behavior B, which internal mechanisms M and A are each able to produce. Figure 7, however, makes no mention of the behaviors that first- and second-order intentionality underwrite. This is the key to understanding why the analyses come out differently.

The point behind Figure 4 is this: if two species exhibit a homologous behavior, each of them must also possess an internal mechanism (M or A) for producing the behavior. It is more parsimonious to attribute the same mechanism to both Self and Other than to attribute different mechanisms to each. This pattern of argument can be applied to the problem depicted in Figure 7 by making explicit the behavioral consequences that first- and second-order intentionality are supposed to have. Suppose that first-order intentionality allows the organisms that have it to exhibit a range of behaviors B1. When a species that has first-order intentionality evolves second-order intentionality, this presumably augments the behaviors that organisms in the species are able to produce; the repertoire expands to

B1&B2.[22] If human beings exhibit B1 because they have first-order intentionality, we parsimoniously explain the fact that chimps also exhibit B1 (if they do) by saying that chimps have first-order intentionality. But there is no *additional* gain in parsimony to be had from attributing second-order intentionality to them as well, unless we also observe them producing the behaviors in B2.[23]

Cognitive ethologists are trying to find behaviors that chimps will produce if they have second-order intentionality but will not produce if they possess first-order intentionality only.[24] Identifying behaviors of this sort would permit an empirical test to be run concerning whether chimps have a theory of mind. It certainly is desirable that such behaviors should be found. Even if chimps exhibit B2, however, the question arises of why we should explain this by attributing second-order intentionality to them, rather than some alternative mechanism A. Of course, A will not be the trait of purely first-order intentionality, but presumably there are more than two options to be considered here. It is at this point that the principle of parsimony makes its entrance; if chimps exhibit behavior B2, this is better explained by the hypothesis that they have second-order intentionality than by the hypothesis that they have alternative mechanism A.

III

Why should we trust the parsimony arguments just described? Is parsimony an inferential end in itself or is there some deeper justification for taking parsimony seriously? In the present circumstance

---

[22] Some of the most interesting experiments on whether chimps have a theory of mind have been carried out by Daniel Povinelli; he looks for a B2 and finds that it is absent. I discuss his "knower/guesser" experiment in "Black Box Inference." More recently, Povinelli and Steve Gambrone have argued that "a novel psychological system for generating and sustaining higher-order representations, including the representation of other minds, may have evolved in the human lineage without *radically* altering our *basic* behavior patterns [italics mine]"—see their "Inferring Other Minds: Failure of the Argument by Analogy," *Philosophical Topics* (forthcoming). The words I have italicized in this quotation are important; the authors do not deny that there are behavioral differences between present day human beings and chimps that reflect the fact that the former have a theory of mind while the latter do not. Their suggestion is that the *initial appearance* of second-ordering intentionality may have had little or no immediate effect on behavior, but may have been a *building block* that allowed more substantial behavioral divergence to occur later.

[23] Suppose there is a behavior that human beings sometimes produce by using first-order intentionality only and sometimes by using second-order intentionality. What inference should we draw if we observe chimps producing this behavior? Again, there is no difference in parsimony between (SAME) and (DIFF).

[24] In fact, a behavioral test need not take this extreme form. It would suffice to find behaviors that are *probable* if subjects have a theory of mind but *improbable* if they have first-order intentionality only. The behavior need not be *impossible* in the absence of a theory of mind.

at least, there is no need to regard parsimony as something that we seek for its own sake.[25] My suggestion is that parsimony matters in problems of phylogenetic inference only to the extent that it reflects *likelihood*.[26] Here, I am using the term 'likelihood' in the technical sense introduced by R. A. Fisher. The likelihood of a hypothesis is the probability it confers on the observations, not the probability that the observations confer on the hypothesis. The likelihood of H, relative to the data, is $\Pr(\text{Data} \mid H)$, not $\Pr(H \mid \text{Data})$.

To see how the likelihood concept can be brought to bear on cladistic parsimony, consider Figure 2. It can be shown, given some minimal assumptions about the evolutionary process, that the data depicted in that figure are made more probable by the (RS)C hypothesis than they are by the R(SC) hypothesis. These assumptions are as follows:

Heritability: character states of ancestors and descendants are positively correlated; that is, $\Pr(\text{Descendant has a wing} \mid \text{Ancestor has a wing}) > \Pr(\text{Descendant has a wing} \mid \text{Ancestor lacks a wing})$.

Chance: all probabilities are strictly between 0 and 1.

Screening-off: lineages evolve independently of each other, once they branch off from their most recent common ancestor.

What we have here, in its essentials, is a proof that Hans Reichenbach[27] gave in connection with his "principle of the common cause."

I think that most biologists would agree that these three assumptions hold pretty generally. The first of them does *not* say that descendants probably end up resembling their ancestors. The claim is *not* that stasis is more probable than change—that $\Pr(\text{Descendant has a wing} \mid \text{Ancestor has a wing}) > \Pr(\text{Descendant lacks a wing} \mid \text{Ancestor has a wing})$. Rather, the claim is that, if a descendant has a wing, that this result would have been more probable if its ancestor had had a wing than it would have been if its ancestor had lacked a wing.[28] The last assumption, it should be noted, is not exceptionless;

---

[25] The same point can be made in connection with the use of a parsimony criterion in problems of model selection, including curve-fitting problems; see Malcolm Forster and my "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions," *British Journal for the Philosophy of Science*, XLV (1994): 1-36.

[26] I defend this suggestion in my *Reconstructing the Past*.

[27] In *The Direction of Time* (Berkeley: California UP, 1956). For discussion, see chapter 3 of my *Reconstructing the Past*.

[28] The concept of heritability used here is closer to *narrow sense* heritability than to *broad sense* heritability; see D. S. Falconer and Trudy F. C. Mackay, *An Introduction to Quantitative Genetics* (London: Longmans, 1996, 4th ed.). It neither implies nor is implied by the idea of *genetic determination*. A trait can be heritable and still be

after all, there are ecological circumstances in which a lineage's evolving a trait influences the probability that other contemporaneous lineages will do the same. But even here, the assumption of independence is often true; and when it is false, it usually can be weakened without materially affecting the qualitative conclusions I want to draw.

Not only does likelihood provide a framework for understanding the role of parsimony considerations in phylogenetic inference; it also has implications about the Self and Other problem depicted in Figure 4. If traits M and A obey the assumptions listed, the following inequality is a consequence:

(P) Pr(Self has M | Other has M) > Pr(Self has M | Other has A).

This proposition says that Self and Other are correlated (nonindependent) with respect to the traits M and A. It also says that there is a likelihood justification for anthropomorphism.[29] The observation that Self has M is rendered more probable by the hypothesis that Other has M than by the hypothesis that Other has A. Such differences in likelihood are generally taken to indicate a difference in support—the observation *favors* the first hypothesis over the second.[30]

The likelihood concept also throws light on De Waal's proviso—that parsimonious anthropomorphism is on firmer ground for our near relatives than it is for those individuals to whom we are related more distantly. We may translate this into the claim that the two

---

influenced by the environment. And a trait can be genetically determined and still not be heritable; the traits male (XY) and female (XX) provide examples.

[29] Hybrids aside, two species have a unique species that is their most recent common ancestor. Sexually reproducing individuals are not like this, however. Full-sibs have *two* parents as their most recent common ancestors. And first cousins usually overlap only partially in the set of ancestors they have two generations back; each has four grandparents, but (usually) only two of them are shared. How, then, does the Reichenbachian picture of common causes apply to human genealogies? The simplest way to connect them is to think of the common causes as sets of individuals, not as singletons. Thus, two full-sibs have the same parental *pair* as a common cause. And two first cousins can be thought of as tracing back to a set of six individuals; this is the set of all their grandparents, including the ones they do not share. Standard Mendelian genetics assures us that the state of this set screens off one cousin's genotype from the other's. Of course, this set is not limited to the two cousins' *common* ancestors. It can be shown that the common ancestors do screen off, however, since the ancestors in the set who are not shared influence one cousin's genotype but not the other's; see my and Martin Barrett's "Conjunctive Forks and Temporally Asymmetric Inference," *Australasian Journal of Philosophy*, LXX (1992): 1-23.

[30] See Royall; and my "Testability."
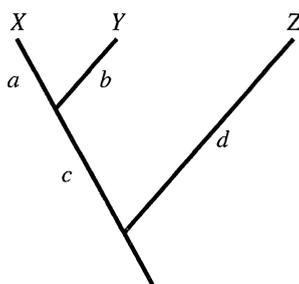
Figure 8

probabilities compared in proposition (P) are more different when Self and Other are closely related and become more similar as the relationship becomes more distant. This thesis is illustrated by Figure 8, in which X and Y are more closely related to each other than either is to Z. The lower-case letters in the tree's interior are path coefficients, which entail that the correlation of X and Y (r_{XY}) and the correlation of X and Z (r_{XZ}) have the values $r_{XY} = ab$ and $r_{XZ} = acd$. Notice that $r_{XY} > r_{XZ}$ if and only if $b > cd$. This inequality need not be true, but it does follow from an assumption that often figures in evolutionary discussions. This is the assumption of *uniform rates*—that a given evolutionary event has the same probability of occurring in different contemporaneous branches of a tree. In the present example, uniform rates means that $a = b$ and $d = ac$, which suffices to insure that X and Y are more strongly correlated than are X and Z. Thus, De Waal's proviso[31] is not true unconditionally; but when it is true, it need not be added as an independent constraint on parsimony arguments (which have no machinery for taking account of recency of divergence). The proviso flows from a likelihood analysis.

[31] I hope it is clear that this claim does not underwrite racist, nationalist, or species-ist conclusions. The argument does not provide a reason for denying that individuals outside one's own "group" have minds. For one thing, it is a mistake to think of one's self as belonging to just one group; each organism belongs to multiple, nested groups. For another, the argument presented here does not concern acceptance or rejection. And, finally, it is important to remember that Self-to-Other inference is not the only pathway by which we form beliefs about the internal states of others. There is, in addition, the possibility of strictly third-person behavior-to-mind inference. Much of our confidence concerning the mental states of others presumably rests on this second sort of inference. The point I am making about the *incremental* Self-to-Other problem is that the increment provided by knowledge of Self falls off as genealogical relatedness becomes more distant.

IV

Genealogical relatedness suffices to justify a likelihood inference from Self to Other, if the Reichenbachian assumptions that I described hold true. But is genealogical relatedness *necessary* for this extrapolation to make sense? Proposition (P) says that Self and Other are correlated with respect to trait M. What could induce this correlation? Reichenbach argued that whenever there is a correlation of two events, either the one causes the other, or the other causes the one, or the two trace back to a common cause. Considerations from quantum mechanics suggest that this is not always the case,[32] and doubts about Reichenbach's principle can arise from a purely classical point of view as well.[33] If we are not prepared to suppose that mentalistic correlations between Self and Other are brute facts, however, and if Self's having M does not causally influence whether Other has M (or vice versa), then Reichenbach's conclusion seems reasonable, if not apodictic; if Self and Other are correlated, this should be understood as arising from a common cause.

Genealogical relatedness is one type of common-cause structure. It can induce the correlation described in (P) by having ancestors transmit genes to their descendants, but there are alternatives that we need to recognize. To begin with, parents exert *non*genetic influences on their offspring through teaching and learning. For example, children have a higher probability of speaking Korean if their parents speak Korean than if their parents do not, but this is not because there is a gene for speaking Korean. And there are nongenetic connections between parents and offspring that do not involve learning, as when a mother transmits immunity to her children through her breast milk.

Correlations between Self and Other also can be induced by common causes when Self and Other are not genealogically related. If students resemble their teachers, then students of the same teacher will resemble each other. Here, learning does the work that genetic transmission is also able to do. Similarly, Self and Other can be correlated when they are influenced by a common environmental cause that requires no learning. For example, if influenza is spreading through one community, but not through another, then the fact

---

[32] See Bas van Fraassen, "The Charybdis of Realism: Epistemological Implications of Bell's Inequality," *Synthese*, LII (1982): 25-38.

[33] See van Fraassen, "Rational Belief and the Common Cause Principle," in Robert McLaughlin, ed., *What? Where? When? Why? Essays in Honor of Wesley Salmon* (Boston: Reidel, 1982), pp. 193-209; Nancy Cartwright, *Nature's Capacities and Their Measurement* (New York: Oxford, 1989); and my *Reconstructing the Past*, chapter 3.
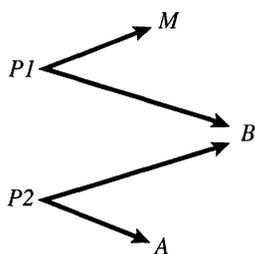
Figure 9

that I have the flu can be evidence that Other does, too, if the two of us live in the same community.

I list these alternatives, not because they apply with equal plausibility to the problem of other minds, but to give an indication of the range of alternatives which needs to be considered. Genealogical relatedness is only an example; the fundamental question is whether there are common causes impinging on Self and Other that induce the correlation described in (P). If there are, then there will be a likelihood justification for extrapolating from Self to Other.[34]

V

What, exactly, does cladistic parsimony and its likelihood analysis tell us about the problem of other minds? When I cry out, wince, and remove my body from an object inflicting tissue damage, this is (usually) because I am experiencing pain. When other organisms (human or not) produce the same set of behaviors, is this evidence that they feel pain? This hypothesis about Other *is* more parsimonious (if the behaviors are homologous and there are no known relevant neurophysiological differences), and it *is* more likely (if Reichenbachian assumptions about common causes apply). Does that completely solve the problem of other minds, or does there remain a residue of puzzlement?

One thing that is missing from this analysis is an answer to the question—*how much* evidence does the introspected state of Self provide about the conjectured state of Other? I have noted that the likelihoods in proposition (P) become more different as Self and

[34] Although I have described M and A as possible causes of the behavior B, this is not essential for the parsimony or likelihood arguments I have presented. Suppose that M and A are *epiphenomenal consequences* of the physical states P1 and P2, and are related to the behavior B as shown in Figure 9. If P1 suffices for M and B while P2 suffices for B and A, parsimony and likelihood are relevant to deciding whether M or A should be attributed to Other, given that Self has M.

Other become more closely related, but this comparative remark does not entail any quantitative benchmarks. It is left open whether the observation *strongly* favors one hypothesis over the other, or does so only *weakly*.

Another detail that I have not addressed is how probable it is that Other has *M*. If I have *M* when I produce behavior *B*, and Other exhibits *B*, is the probability greater than ½ that Other has M as well? The previous discussion helps answer that question, in that principle (P) is equivalent to the claim that Self's having *M* raises the probability that Other has *M*. Whatever the prior probability is that Other has *M*, the posterior probability is greater. Whether additional information can be provided that allows the value of that posterior probability to be estimated is a separate question.

A genealogical perspective on the problem of other minds helps clarify how that problem differs from the behavior-to-mind problem discussed at the outset. At first glance, it might appear that it does not matter to the problem of other minds whether the other individual considered is a human being, a dog, an extraterrestrial, or a computer. In all these cases, the question can be posed as to whether knowledge of one's own case permits extrapolation to another system that is behaving similarly. We have seen that these different formulations receive different answers. I share ancestors with other human beings, and I share other, more remote, ancestors with the nonhuman organisms found on earth. If there are creatures on other planets that evolved independently of life on earth, however, then I share no common ancestors with them. In this case, extrapolation from Self to Other does not have the genealogical justification I have described. I would go further and speculate that *it has no justification at all.* This does not mean that we should never attribute mental states to such creatures. What it does mean is that we must approach such questions as instances of the purely third-person behavior-to-mind inference problem. Similar remarks also may apply to computers. When they behave similarly to us (perhaps by passing an appropriate Turing test or by playing a competent game of chess), we may ask what causes them to do so. We have no ancestors in common with them; rather, we have constructed them so that they produce certain behaviors. Does this fact about the design of computers provide a reason to think that the proximate mechanisms behind those behaviors resemble those found in human beings? Arguably not. For extraterrestrials and (arguably) for computers, extrapolation from one's own case will not be justified. Indeed, this point applies to organisms with whom we do share ancestors, if the behaviors we have in common with them are not homologies (Figure 5). And even when Self and Other do exhibit a behavioral homology, if Self and Other are known to deploy

different neural machinery for exhibiting that behavior, the extrapola-
tion of *M* from Self to Other is undermined (Figure 6).

A probabilistic representation of the problem of other minds shows
that the usual objection to extrapolating from Self to Other is in fact
irrelevant, or, more charitably, that it rests on a factual assumption about
the world that we have no reason to believe. The question is not whether
introspected information about one's own mind provides lots of data or
only a little. Rather, the issue is how strong the correlation is between
Self and Other. Consider two urns that are filled with balls; each ball has
a color, but the frequencies of different colors in the urns are unknown.
The urns may be similar or identical in their compositions, or they may
be very different. If I sample one ball from the first urn and find that it
is green, does this provide substantial evidence concerning the second
urn's composition? If I sample a thousand balls from the first urn, does
this allow me to say any more about the second? Everything depends on
how the two urns are related. If they are *independent*, then samples drawn
from the first, whether they are small or large, provide no information
about the second. But if they are not independent, then even a small
sample from the first may be informative with respect to the second.
There is nothing wrong with *asking* whether knowledge of Self supports
a conclusion about Other. But the skeptical *assertion* that it does not
involves a factual claim about the world. A claim of independence is no
more a priori than a claim of correlation. If the relevant mental states of
Self and Other are joint effects of a common cause (with the properties
that Reichenbach described), then the skeptical assertion is false.

The argument I have presented is intended to show how certain
propositions are justified; I have not addressed the question of whether
people are justified in believing this or that proposition. My own
mental state can be an indicator of the mental states of others,
whether or not I know that this is true, or understand why it is true.
But what is required for people to be justified in extrapolating from
Self to Other? Must they know the fine points of phylogenetic parsi-
mony or of likelihood analysis? Or does it suffice that a given extrap-
olation is sanctioned by parsimony or likelihood considerations? This
is the epistemological thicket in which internalist and externalist
views of justification are in contention.[35] I shall not try to evaluate
these different approaches to the concept of justification, nor to say
in any detail how they are related to the ideas I have developed in this

[35] For discussion, see Laurence BonJour, "Externalism/Internalism," in Jonathan
Dancy and Ernest Sosa, eds., *A Companion to Epistemology* (Cambridge: Blackwell,
1992).

paper, but I do want to explore one line of questioning that arises from an internalist point of view.

Is it possible for me to figure out that proposition (P) is true, if I do not already know whether Other has M or A? Surely, I can. I can tell whether M is heritable by looking at still other individuals who are known to have either M or A and see how they are related genealogically. But is it possible to determine that (P) is true without knowing anything about which individuals (other than one's self) have M and which have A? Without this information, how can I tell whether the traits are heritable?[36] Well, knowledge in some strong philosophical sense is probably unnecessary, but perhaps judgments about heritability require one to have reasonable opinions, however tentative, about which individuals have which traits. Even so, the solution to the problem of other minds that I have suggested would not be undermined. The incremental version of the problem, recall, asks whether knowledge of my own case should make a difference in the characteristics I attribute to others. It is not required that I conceive of myself as beginning with no knowledge at all concerning the internal states of others.

We learned from Good that there is no saying whether a black raven confirms the generalization that all ravens are black unless one is prepared to make substantive background assumptions. The mere observation that the object before you is a black raven is not enough. The same point, applied to the problem of other minds, is that the mere observation that Self and Other share B and that Self has M is not enough. Further assumptions are needed to say whether these observations confirm the hypothesis that Other has M. Recognizing this point in the ravens paradox does not lead inevitably to skepticism, and it should not have that effect in the case of Self-to-Other inference. The problem of other minds should not be shackled with the "methodological fiction" that Hempel imposed on the ravens paradox. When the fetters are broken, the problem of other minds turns into the problem of searching out common causes.

<div align="right">ELLIOTT SOBER</div>

University of Wisconsin/Madison
London School of Economics and Political Science

---

[36] To be sure, it is possible to tell whether the behavior B is heritable (since B is observable), but the heritability of B is neither necessary nor sufficient for the heritability of M and A. I thank Branden Fitelson and Richard Lewontin for helping me clarify this point.