

# *The Evolution of Mind*

Chapter 9, Pages 224-242

"Morgan's Canon"

Elliott Sober

*Edited by*

DENISE DELLAROSA CUMMINS

AND COLIN ALLEN

1998

## 9

## Morgan's Canon

ELLIOTT SOBER

When an ant in the species *Solenopsis saevissima* dies, its fellow workers carry the body out of the nest to discard it. Why do they do so? One possibility is that the workers believe that the immobile organism is dead, and they have the desire to rid the nest of dead individuals. A better explanation is provided by the fact that the dead ant exudes oleic acid. Workers are disposed to pick up and discard anything that smells of this compound, even living ants that perverse biologists have daubed with the tell-tale substance (Wilson, 1971, p. 279).

When a piping plover (*Charadrius melodus*) sees a predator approach its nest, it will produce a broken wing display, dragging its intact wing along the sand as if the wing were injured (Ristau, 1991). Why does the bird do this? One possibility is that the plover wants to protect its young and believes that the display will induce a false belief in the predator. A better explanation is that the plover wants to protect her young, and believes that the broken wing display will have that effect.

In the first of these examples, a nonmentalistic explanation is preferable to a mentalistic explanation; the hypothesis that the ants are engaging in a fixed action pattern, triggered by a chemical trace, seems more plausible than the hypothesis that they are acting on the basis of beliefs and desires. In the second example, we prefer to explain the plover's behavior by attributing to it a belief about its own action, rather than by attributing to it a belief about the mind of the predator it confronts. Both these inferences obey a principle that C. Lloyd Morgan espoused in his *Introduction to Comparative Psychology*: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Morgan, 1894, p. 53).<sup>1</sup> Morgan called this his canon, and

the name has stuck. In this paper, I want to explore what the canon says and whether it is justified.

The two examples I have sketched (both drawn from Dennett, 1987) may make Morgan's principle seem self-evident. However, there are two considerations, each of them biological in character, that suggest that the canon is less than transparently obvious.

First, there is Morgan's use of the terms "higher" and "lower." Morgan thought of himself as developing an evolutionary perspective on the explanation of behavior. Indeed, he believed that his canon depends on specifically Darwinian ideas, not on a general maxim that bids us prefer simpler theories (Morgan, 1894, p. 54). Yet, when we consult the words of the master, we find Darwin writing himself a memo to never say "higher" and "lower" (Ghiselin, 1969, p. 70). This was no mere passing scribble, but characterized a deep and enduring implication of the hypotheses that Darwin developed. The theory of evolution by natural selection undermines the idea of a linear scale of nature in which each stage is either higher or lower than every other. Darwin replaced the ladder with the tree; lineages diverge from each other and develop adaptations that suit them to their peculiar conditions of life. In this framework, it makes no sense to ask whether the ant's use of pheromones is higher or lower than the plover's use of deception.<sup>2</sup> Both have their place in life's diversity.

These worries about the concepts of *higher* and *lower* are not assuaged by what Morgan says. Although Morgan was less than totally explicit about what these concepts mean, it is clear that he was very much in the grip of the Spencerian doctrine that evolution always marches from simple to complex (Gottlieb, 1979, p. 150; Boakes, 1984, p. 40). To the degree that Morgan's canon depends on this claim of directionality, the canon is in trouble. Although life started simple, and thus had to show a net increase in complexity (it had nowhere to go but up), the history of life is peppered with cases of evolutionary simplification. For example, the evolution of parasites typically involves a transition from complex to simple, as the parasite loses features of its free-living ancestor. So the first reason to pause over Morgan's canon is his use of the concepts of higher and lower. How should they be understood; and, once they are clarified, how justified does the canon turn out to be?

A second reason for wondering about the correctness of Morgan's canon is provided by patterns of reasoning frequently used in comparative biology. Comparative biologists attempt to infer the phylogenetic relationships that different species bear to each other from data concerning their similarities and differences. They also use hypotheses about phylogenetic relationship to help infer the characteristics of extinct species that cannot be ascertained by direct observation. These biologists usually do not think about psychological characteristics; more commonly, they focus on morphological, physiological, genetic, and molecular traits. Nonetheless, their modes of reasoning can be applied perfectly well to psychological features. I suggest that if Morgan's canon makes sense, it should make sense for any trait, whether the trait is psychological or not. For example, if an organism

is able to digest a particular protein, we may ask what internal mechanism enables it to do so. Morgan tells us to prefer a hypothesis that attributes a lower digestive mechanism over a hypothesis that attributes a higher mechanism. Comparative psychology is a branch of comparative biology.

With this point in mind, let's consider the inference problem depicted in Figure 9.1. We know by observation that the eggs of modern turtles, birds, and mammals have amniotic sacs, whereas the eggs of modern amphibians and fish do not. Now let's consider the protorothyrids, a group of small fossil reptiles thought to be closely related to the group that includes turtles, birds, and mammals (Carroll, 1988, pp. 192–198). Since fossils preserve hard parts and nothing else, we cannot tell by observation whether protorothyrids possessed an amnion. Should we attribute an amniotic sac to them, or not? Morgan's canon apparently says that we should decline to do so. If having an amniotic sac is "higher" and lacking a sac is "lower," and if our observations do not decide between these two hypotheses, then the canon evidently recommends the conclusion that the fossil resembled fish and amphibia in this regard.

Many comparative biologists would find this preference for the lower hypothesis less than compelling. Systematists who call themselves "cladists" would evaluate the hypotheses by applying considerations of parsimony, which

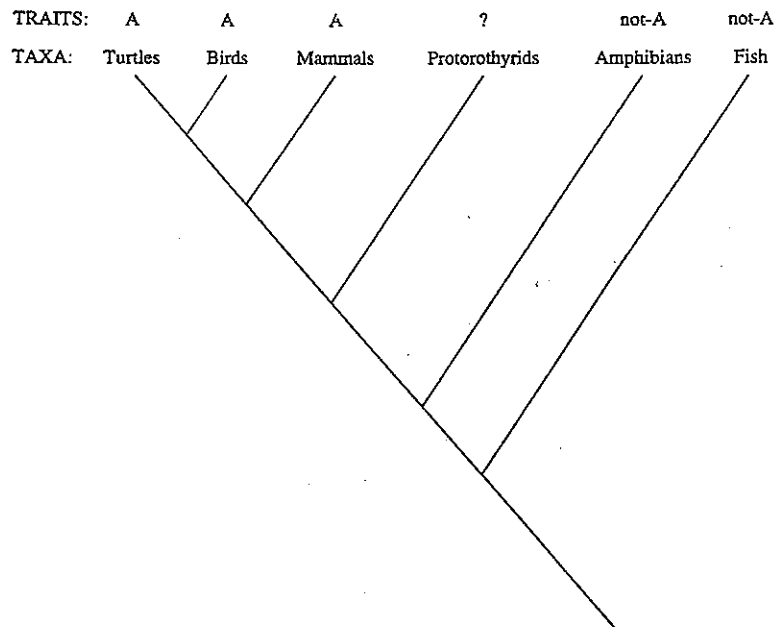


Figure 9.1 The protorothyrids are closely related to turtles, birds, and mammals. Do protorothyrids have amnions, or not?

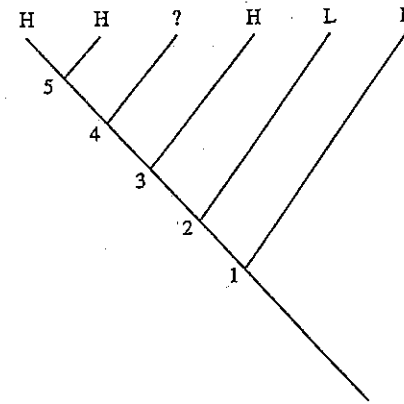


Figure 9.2 The question-mark taxon falls within a genealogical group whose other members are known to have trait *H*. What character state should this taxon be assigned?

they measure by counting the number of independent originations of traits that a hypothesis demands.<sup>3</sup> Notice that the tree depicted in Figure 9.1 requires only a single origination in the trait in question, regardless of whether protorothyrids are assigned an amnion or not. In this example, cladistic parsimony is quite indifferent to the distinction between higher and lower.

Many systematists would not wish to let the question about whether protorothyrids have amnions go at that. Even though cladistic parsimony is an important inferential tool, it is not the only game in town. Here is a fairly conventional piece of reasoning that biologists use to solve this problem: the amnion is thought to have evolved because it helps prevent eggs from drying out. The sac allows organisms to lay eggs on land, thus avoiding an aquatic larval stage. The fossils of protorothyrids suggest that they were land-dwelling and so they probably laid their eggs on land. So, in all likelihood, protorothyrid eggs had amnions. The point of this example is that biologists assign a character state to protorothyrids by considering specific biological details; they do not invoke a general preference for "lower" over "higher."

Figure 9.2 depicts a different problem. Once again, we are asked to assign a character state to a taxon whose phylogenetic relationship to other taxa is known. However, in this instance, the question mark taxon falls within a group that is otherwise characterized by the "higher" character state *H*.<sup>4</sup> The most parsimonious assignment of character states to the ancestors represented as interior nodes in this phylogenetic tree is 1=*L*, 2=*L*, 3=*H*, 4=*H*, and 5=*H*. This set of assignments requires that the *H* character state evolved only once. Parsimony favors assigning *H* to the question mark taxon as well. In this problem, the principle of cladistic parsimony leads us to prefer "higher" over "lower" because the target taxon belongs to a phylogenetic group whose other members exhibit the higher character state.

The two inferences just discussed differ in structure and in the types of biological information that are relevant. Nonetheless, they have something

in common. With biological information of the requisite type, a character state can be assigned to the target taxon. This assignment is made in accordance with principles that embody no particular preference for "lower" over "higher." But suppose that such biological details were not available in the inference problems just described. It is here that Morgan's principle might be expected to play the role of a tie-breaker—of discriminating between hypotheses that fit the available evidence equally well. However, when no biological evidence is available, I would expect many biologists to remain agnostic about the character state of the question mark taxon. Morgan's canon reaches the wrong conclusion in the two problems I just sketched when biological details are available; in similar problems in which biological details are *not* available, the canon seems to license inferences in which the proper conclusion is that no inference can be drawn.

These critical remarks are not meant to show that Morgan's canon is mistaken. Rather, the point so far is that the canon presents a puzzle. The problem is to understand what "higher" and "lower" mean, and to see why it makes sense to favor lower over higher. At the end of this paper, I'll try to construct a justification of Morgan's idea. But before I get to that, I want to describe the role that the canon played in Morgan's own thought and also examine in detail the justifying argument that Morgan himself constructed. Morgan's canon has had a long and influential history in psychology. As the examples at the beginning of this paper suggest, the canon was a weapon that behaviorists used against hypotheses that postulate inner mental states; it currently plays a role in cognitive ethology, guiding inference towards hypotheses that are "conservative" in the psychological mechanisms they assign. It is worth trying to understand the canon in its historical context. It also is worth seeing whether the canon can be defended in ways that its originator never imagined.

To understand Morgan's canon historically, we must understand what Morgan was reacting *against*. Darwin had argued for the mental continuity of human and nonhuman organisms. His chosen successor, Romanes, continued to emphasize this idea. Darwin's objective was to show that evolutionary ideas apply to mental characteristics no less than they apply to morphology and physiology. If all living things are related genealogically, we can locate the emergence of novelties in the interior branches of phylogenetic trees like the one depicted in Figure 9.1. According to this evolutionary conception, new traits emerge from old ones; we should be able to find the vestiges of ancient forms in more modern adaptations.

Darwin and Romanes defended this point about psychological evolution by relating anecdotes about animal behavior that were saturated with anthropomorphism (Richards, 1987). For example, in Chapter Two of *The Descent of Man*, Darwin tells stories about animal behavior to support the claim that language, self-consciousness, an aesthetic sense, and belief in God are qualitatively similar (though not identical) to mental faculties found in nonhuman organisms. Morgan also wished to defend the evolutionary hypothesis that all life is genealogically related, but he saw that the case for

evolution does not require one to gloss over the differences that separate human beings from the rest of nature. One branch of a phylogenetic tree can develop novelties that do not emerge on others; a shared genealogy does not require that there be no qualitative differences among the traits exhibited by related species (Gottlieb, 1979, p. 150).

Although Morgan's insight accords well with a modern evolutionary point of view, there is something decidedly unmodern about Morgan's ideas on the foundations of psychology. Like many psychologists writing at the time, Morgan maintained that attribution of mental states to others depends on an introspective examination of one's self. When I raise a cup to my lips, this is because I believe that the cup contains a palatable liquid that I desire to drink. When I see another human being perform the same action, I infer a similar mental cause. Morgan saw that this pattern of inference extends across species boundaries. What Morgan termed the *double inductive method* leads one to interpret the behavior of organisms in other species as stemming from the same causes that move human beings to action. The net result is that my interpretation of nonhuman organisms is prey to a bias—the bias of *anthropomorphism*.<sup>5</sup>

This bias requires a counterbalance, and that was the role that the canon played in Morgan's thought (Burghardt, 1985, p. 912). Morgan believed that the *simplest* hypothesis would be that other organisms are just like us. If I drink water because I believe that water is thirst quenching and I want to stop being thirsty, then the simplest inference is that other organisms drink water for the very same reason. It is Morgan's canon that leads one to ask whether drinking behavior in other species can be explained by a psychological mechanism that is less elevated than the beliefs and desires that animate human beings. If the behavior *can* be so explained, then it *should* be so explained. We should conclude that other organisms are *not* like us psychologically. The proximate mechanism that drives drinking behavior in human beings differs from the mechanisms at work in other creatures.

Seen in this historical context, it is evident that Morgan's canon served a useful function; it provided a needed corrective to the anthropomorphism that introspective methods and uncritical anecdote-mongering tended to engender. However, this point is not enough to justify the canon. Although the canon helps one avoid the bias of anthropomorphism, the question needs to be asked whether it introduces an opposite bias of its own. If other creatures really are like us in certain respects, perhaps the canon will lead one to miss this fact about nature. The canon would not make sense if it merely avoided one bias by embracing another.

At this point, it is worth considering the idea that Morgan's canon is a version of Occam's razor—the principle of parsimony. As I've mentioned, Morgan claimed that the canon leads one to reject a simple theory, not accept it. Nonetheless, many commentators have thought that the canon is justified because of its connection to the principle of parsimony. For example, Skinner (1938) says: "Darwin, insisting upon the continuity of mind, attributed mental faculties to subhuman species. Lloyd Morgan, with

his law of parsimony, dispensed with them in a reasonably successful attempt to account for characteristic animal behavior without them" (4).<sup>6</sup> Boring (1950) also thinks that the canon is a version of the razor, but denies that the principle of parsimony is legitimate when the problem is to infer the mental capacities of nonhuman organisms. Boring says that "nature is notoriously prodigal; why should we interpret it only parsimoniously?" (p. 474). Boring's criticism, I suggest, involves a misunderstanding of what the principle of parsimony asserts. Roughly speaking, Occam's razor tells one to accept the simplest theory that fits the evidence; it places no upper bound on how complex that theory will have to be. Even if nature is prodigal, Occam's razor tells us to find the simplest theory that is consistent with its observed prodigality (Sober, 1988).

Boring is not the only psychologist to have thought that Morgan's canon is defective because it is a version of Occam's razor. Walker (1983) provides an interesting elaboration of this line of argument:

Clearly, because sticklebacks and chimpanzees both build nests, we should not be obliged to believe that the psychological processes available to the chimpanzee are the same as those utilized by the stickleback. . . . There is a very general, and a very difficult problem behind all this, which infects the roots of behaviourist systems of explanation of animal activities. It may be described as the "same behavior—same mechanism" fallacy. (p. 57)

It might be replied that the similarity in behavior here described between sticklebacks and chimpanzees is superficial; a fuller profile of their respective behaviors would show that the similarities are slight. So, postulating the same psychological processes is in fact *not* demanded by Morgan's canon. In contrast, when two organisms exhibit a detailed set of behavioral similarities, there seems to be a presumption in favor of postulating similar underlying processes. This apparently is why we explain behavioral similarities among conspecifics by saying that they have similar psychological mechanisms. What is true for conspecifics also is true for different species in the same genus, and so on, up the taxonomic hierarchy.<sup>7</sup>

The conclusion I draw is that Walker's point about chimp and stickleback nest building does not undermine the idea that common cause explanations are often preferable to explanations that postulate separate causal mechanisms (Sober, 1994b). Still, the question remains of what this has to do with "higher" and "lower." In this regard, it is worth asking whether Morgan's canon really is an instance of Occam's razor at all. Newbury (1954), for example, suggests that the razor "is applied when we adhere to a *paucity* of assumptions, whereas Morgan's Canon refers to *lower* processes of development" (p. 72). Newbury's point is an excellent one. If parsimony is assessed by counting mechanisms, then attributing a single lower mechanism and attributing a single higher mechanism are equally parsimonious. And how should we compare a hypothesis that explains a set of behaviors by postulating a single higher mechanism and a hypothesis that explains those behaviors by postulating twelve separate, lower, mechanisms? The

hypothesis that involves a single higher mechanism seems to be simpler than the hypothesis of a dozen lower mechanisms. Pending clarification of what "higher" and "lower" mean, it is quite unclear how Morgan's canon is related to Occam's razor.

As I have already mentioned, Morgan argues that his canon rests on a specifically evolutionary foundation. He formulates the problem of justification by asking the reader to consider three "divergently ascending grades of organisms." Species *a*, man, has ascended to a higher level than *b*, and *b* has risen higher than *c*. Each of these organisms may exhibit, to some degree, each of three "ascending faculties or stadia in mental development," numbered 1, 2, and 3. How might these three faculties be represented in the three taxa? Morgan describes three possible patterns by which psychological faculties might be distributed across species; Morgan calls each distribution pattern a "method." I've reproduced his graphical representation of the problem (Morgan 1894, p. 56) as my Figure 9.3.<sup>8</sup>

Morgan calls the first possibility the *Method of Levels*. Here, "the faculties or stadia are of constant value. In the diagram, *b* has not quite reached the level of the beginning of the third or highest faculty, while *c* has only just entered upon the second stadium" (p. 57). The Method of Levels apparently says that an organism must attain a certain level of development of a lower faculty before it can have any trace whatever of a higher faculty.

The second alternative Morgan dubs the *Method of Uniform Reduction*. In this arrangement "in both *b* and *c* we have all three faculties represented in the same ratio as in *a*, but all uniformly reduced" (p. 57). The idea here is that a lower organism has all the faculties that a higher organism possesses, but it has them developed to a lesser extent.

The third alternative is the *Method of Variation*, "according to which any one of the faculties 1, 2, or 3, may in *b* and *c* be either increased or reduced relative to its development in *a*" (p. 57). This pattern seems to be the least constraining of the three; evidently, the method of variation is the method of *anything goes*.

Morgan summarizes how the above methods differ by asking us to suppose that:

*b* represents the psychic stature of a dog. Then, according to the interpretation on the method of levels, he possesses the lowest faculty (1) in the same degree as man; in the faculty (2) he somewhat falls short of man; while in the highest faculty (3) he is altogether wanting. According to the interpretation on the method of uniform reduction he possesses all the faculties of man but in reduced degree. And according to the interpretation on the method of variation he excels man in the lowest faculty, while the other two faculties are both reduced but in different degrees. (p. 57)

Morgan then asserts that the process of evolution by natural selection entails that "it is the third method . . . which we should expect *a priori* to accord most nearly with observed facts." He notes that "in the diagram by

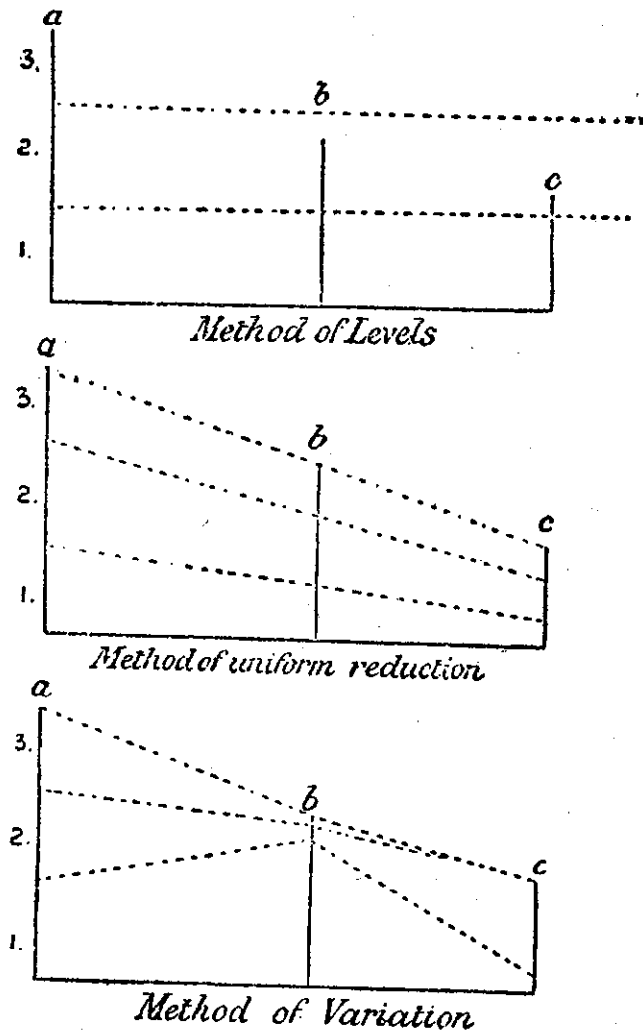


Figure 9.3 Morgan's depiction of three alternative "methods" that describe how psychological faculties 1, 2, and 3 might be distributed among three species *a*, *b*, and *c*.

which the Method of Variation is illustrated, the highest faculty 3 is in *c* reduced to zero;" the total absence of higher faculties in lower organisms is entirely possible.

Morgan now is able to bring his argument for the canon to its conclusion. If the Method of Variation is correct, then "it is clear that any animal may be at a stage where certain higher faculties have not yet been evolved from their lower precursors; and hence we are logically bound not to as-

sume the existence of these higher faculties until good reasons shall have been shown for such existence" (p. 59).

It is here that we can see the slippage in Morgan's logic. The Method of Variation does say that it is possible for an organism to have lower but not higher faculties. However, the Method also seems to allow for the possibility that an organism will have a higher faculty but not a lower one. This alternative happens not to be represented in Morgan's diagram; it is depicted in Figure 9.4. If "any . . . of the faculties . . . may in *b* and *c* be either increased or reduced relative to its development in *a*," then *a* may have none of a lower faculty that is found in *b*.

Although Morgan's Method of Variation is too permissive to justify the canon, it is possible to construe the concepts of higher and lower so that this objection to Morgan's argument can be finessed. In the problems depicted in Figures 9.1 and 9.2, "lower" and "higher" were mutually exclusive characteristics. Let us now stipulate new meanings for those terms:

Trait *X* is higher than trait *Y* if and only if *X* entails *Y*, but not conversely.

Lower traits can be present without higher traits, but higher traits cannot be present without lower ones. For example, if (as Morgan thought) "abstract reasoning" is higher than "sense perception," then a creature that reasons must be capable of sense perception, but it is possible for an organism to have sense perception without being able to reason abstractly. If we define "higher" and "lower" in this way, there is no Spencerian implication that evolution always moves in the direction of adding higher traits to lower ones. A parasite may lose some of the "higher" characteristics of its free-living ancestors. What an organism cannot do is lose lower traits while retaining higher ones.

This definition seems to accord fairly well with at least some of what Morgan says, and it entails that it is more probable that an organism has a

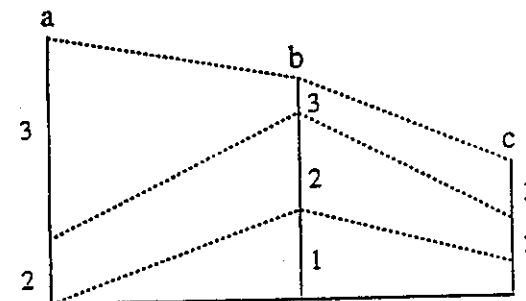


Figure 9.4 A pattern of distribution that is not depicted in Morgan's figure. Here a higher species (*a*) lacks a lower faculty (1) that is present in lower species (*b* and *c*).

lower faculty than that it has a higher faculty, relative to any evidence one might have available:

If Higher entails Lower, but not conversely, then  
 $\Pr(\text{Higher}|\text{Evidence}) < \Pr(\text{Lower}|\text{Evidence})$ .

Ironically, there is no need to appeal to evolutionary processes if this is what “higher” and “lower” mean. The proposed definition entails that Lower is more probable than Higher, even if organisms are separately created by God.

It is no objection to Morgan’s canon that its justification does not depend on the evolutionary matters that Morgan thought germane. However, the justification we are considering is limited in an important respect. Even when Higher entails Lower, but not conversely, it is still an open question how probable it is that a species should have both Higher *and* Lower, rather than Lower only. Figure 9.5 depicts a group of species in which every organism with the Higher trait also possesses the Lower trait, but not conversely. If we sample a species at random from the tips of the tree, it is more probable that the species will have Lower than that it will have Higher. However, the same sampling procedure entails that  $\Pr(\text{Higher} \& \text{Lower}) > \Pr(\text{-Higher} \& \text{Lower})$ . In a phylogenetic group of this sort, Morgan’s canon would be misleading, if the canon is taken to require one to *assert* the presence of Lower and *deny* the presence of Higher. If Lower and Higher are internal mechanisms that might produce a behavior, and at least one of them is needed to produce the behavior one observes, then the most probable inference is that the organism possesses *both* of them.

In spite of the point illustrated by Figure 9.5, there is a special circumstance in which it is possible to explain why (Higher & Lower) is a less probable hypothesis than (-Higher & Lower). This is the case in which Higher entails Lower *and* the two internal mechanisms are behaviorally equivalent—not only are both able to explain the behaviors one presently observes, but they have the same implications about all possible behaviors.<sup>9</sup> As noted earlier, it is not an implication of probability theory that having exactly one mechanism is more probable than having exactly two. However, there is an evolutionary reason to expect an organism to deploy one mechanism rather than two if their behavioral consequences are exactly the same. The reason is *energetic efficiency*. It costs energy to build and maintain a machine. This is as true for organisms as it is for automobiles. If one machine produces precisely the same range of behaviors as would occur if that machine were supplemented by a second machine, then there is no adaptive reason why that supplement should evolve. In fact, the *internal costs* (Sober, 1994a) involved suggest that the supplement should *not* evolve. Here, it is a specifically evolutionary consideration that justifies a preference for monism over pluralism with respect to internal mechanisms. Maybe Morgan would have been pleased.

The argument just presented applies to the problem of discriminating between behaviorally equivalent hypotheses. I want to emphasize, however, that this is rarely a situation that ethologists know themselves to confront.

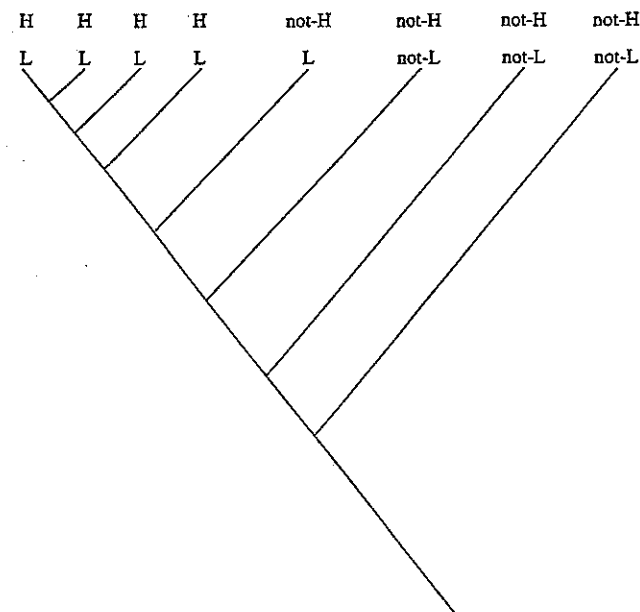


Figure 9.5 Since the presence of H entails the presence of L, but not conversely,  $\Pr(H) < \Pr(L)$ . However, in this example, it is nonetheless true that  $\Pr(H \& L) > \Pr(\text{-H} \& L)$ .

Biologists construct alternative hypotheses that may be able to explain the data at hand; that these hypotheses do equally well with respect to all possible data, however, is not at all obvious. One should be circumspect about endorsing the claim that two alternative mechanisms are in fact behaviorally equivalent; it is difficult to survey in an instant all the environmental circumstances that may arise. Mechanisms that seem to perform similarly in one range of environments may turn out to perform differently in others.

When two devices influence the same behaviors, causing the organism to perform a given behavior in a given environmental circumstance, it is often true that two control devices do a better job than one. When internal mechanisms are subject to error, redundancy can increase precision (Sober & Wilson, 1997). I suspect that it is virtually never the case that a pair of mechanisms is behaviorally equivalent with one of them acting alone. Even if two mechanisms are more energetically costly than just one, there still may be a fitness advantage in having the two rather than the one. No general principle can be stated about this tradeoff; whether one is better than two will depend on details that may vary from trait to trait and from species to species.

For this reason, I want to consider a different approach to the problem of explicating and justifying Morgan’s canon. Figure 9.6 reproduces a cartoon by S. Gross that appeared in *The New Yorker* (September 4, 1995,

42). Some penguins stand in a group, watching one of their number flying overhead. The airborne bird is saying, "We just haven't been flapping them hard enough." Gross's cartoon suggests that he may have read Chomsky's *Rules and Representations*, in which Chomsky (1980, p. 239) asks whether nonhuman organisms possess something like the human capacity for language. Here is his answer:

It is conceivable, but not very likely. This would constitute a kind of biological miracle, rather similar to the discovery on some unexplored island of a species of bird that had never thought to fly until instructed to do so through human intervention. . . . It is difficult to imagine that some other species, say the chimpanzee, has the capacity for language but has never thought to put it to use.

Gross' joke is based on Chomsky's principle—a useful ability does not go unused.

To bring this idea into contact with Morgan's canon, let's consider a new definition of the concepts of *higher* and *lower*:

One internal mechanism is higher than another if and only if the behavioral capacities entailed by the former properly include the behavioral capacities entailed by the latter.



"We just haven't been flapping them hard enough."

Figure 9.6 An illustration of the principle that useful capacities get used. Drawing by S. Gross; © 1995. *The New Yorker Magazine, Inc.*

According to this construal, "higher" faculties allow an organism to do more. The present proposal does not assume that "higher" faculties always evolve after "lower" ones. Nor does it assume that an organism with a "higher" mechanism also must possess a "lower" mechanism; even though the behaviors provided by the higher mechanism properly contain the behaviors provided by the lower mechanism, there is no implication relation between the mechanisms themselves.

Let us suppose that the internal mechanisms,  $H$  and  $L$ , have the following behavioral consequences.  $H$  allows an organism to perform the behaviors in set  $B_1$  and those in set  $B_2$ . In contrast, if an organism has  $L$  but lacks  $H$ , then its behavioral repertoire shrinks to  $B_1$ . Let us assume, with Gross and Chomsky, that what an organism *can* do, it *will* do, given the right circumstances.

What can we infer about the presence of  $H$  and  $L$ , if we see an organism perform the behaviors in  $B_1$ ? I suggest that no conclusion can be drawn as to whether the organism has  $H$  alone, or  $L$  alone, or both together. However, suppose we observe the organism for a long time under varied circumstances, never seeing it perform the behaviors in  $B_2$ . I take it that we now are entitled to an additional inference: since we have not observed the organism perform these behaviors, we have observed it not performing them. We have observed a nonevent. This involves something more than merely not observing an event. Someone who has never looked at the organism has failed to observe it perform  $B_2$ . However, only those who have observed the organism are entitled to say that they have seen the organism not perform  $B_2$ . If we now have the premiss  $\neg B_2$ —that these behaviors are outside the organism's repertoire—then we can deduce  $L \& \neg H$  (assuming that  $H$  and  $L \& \neg H$  are the only two alternatives). There is no need for a special principle of nondeductive inference to derive this consequence. Deductive logic suffices.<sup>10</sup>

It is important to bear in mind that this line of reasoning requires that one has observed the organism in "appropriate circumstances." As far as we know, the human beings who lived 100,000 years ago were anatomically modern. Their fossil traces are indistinguishable in shape from the bones that modern human beings possess. Let us suppose, then, that their brains were just as big as ours, and that they had all the same cognitive capacities. In spite of this similarity, an observer of these organisms would notice that they fail to do a number of things that modern people do. Apparently, these human beings had no decorative arts; the oldest examples arose some 60,000 years later, with most examples (e.g., cave paintings) appearing a mere 10,000 to 17,000 years ago (Bahn, 1992). Agriculture and written language are even more recent developments. However, it does not follow from these behavioral facts that the human beings of 100,000 years ago lacked the capacities we have. It is perfectly possible that the circumstances in which they lived explain why some of their capacities never surfaced in behavior.

So, on second thought, the idea underlying Gross's cartoon and Chomsky's claim is not entirely unproblematic. Capacities are manifested



in behavior only when the circumstances are right. And a pervasive theme in the history of life is the existence of evolutionary spin-off. Perhaps our big brain evolved in ancestral human beings for one set of reasons, but then, in novel circumstances, this big brain has allowed us to do other things. We now use our brains to do science, philosophy, and the arts, but this is not why the big brain evolved. Our ancestors had the same big brains we have, but their behavior was quite different. This point must always temper the inference that an organism's failure to exhibit a behavior shows that the organism is incapable of performing the behavior.

The argument I have given uses a definition of "higher" and "lower" according to which the behavioral capacities of a higher internal mechanism properly include those associated with a lower mechanism. Although this definition fits some examples in which we may be disposed to talk of higher and lower, it fails to fit others. Perhaps the ability to fly is "higher" than the ability to glide, but even so, there may be things that gliders can do that flyers cannot. Fortunately, there is no need to insist on the definition of higher and lower that I have used, since the epistemological point remains the same even for internal mechanisms whose associated behavioral capacities overlap only partially. If one observes a behavior that is part of the repertoires associated with both *L* and *H*, then no conclusion can be drawn. However, if the organism fails to perform a behavior that should have occurred if it had *H*, but could not have occurred if it had *L* alone, this is evidence that favors *L* and goes against *H*. We may leave behind the question of how "higher" and "lower" should be defined and recognize that this simple point is fundamental.<sup>11</sup>

Let me bring this analysis of Morgan's canon down to earth by applying it to the piping plovers that I discussed at the beginning of the chapter.

When a predator approaches their nest, piping plovers move away from the nest and peep loudly. They engage in "false nesting"; that is, they make it look as if they are sitting on their eggs when, in fact, they are just sitting on sand. And they produce the broken-wing display (Ristau, 1991). Consider two possible explanations of these behaviors. The first hypothesis (*H*<sub>1</sub>) says that the birds want to protect their young, and believe that these actions will have that effect. The second hypothesis (*H*<sub>2</sub>) is that the plovers want to protect their young and believe that deceiving the predator is the best way to accomplish this; they perform the actions listed because they think that these actions will deceive the predator and thereby protect their young. This second hypothesis attributes second-order intentionality (Dennett, 1987) to the birds—the hypothesis says that they form representations about the mental states of other organisms. The first hypothesis limits itself to a claim of first-order intentionality; it says only that the birds have beliefs about nonmental subjects.

Most philosophers believe that second-order intentionality entails first-order intentionality, but not conversely. For example, if a plover is able to form beliefs about what predators *think* about whether its wing is broken, then the bird is also able to form beliefs about whether its wing is in fact

broken. However, the converse does not hold; it is possible to form beliefs about wings without also forming beliefs about beliefs about wings.

Morgan's canon, applied to this problem, suggests that we should endorse the first hypothesis, and reject the second. Since *H*<sub>2</sub> entails *H*<sub>1</sub>, but not conversely, it is a theorem of probability theory that  $\Pr(H_1) > \Pr(H_2)$ . However, nothing follows, so far, as to whether  $\Pr(H_1 \& \neg H_2) > \Pr(H_1 \& H_2)$ . Nor is it at all obvious that the conjunction *H*<sub>1</sub> &  $\neg H_2$  is behaviorally equivalent with the conjunction *H*<sub>1</sub> & *H*<sub>2</sub>. So, energetic efficiency cannot be used to argue that the hypothesis of second-order intentionality is false. Presumably, the second-order ability underwrites behavioral capacities not provided by first-order intentionality alone.

It is tempting to think that some nondeductive principle of inference needs to be imported here; since the plover's behavior can be explained by both hypotheses, some such principle may seem to be needed to underwrite the conclusion that plovers have first-order intentionality and lack the second-order ability. I agree that the more "conservative" hypothesis is the one that is more plausible; however, I suggest that no special principle of inference is needed.

It isn't the plover's broken-wing behavior that is doing the work here, but the fact that the plover is *not* producing the sorts of behavior that only second-order intentionality would be able to generate. It is this *nonevent* that ethologists implicitly take into account when they favor the hypothesis that only first-order intentionality is present. Let me try to persuade the reader of this diagnosis with a thought experiment. Suppose I told you, "I am thinking of an organism that is able to protect its young by performing actions that have the effect of misleading predators." I then ask: "Am I thinking of an organism that has first-order intentionality only, or am I thinking of one that has first- and second-order intentionality both?" My opinion is that you should decline to answer my question. For all you know, I am thinking of a human being; alternatively, I may be thinking of a plover. When the problem is stripped of background information in this way, there is little inclination to think that a behavior that could occur with only first-order intentionality is in fact produced in that way.

It may be replied that there are many organisms that have first-order intentionality, but only a few that have second-order intentionality. I agree that this consideration would be relevant, if you thought I had chosen my example by drawing an organism at random. However, you have no reason to think this. Furthermore, Morgan's canon cannot rely on the assumption that it is more common for organisms to have lower abilities alone than for them to have lower and higher abilities together. This may be true for the example of first- and second-order intentionality, but there is no reason why the claim should be true in general. In any event, I take it that the canon is intended to guide our inferences about internal mechanisms when we don't have this type of prior knowledge.

If this reconstruction of the meaning and justification of Morgan's canon is correct, then the canon has been widely misunderstood. Morgan

thought that the canon rests on a specifically evolutionary rationale; however, we have seen that his evolutionary justification does not work, and the justification I am suggesting does not depend on any evolutionary assumption. Many others have seen the canon as an instance of Occam's razor, the principle of parsimony. But there is no *need* to understand the canon in this way. If we take Morgan's terms "lower" and "higher" to mark a difference in the behavioral capacities that a psychological mechanism provides, and if we are entitled to expect an organism to exhibit its behavioral capacities in suitable circumstances, then observed behavior will sometimes tell us to attribute a lower rather than a higher psychological mechanism. It isn't that observing  $B_1$  tips the scale in favor of  $L$  and against  $H$ , with the principle of parsimony underwriting the inference. Rather, it is the observation of the nonoccurrence of  $B_2$  that points to  $L$  and away from  $H$ , with the inference proceeding deductively. Construed in this way, Morgan's canon makes sense, but neither evolution nor parsimony is the reason why.

### Notes

I am grateful to Colin Allen, Robin Andreasen, Andre Ariew, Marc Bekoff, Len Berkowitz, Ned Block, Paul Bloom, Denise Cummins, Berent Enç, Ted Everett, David Papineau, Larry Shapiro, Peter Godfrey-Smith, Chris Stephens, and Denis Walsh for useful discussion, and to Denis Walsh, in addition, for the example of the protorothyrids.

1. In the book's second edition, Morgan (1903) rephrases the canon and then adds an important clarification: "To this, however, it should be added, lest the range of the principle be misunderstood, that the canon by no means excludes the interpretation of a particular activity in terms of the higher processes, if we already have independent evidence of the occurrence of these higher processes in the animal under observation" (p. 59). If methodological behaviorism is the view that human behavior should be explained without appealing to inner mental states, even though we know that human beings in fact occupy such states, then the rider that Morgan appended entails that the canon cannot be used to justify methodological behaviorism. Similar remarks apply to the use of Morgan's canon in spurious justifications of psychological egoism. See Chapter 10 of Sober and Wilson (1998) for discussion.

2. This is not to say that Darwin disavowed the idea of evolutionary progress; see Ospovat (1981, chapter 9). Sober (1994e) discusses modern evolutionary theory's attitude to this concept; see also the essays in Nitecki (1988).

3. For an explanation of the conceptual framework of cladistic parsimony, see Sober (1988).

4. The group comprised of the question mark taxon and the taxa that have the character state  $H$  are said to be *monophyletic*; a monophyletic group is composed of an ancestor and all its descendants. The members of the monophyletic group just described are more closely related genealogically to each other than any of them is to the taxa in the figure that lie outside.

5. Morgan's brief for the indispensability of introspection (which rehearses a traditional solution to the philosopher's problem of other minds) arguably pertains to the *context of discovery*, not the *context of justification*. If so, the eviden-

tial warrant of third person attributions of mental states may depend not at all on introspection. In Sober (1995), I argue that the analogy argument for the existence of other minds is better construed as an *abductive* argument than as an argument based on *inductive sampling*. Introspection plays no essential role in the former; moreover, the standard objection to the argument—that the sample size is too small (i.e.,  $n=1$ )—has no bite when the argument is understood in this way.

6. See note 1.

7. There is no precise degree of similarity that is necessary or sufficient to postulate a common underlying mechanism. Some similarities that unite chimps and sticklebacks *are* homologies, though building a nest is not one of them. See Sober (1988) for further discussion.

8. It is an unfortunate feature of Morgan's graphical representation that the overall "highness" of a species is a sum of the degree to which it has developed different mental faculties. This implies a type of commensurability that is quite unjustifiable—namely, that a "unit" of change in one faculty has the same effect on "highness" as a "unit" of change in any other.

9. In saying that the mechanisms are *behaviorally* equivalent, I am not saying that the two hypotheses are *observationally* equivalent. After all, there is more to observation than observing an organism's behavior; one can open the black box and observe the organism's physical make-up. I would be loathe to endorse a principle for discriminating between observationally equivalent theories, but I have no problem with the idea that there may be reasons for saying that two behaviorally equivalent theories differ in plausibility; see Sober (1994c) for details.

10. The argument here resembles the one developed in Sober (1994d), in which two influential arguments in evolutionary biology that appeal to the principle of parsimony are reformulated so that no such appeal is required.

11. It is worth noticing that this construal of Morgan's Canon does not license a wholesale preference for purely physiological explanations over explanations that attribute representational content. Suppose that  $H$  and  $L$  are psychological mechanisms, each involving the formation of representations. Suppose further that the organism under study implements  $H$  if it has physiological structure  $P_1$  and implements  $L$  if it has physiological structure  $P_2$ . Morgan's canon concerns the choice between  $H$  and  $L$  and the choice between  $P_1$  and  $P_2$ ; physiological hypotheses entail behavioral capacities just as much as representational hypotheses in psychology do. However, the canon does *not* entail a preference for  $P_1$  over  $H$ ; these are equally "high" in any reasonable sense of the term.

### References

- Bahn, P. (1992). Ancient Art. In S. Jones, R. Martin, & D. Pilbeam (Eds.), *Human evolution* (pp. 361–364). Cambridge: Cambridge University Press.
- Boakes, (1984). *From Darwin to Behaviorism*. Cambridge: Cambridge University Press.
- Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton Century Crofts.
- Burghardt, G. (1985). Animal Awareness—Current Perceptions and Historical Perspective. *American Psychologist*, 40, 905–919.
- Carroll, R. (1988). *Vertebrate paleontology and evolution*. New York: W. H. Freeman.
- Cheney, D., & Seyfarth, R. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.

- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Darwin, C. (1871/1981). *The descent of man and selection in relation to sex*. Princeton: Princeton University Press.
- Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Ghiselin, M. (1969). *The triumph of the Darwinian method*. Berkeley: University of California Press.
- Gottlieb, G. (1979). Comparative Psychology and Ethology. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 147–174). Hillsdale, NJ: Erlbaum.
- Heyes, C. (1994). Social Cognition in Primates. In N. J. Mackintosh (Ed.), *The handbook of perception and cognition, vol. 9, Animal learning and cognition* (pp. 281–304). New York: Academic Press.
- Morgan, C. Lloyd (1894). *An introduction to comparative psychology*. London: Walter Scott.
- Morgan, C. Lloyd (1903). *An introduction to comparative psychology* (2nd ed.). London: Walter Scott.
- Newbury, E. (1954). Current Interpretation and Significance of Morgan's Canon. *Psychological Bulletin*, 51, 70–74.
- Nitecki, M. (Ed.) (1988). *Evolutionary progress*. Chicago: University of Chicago Press.
- Ospovat, D. (1981). *The development of Darwin's theory*. Cambridge: Cambridge University Press.
- Richards, R. (1987). *Darwin and the emergence of evolutionary theories of mind and behavior*. Chicago: University of Chicago Press.
- Ristau, C. (1991). Aspects of the Cognitive Ethology of an Injury-Feigning Bird, the Piping Plover. In C. Ristau (Ed.), *Cognitive ethology—the minds of other animals* (pp. 91–126). Hillsdale, NJ: Erlbaum.
- Skinner, B. (1938). *The behavior of organisms*. New York: Appleton Crofts.
- Sober, E. (1988). *Reconstructing the past—parsimony, evolution, and inference*. Cambridge, MA: MIT Press.
- Sober, E. (1994a). The Adaptive Advantage of Learning and A Priori Prejudice. In *From a biological point of view* (pp. 50–70). Cambridge: Cambridge University Press.
- Sober, E. (1994b). The Principle of the Common Cause. In *From a biological point of view* (pp. 158–174). Cambridge: Cambridge University Press.
- Sober, E. (1994c). Contrastive Empiricism. In *From a biological point of view* (pp. 114–135). Cambridge: Cambridge University Press.
- Sober, E. (1994d). Let's Razor Ockham's Razor. In *From a biological point of view* (pp. 136–157). Cambridge: Cambridge University Press.
- Sober, E. (1994e). Progress and Direction in Evolution. In J. Campbell (Ed.), *Progressive evolution?* Boston: Jones & Bartlett.
- Sober, E. (1995). *Core questions in philosophy* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Sober, E., & Wilson, D. (1998). *Unto others—The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Walker, S. (1983). *Animal thought*. London: Routledge.
- Williams, G. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.
- Wilson, E. (1971). *The insect societies*. Cambridge: Harvard University Press.

## 10

## Do's and Don'ts for Darwinizing Psychology

LAWRENCE A. SHAPIRO

Darwin's claims about the evolution of *homo sapiens* by natural selection apply as much to human minds as they do to human morphology. Many of our morphological similarities to members of other species owe to our shared ancestry. Likewise, although we have no direct acquaintance with the minds of nonhuman animals, our common ancestry should lead us to expect remarkable mental similarities. This continuity in human and nonhuman minds, repellent as it may have appeared to an unsuspecting nineteenth-century audience, is a reason for celebration in our life sciences today. Ethology, which once limited itself to cumbersome, mechanistic explanations of animal behavior, now avails itself of cognitive psychology's mentalistic vocabulary to study topics such as memory and imagery in animals (Yoerg & Kamil, 1991). Cognitive psychology, correlatively, gains from the stepping stones that investigations of nonhuman minds afford (e.g., Parker & Gibson, 1979). In short, Darwin's discovery that human and nonhuman minds differ in degree rather than kind (provided they are phylogenetically related) allows the possibility that the study of any mind will inform the study of any other mind, regardless of species. For this latter point alone, Darwin's importance in the history of psychology is most widely, and easily, recognized.<sup>1</sup>

This said, I wish now to suggest that the theory of natural selection makes another significant contribution to psychology that only few have appreciated despite the fact (maybe because of the fact?) that it is even more obvious than the connection between human and nonhuman minds. What Darwin did for psychology is to license and ground the ascription of teleo-