1-1-1997

Individualism and Evolutionary Psychology (or: In Defense of 'Narrow' Functions)

David J. Buller

Original Citation
Buller, David J. 1997. "Individualism and Evolutionary Psychology (or: In Defense of 'Narrow' Functions)." Philosophy of Science 64: 74-95.

Individualism and Evolutionary Psychology (or: In Defense of "Narrow" Functions)*

David J. Buller†‡
Department of Philosophy, Northern Illinois University

Millikan (1993) and Wilson (1994) argue, for different reasons, that the essential reference to the environment in adaptationist explanations of behavior makes (psychological) individualism inconsistent with evolutionary psychology. I show that their arguments are based on misinterpretations of the role of reference to the environment in such explanations. By exploring these misinterpretations, I develop an account of explanation in evolutionary psychology that is fully consistent with individualism. This does not, however, constitute a full-fledged defense of individualism, since evolutionary psychology is only one explanatory paradigm among many in psychology.

1. Introduction. Psychological individualism is the principle that an individual's psychological states supervene on that individual's (current) internal physical states. Although individualism can appear to be an ontological thesis, it is actually a methodological constraint on psychological explanation deriving from a commitment to causal explanations of behavior. The idea is this. Conditions in an individual's external environment either have had an effect on that individual's internal states or have not. If they have not, they could not have had an effect on the behavior produced by that individual; so they are ex-

*Received July 1995; revised November 1995.

†Send reprint requests to the author, Department of Philosophy, Northern Illinois University, DeKalb, IL 60115, USA.

‡For thoughtful input on beta versions of this paper (not all of which I could incorporate while respecting word limits), I would like to thank Justin D'Arms, Valerie Hardcastle, Ruth Millikan, Elliott Sober, two anonymous referees, and an audience at Northwestern University (especially Arthur Fine, Christopher Horvath, David Hull, and Lenny Moss).

Philosophy of Science, 64 (March 1997) pp. 74–95. 0031-8248/97/6401-0004$2.00

planatorily irrelevant. If they have, then their effect on the individual's behavior has been mediated by their effect on internal states. However, since precisely *how* external conditions affect internal states is partly determined by the current global economy of an individual's internal states, the same external conditions may have different effects on the internal states of different individuals (or the same individual at different times). So it is not the external conditions that explain behavior, but rather the internal states they affect. Consequently, even if external conditions have affected internal states, reference to those external conditions is explanatorily uninformative. Either way, then, a *complete* causal explanation of an individual's behavior cites only states that supervene on that individual's internal physical states.

Millikan (1993) and Wilson (1994) argue, on different grounds, that individualism is inconsistent with explanation in evolutionary psychology. Their reasons stem from the fact that evolutionary psychology appears to explain behavior as the product of evolution by natural selection. This, in turn, appears to involve viewing some given type of behavior or behavioral pattern as an *adaptation*—a heritable phenotypic character that, within the particular environment prevailing at some point in the evolutionary history of a species, conferred a selective advantage on individuals exhibiting it. Both Millikan and Wilson, in different ways, take this essential reference to the environment to make adaptationist accounts of behavior inconsistent with individualism.

In what follows, I will show that their arguments are based on misunderstandings of the explanatory role of reference to the environment and that, consequently, explanation in evolutionary psychology is fully consistent with individualism. I do not, however, intend this as a defense of individualism. For I believe in the *disunity* of psychology: Psychology is not the monomorphic enterprise that philosophical talk of "the nature of psychological explanation" implies, but is a motley of research programs with differing explanatory goals. Thus, showing that one research program within psychology is individualistic does not show that all forms of psychological explanation are. My goal, rather, is merely to elucidate the nature of explanation in *evolutionary* psychology.[1]

2. "Evolutionary Explanations of Behavior"? Wilson argues simply that "evolutionary explanations of behavior" typically cite "the pressures

1. I will discuss only psychological explanations based on traditional neo-Darwinism, not explanations based on *developmental systems theory* (see, e.g., Griffiths and Gray 1994); for the overwhelming majority of work in evolutionary psychology is based on neo-Darwinian evolutionary theory, and it is not yet clear precisely how developmental systems theory will lead to a different form of *psychological* explanation.

of selection causally responsible for the existence of a given behavior." Such explanations are inconsistent with individualism, he concludes, "since selection pressures do not supervene on the intrinsic, physical properties of individuals" (1994, 59–60). This argument fails for three reasons, discussion of which will serve to focus the issues.

First, selection pressures are not "causally responsible for the existence" of any phenotypic traits, including behaviors. Evolution by natural selection involves two processes, repeating every generation: (1) a process that produces heritable phenotypic variation and (2) the process of *selection*, which "winnows" that variation through "environmental demands" (Amundson 1994, 570; Mayr 1988, 98). Process (1) itself involves two subprocesses: *reproduction*, wherein genes are transmitted from individuals in one generation to those in the next, and *development*, wherein the genome of each individual in the new generation (epigenetically) produces its phenome. Selection pressures only enter into process (2) as the environmental demands that winnow the variation *already* produced in process (1). In winnowing that variation, selection alters the frequencies of the genes available to be transmitted through reproduction. This, in turn, alters the frequencies of the types of developmental process that will occur in the next generation and, hence, the frequencies of the phenotypic characters produced by those developmental processes. Thus, selection is only *a cause of changes in the frequencies* of phenotypic traits in some population; it does not causally produce any phenotypic traits of individuals in that population (Cummins 1975, 750–751; Endler 1986, 46 and 241; Maynard Smith 1993, 20; Mayr 1988, 98; Sober 1984, 149–152).

Contrary to Wilson's interpretation, when evolutionary psychologists speak of explaining the "existence" of a behavior, they are referring to reconstructing the history of *cumulative modifications* to previous characters that culminated in the current form of behavior (Alcock 1993, 1–6). Thus, such "existence" explanations account for the *genealogy*, not the etiology, of some behaviors. And the modifications cited in them also are not caused by selection; they are (typically) caused by mutation and then *persist* because of being selected.

Neander (1995b), however, argues that such cumulative evolution (via cumulative selection of "gene sequences") shows that selection *is* a cause of an individual's phenotypic traits. Neander's argument is clearest if we consider a simple example of a hypothetical species of haploid, uniparental organisms. Suppose that all individuals in the current generation of this species have the gene sequence ⟨A2,B2,C2⟩, which has evolved from ⟨A1,B1,C1⟩ through successive mutations (A1 to A2, etc.), where selection so strongly favored each new sequence that it was driven to fixation generations before another mutation. Sup-

pose further that the sequence ⟨A2,B2,C2⟩ produces (during normal development) the phenotypic trait T. Thus, to explain why an individual, S, in the current generation possesses T, we would cite the developmental process by which ⟨A2,B2,C2⟩ produces T. But Neander argues that we could provide a more complete causal explanation by citing the reproductive process by which S inherited ⟨A2,B2,C2⟩ from its parent. An even more complete causal explanation, she argues, would trace the chain of inheritance of ⟨A2,B2,C2⟩ back through S's lineage. Eventually, however, we will have to explain the causal *origin* of the ⟨A2,B2,C2⟩ sequence, with some individual Z, by citing the fact that C1 mutated to C2. But, Neander argues, mutation is only *part* of the causal explanation of the origin of ⟨A2,B2,C2⟩. The other part of the explanation involves selection, since ⟨A2,B2,C2⟩ would not have arisen by mutation of C1 to C2 with Z unless there had been prior selection for ⟨A2,B2,C1⟩. If only ⟨A2,B1,C1⟩ had been available in the generation before Z, the production of ⟨A2,B2,C2⟩ would have required the mutation of B1 to B2 in addition to C1 to C2. Since the probability of two mutations occurring is much smaller than the probability of only one, the fact that selection drove ⟨A2,B2,C1⟩ to fixation by the generation before Z—as opposed to favoring ⟨A2,B1,C1⟩ or favoring neither over the other—greatly increased the probability that ⟨A2,B2,C2⟩ would occur. (Similarly, the probability increased when there was selection for the A2 mutation.) The fact that selection thus increased the probability that ⟨A2,B2,C2⟩ would occur, Neander argues, shows that it is a causal factor in the *origin* of the ⟨A2,B2,C2⟩ sequence and, hence, that it is a historical cause of S's possessing T.

But the fact that selection for ⟨A2,B2,C1⟩ increased the probability that ⟨A2,B2,C2⟩ would occur is in fact *not* relevant to explaining *S*'s possessing T. For selection did not make it more probable that *Z* would have ⟨A2,B2,C2⟩, *rather than some other individual* in Z's generation or a subsequent generation. Selection for ⟨A2,B2,C1⟩ thus did not increase the probability *that Z would have the* ⟨A2,B2,C2⟩ *sequence*. And, if it is supposed to be selection's increasing the probability of an occurrence that implicates selection in the causation of that occurrence (as per Neander's argument), selection for ⟨A2,B2,C1⟩ was not a cause of *Z*'s having ⟨A2,B2,C2⟩. An explanation of how *Z* came to have ⟨A2,B2,C2⟩ need thus cite only the fact that Z's parent had ⟨A2,B2,C1⟩ and that there was mutation of C1 to C2 in the production of Z. (Neander would argue that selection was a cause of Z's parent's having ⟨A2,B2,C1⟩; but my counterargument, *mutatis mutandis*, would again apply.) If selection was not a cause of *Z*'s having ⟨A2,B2,C2⟩, however, it was not among the historical causes of *S*'s having T. Only if selection increased the probability that ⟨A2,B2,C2⟩ would occur *in S's lineage*, rather than

elsewhere, would it be a cause of *S*'s possessing T; for, had ⟨A2,B2,C2⟩ originated with an individual *not* in S's lineage, S would not have inherited that sequence and, hence, would not possess T. In short, the increased probability that ⟨A2,B2,C2⟩ would occur was strictly a *population level* increase and did not increase the probability that any particular *individual* in that population would have ⟨A2,B2,C2⟩. Thus, selection is not a cause of any particular individual's phenotypic traits. (See Sober 1995 for further related criticisms of Neander's position.)

Returning to Wilson's argument, the second problem is that not only does selection not causally produce behaviors, strictly speaking it does not *select for* behaviors either. For there is only selection for *heritable* phenotypic characters (Sober 1984, 151); and it is not behaviors, but the dispositions and abilities to perform them, that are inherited. For example, a gazelle has not failed to inherit the stotting response if it fortuitously goes through life without ever spotting a predator; thus, it is not stotting itself, but the *disposition* to stott when a predator is spotted, that is inherited. Dispositions and abilities, however, cannot be inherited apart from mechanisms that subserve them. Thus, strictly speaking, it is mechanisms that produce and control behavior that are selected for—what are standardly called "proximate mechanisms." To put this another way, since it is genes that are transmitted from one generation to the next, and genes just build bodies, genes do not "code for" behaviors, but only for behavior control mechanisms (Dawkins 1995, 57–58). Consequently, as Cosmides and Tooby say: "To speak of natural selection as selecting for 'behaviors' is a convenient shorthand, but it is misleading usage. . . . Natural selection cannot select for behavior per se; it can only select for mechanisms that produce behavior" (1987, 281). So evolutionary psychology actually explains the evolution *of proximate mechanisms*. Behaviors of interest to evolutionary psychologists are then explained in terms of the functioning of these mechanisms.

Third, it is not the case that the only difference between explaining the evolution of proximate mechanisms and explaining how they function to produce behavior lies in the proximity to the behavior—that the evolutionary account explains what causally produces a proximate mechanism and the latter account explains how that mechanism causally produces the behavior—and that the evolutionary account thus provides a distal causal explanation of the behavior. For proximate mechanisms are also phenotypic characters. As such, selection no more causally produces them than it causally produces any other phenotypic character. Consequently, selectionist explanations of proximate mechanisms do not provide even distal causal explanations of behavior (cf. Dretske 1988, 95).[2]

2. Wilson (pp. 73–74) endorses the possibility that evolutionary psychology provides

These three points are significant. For individualism is a methodological constraint on causal explanations of why particular individuals behave in particular ways at particular points in time. To argue, then, that evolutionary psychology violates individualism by citing environmental selection pressures is beside the point; for selection pressures are not part of a *causal* explanation *of behavior*. The only causal explanations in which selection pressures feature are those of *changing frequencies of types* of proximate mechanism and, hence, of changing frequencies of the types of behavior produced by those mechanisms under certain conditions. These explanations are non-individualistic; but, since they are not causal explanations of why individuals behave as they do, they fall outside the intended methodological purview of individualism. In sum, an adaptationist account of how proximate mechanisms were shaped concerns only *how* they were shaped; and they can be shaped by non-individualistic processes, while the mechanisms so shaped, and how they contribute to the causal production of behavior, may still be wholly individualistic phenomena.

There is, nonetheless, a point of contact between evolutionary psychology and individualism. For evolutionary psychology "involves the exploration of the naturally selected 'design' features of the mechanisms that control behavior" (Tooby 1988, 67). In exploring these design features, evolutionary psychology explains the adaptive strategies that proximate mechanisms serve, and this in turn "gives precise meaning to the concept of function for proximate mechanisms" (Cosmides and Tooby 1987, 283). Thus, insofar as proximate mechanisms are functional items (functioning to produce and control behavior), there is a question about whether *their functions* can be individuated individualistically. The issue, then, is not whether an evolutionary account of proximate mechanisms cites non-individualistic phenomena; it is whether that account forces a non-individualistic functional *individuation* of proximate mechanisms.

3. The Functions of Proximate Mechanisms. Millikan contends that the functions of naturally selected proximate mechanisms *are* non-individualistic. Since her reasons are widely articulated (in 1984 and 1993, with her objections to individualism becoming focused in 1993, Chapters 7 and 8), I will present a brief rationally reconstructed argument.

Dretskean *structuring causes* of behavior. A Dretskean structuring cause brings it about that a particular internal state causes a particular movement in an individual (Dretske 1988, 33–42). But, *that some internal state causes some movement* in an individual is as phenotypic a trait as anything. Consequently, selection does not satisfy Dretske's concept of a structuring cause, only development does (as Dretske argues, pp. 47 and 92–93).

Millikan's first premise is a version of the *etiological theory of functions*: a biological item *I* has as a (proper[3]) function the production of an effect *E* just in case having produced effects of type *E* in the past caused the proliferation of items of type *I* (1984, 26; 1993, 13–14). Thus, it is a function of a proximate mechanism to produce some effect just in case the production of that effect figures in an account of why there was selection for mechanisms of that type (1993, 35–39).

The second premise is that proximate mechanisms were selected because of their environmentally distal effects. To illustrate this, consider an example that Millikan frequently employs. A chameleon possesses a proximate mechanism, *M*, which produces the following series of effects, running from proximal to very distal: It (E1) redistributes pigments in the chameleon's skin, which (E2) makes the coloring of the chameleon match its background, which (E3) makes the chameleon virtually invisible, which (E4) makes the chameleon avoid predation. Millikan contends that only E4 explains why there was selection for *M*s (in that any account that omitted reference to E4 would be unable to explain why *M*s were selected). So it is a function of *M* to produce E4. Of course, *M* produces E4 *by* producing E1, which effects E2, which effects E3, which effects E4. So the account of how production of E4 brought about proliferation of *M*s would also cite the production of E1–E3 (although E1–E3 contributed to that proliferation only by effecting E4). Thus, production of each of these effects is a function of *M*; indeed, the production of effects E1–E4 are what Millikan calls "serial functions" of *M* (1984, 35).

The final premise is that a proximate mechanism's production of distal effects "almost invariably depends upon its having a suitable surrounding environment" (1984, 30), where the more distal the effect the more its production depends on the environment (1984, 35). This can be seen in the steps from E1 to E4, where each step requires that more environmental conditions be in place. These environmental conditions are what Millikan calls the "Normal conditions" without which a proximate mechanism cannot perform its functions (1984, 33–34; 1993, 48 and 160). Since Normal conditions of the environment are thus essential to proximate mechanisms' performing their functions, Millikan concludes, "explaining the operation of these mechanisms requires describing the relations their operations normally bear to the environment" (1993, 137). Thus, environmental Normal conditions are essential to the functional individuation of proximate mechanisms.

3. Millikan's concept of a *proper function* is intended to capture biological functions, which are my focus here. I will use only 'function' for simplicity, however; for as will become clear, my disagreement with Millikan does not turn on terminology.

Before critiquing this view, let me clarify some points. First, undergirding the above argument are two assumptions that I will accept and employ: (1) that the function of a proximate mechanism is individuated individualistically just in case the output that it is the function of the mechanism to produce is itself an individualistic item, and (2) that a behavioral output is individualistic just in case it is described in a way that is insensitive to how that behavioral output is embedded in the environment. Very roughly, (2) circumscribes as individualistic those descriptions of behavior that characterize an organism's proximal responses. These assumptions conjointly entail the following criterion: A proximate mechanism is individuated individualistically just in case the function of that mechanism is to produce a proximal response.

Second, I will also assume a fairly standard version of the etiological theory of functions (see the first premise). My disagreement with Millikan will not be over this theory. Rather, I will argue that the *second premise* is faulty and that, contrary to what Millikan assumes, the etiological theory of functions lends it no support.

Third, according to Millikan's conception of the "serial functions" of a mechanism, the mechanism has as functions to produce earlier items in the series because only by producing them does it succeed in producing the terminal item in the series (on which selection acts). In addition, earlier items in the series function to produce later items, but their having these functional relations derives from the functions of the mechanism that produces the series. Returning to the chameleon example, according to Millikan pigment redistribution (E1) has the function of making the chameleon match its background (E2) only because *M* functions to produce predation avoidance (E4) by producing the series E1–E3 (1993, 56–57).

This conception of serial functions, however, faces the following difficulty. Since *M* does not function to produce E4 directly, the only means of *identifying* that *M* (rather than some other mechanism) functions to produce E4 is by identifying that E3 produces E4 and so on back to *M*'s directly producing E1. So it seems that we must be able to identify that each item in the series functions to produce the next *prior to* identifying that *M* functions to produce the last item in the series. A more natural way of viewing serial functions, then, is as follows: *M* functions to produce E1, which functions to produce E2, and so on until we get to the last item in the series; then *M* functions to produce that item in virtue of functioning to produce earlier items in the series.

If we do view serial functions in this way, however, we cannot conclude that *M* functions to produce the last item in the series in virtue of its producing earlier items. For functionality is not transitive. That is, in general, it is not true that, if the function of *A* is to produce *B*

and the function of *B* is to produce *C*, then the function of *A* is to produce *C*. In a car engine, for example, a function of the alternator is to keep the battery charged and the function of the charged battery is to turn over the starter motor; but it is not a function of the alternator to turn over the starter motor (for the alternator is not even operative until *after* the car has started). Thus, *B* may be an item that is specifiable both functionally *and* non-functionally. In that case, it may be the function of *A* to produce *B* only under some non-functional specification, while only the functional specification of *B* pertains to *C*. The functional individuation of *A*, then, would involve no reference to the production of *C*.

This allows the possibility of viewing serial functions in a way quite opposed to Millikan's. According to this alternative conception, *M* would have as function *only* to produce some proximal effect (under a non-functional description), which (under a functional description) in turn would function to produce the next item in the series, and so on until the penultimate item would function to produce the terminal item. If this is the way we should view a functional series initiated by some proximate mechanism, then that mechanism would be individualistically individuated, since its function would be only to produce a proximal response. Of course, this is only a *possible* way of viewing the functions of proximate mechanisms and the series of effects they initiate. I propose now to argue, contra Millikan, that this is the way we *should* view the functions of proximate mechanisms. My argument will consist in showing, in Sections 3.1–3.4, that four considerations favor this view.

3.1. Functions and System Repairs. The fact that functionality is not transitive is usually obscured, since we are usually attempting to understand the functions of components in large and complex systems, where components are connected precisely so as to allow their proximal effects to have, in turn, far-reaching distal effects within the system. It is often obscured even more because complex systems consist of a hierarchy of *nested* subsystems; thus, which effect we see a component as functioning to produce can be an artifact of which (sub)system within the hierarchy we select as the starting point for functional analysis. This is why we are often tempted to see the production of a distal effect as the function of a component or to see it as having multiple functions corresponding to its many effects within the system hierarchy.

But the function of a component in a hierarchical system can be seen clearly only in cases of *breakdown* of the system in which that component is embedded. In such cases, we follow a localized diagnostic procedure to isolate the cause of the breakdown and then perform one of two types of operation to get the system up and running again: We

either replace or repair a system component *or* restore lost or damaged connections among components. And when we do *not* fix a component it is because, although the system as a whole has broken down, that component has not stopped performing its function. Only a component that requires replacement or repair in cases of embedding system breakdown has stopped performing its function. To illustrate, if my car sputters and stalls when I press the accelerator, but diagnosis reveals a faulty distributor or mistimed engine, the carburetor does not require repair, since it is still performing the function it was *designed* to perform of vaporizing gasoline. Similarly, if the lights do not come on when I flip the switch, but diagnosis reveals a bad bulb or loose connection somewhere between the switch and the light, the switch is still performing the function it was *designed* to perform of channeling current to the proper wire when open. Thus, a breakdown in the system that contains a component does not (necessarily) involve that component's not performing its function. These facts imply the following *principle of functional isolation* (POFI):

> The function of a component is to produce the effect that it produces in *all possible cases* of breakdown of its embedding system in which that component does *not* require repair.

In short, if you don't fix it, it ain't broke (i.e., it's performing its function).

Millikan, however, infers the function(s) of a mechanism from the effect(s) it produces *in Normal conditions*, where saying "that a part or subsystem is in its 'normal conditions' is just another way of saying that it is part of a wider *intact* system, that the *rest* of the system to which it belongs is in place" (1993, 162). Since a breakdown in Normal conditions thus prevents a mechanism from producing some of its Normal effects, Millikan is led to take the Normal conditions *in which* a mechanism functions to be essential to its *performing* its function(s). This is why Millikan says, "if no predator comes by, . . . [the] color state of the chameleon *cannot perform* its proper function" (1993, 57; emphasis added). And, since her theory holds that the function of the color state derives from the function of the mechanism that produces it, her theory entails that the mechanism is also *not performing* its function when it effects a color change in the absence of predators.

But, even granting that the wider predation avoidance system is broken down when it does not *currently* result in avoidance of a predator, it does not follow that the color changing mechanism is not performing its function. For, if fixing the wider predation avoidance system requires only the introduction of a predator to the scene, we are only restoring a broken connection between the color changing mechanism

and the distal environment. Since we are not repairing the mechanism itself, then, the mechanism *is* performing its function even in the absence of predators. Consequently, contra Millikan, the mechanism does not actually function to produce predation avoidance. What then does it function to do? It may seem, at the very least, that it functions to make the chameleon match its background environment. But the chameleon's color changing mechanism is under autonomic control, with cues about ambient light wavelengths picked up directly by the chameleon's skin; blocking these cues by covering a portion of the chameleon's skin results in that portion turning uniformly pale (Young 1981, 308). Such blockage would constitute a breakdown in the matching-the-color-of-the-background system. We get that system working again, however, simply by restoring the connection between the chameleon's skin and the ambient light, not by repairing the color changing mechanism itself. The only condition in which we would repair the color changing mechanism itself (by operating on the chameleon) would be that in which the mechanism did not redistribute pigments in the skin in accordance with the patterns impinging on the surface of the skin. So, the function of the color changing mechanism is actually to produce a *proximal* redistribution of pigments in accordance with *proximal* inputs to the chameleon's skin.

Note that POFI does *not* identify the function of a mechanism via conditions of *the mechanism's* breakdown, but via conditions of its *embedding system* breakdown. Thus, it does *not* entail that the function of a carburetor, say, is to produce the effect that it produces when it does not require repair. A principle entailing that would engender fallacious inferences. For example, if the carburetor requires repair, the engine may sputter; so we would be led to conclude that the function of the carburetor is to produce a smoothly running engine (or worse, that its function is to suppress the sputter). POFI licenses *only* the following type of inference: if the *car engine* is broken down, and if the carburetor does not require repair in order to *fix the engine*, then the carburetor is performing its function. Further, POFI requires that we focus on *all possible* ways in which the engine can break down without the carburetor requiring repair. So, to apply POFI to the case of the carburetor, we must:

(i) consider all possible cases of engine breakdown;
(ii) determine the type of repair required in each case to get the engine running again;
(iii) ignore those cases in which the engine is fixed by (a) repairing the carburetor or (b) repairing some part or connection the breakdown of which resulted in the carburetor's non-operation

(since in these cases the carburetor will produce no effect at all, but due to no problem with it);

(iv) determine the effect that the carburetor produces in *all* the remaining cases.

The function of the carburetor, then, is to produce the effect that we discover in (iv). (And, since POFI is level neutral, it could be applied to any (sub)system within a nested hierarchy, e.g., the fuel system or the whole engine.)

This proposal differs from one made by Neander, who argues that, in cases in which a mechanism generates a series of effects, the "lowest level" in the hierarchy of functional descriptions (the level that describes the item's "most specific" function) is the "preferred" level, since that is the level that corresponds to a description of its "malfunctioning" (1995a, 118–120). I agree with much of Neander's argument, but there are two significant differences between our approaches. First, Neander provides no principled account of *how* to determine which is the "lowest level" in a hierarchy of functional descriptions and thereby determine which is an item's "most specific" function. The proposal I have offered does provide a way of isolating precisely such a lowest level. Second, Neander says that "a part malfunctions when it cannot perform its most specific function" (p. 120). But, if we have no principled way of deciding which is an item's most specific function, we have no principled way of deciding when it is malfunctioning either. So it seems that the appeal to the level at which an item is describable as malfunctioning cannot actually succeed in informing us of which is the "preferred" functional description of an item. My approach avoids this difficulty by focusing not on the breakdown of *the mechanism* in order to isolate its function, but on the breakdown of *the system that embeds it*: We work in toward the function of the mechanism through the breakdowns of its embedding system. Once we have determined the mechanism's function in this way, its breakdown is defined in terms of its failure to perform that function.

It may appear that this involves my approach in a regress. For I rely on the breakdown of an embedding system to isolate the function of one of its subsystems; breakdown of the subsystem is then defined as failure to perform that function. But identifying the embedding system as broken down requires approaching that system through breakdowns of its embedding system, which in turn must be identified as broken down. The etiological theory of functions, however, provides a terminus to this regress. For, by defining functions as selected effects, the "highest level" effect of a mechanism will be an increase in fitness for its possessor. So the regress terminates at that system whose immediate

effect is a fitness increase for the possessor of the proximate mechanism. This system will be at the highest level of the hierarchy of nested systems that embeds the mechanism, and it will include all features of the environment that are implicated in the fitness increase. It is by working back down the hierarchy from this most inclusive system that we determine the function of a proximate mechanism.

3.2. Functions and Control. Some mechanisms *control for* the degree of their effect by suppressing that effect once it reaches some threshold. All mechanisms with automatic shut-off, for instance, are of this type. The present discussion, however, concerns mechanisms that produce a *series* of effects; so the type of control that is of interest here is not control for the degree of an effect, but control for the *propagation* of effects along a series. That is, when a mechanism produces a series of effects, we can ask how far down that series the mechanism controls for the production of those effects. For example, pressing the rewind button of my VCR effects the rotation of the carriage spindle, which effects the rotation of the tape cartridge sprocket, which effects the rewinding of the tape. (Of course, a series of events mediates the pressing of the button and the rotation of the spindle, but ignore these for simplicity.) How far down this series of effects does the rewind mechanism control for the production of those effects? The answer is that, since the rewind mechanism will not produce the rotation of the spindle unless a tape is in the carriage, it controls for the rotation of the cartridge sprocket; but, since it will produce the rotation of the sprocket even if the tape in the cartridge is broken, it does not control for the tape's actually rewinding.

The VCR rewind mechanism exhibits a familiar type of control, which works as follows: Prior to producing one effect in a series, the mechanism "checks" that the conditions necessary for that effect's producing another obtain. In other words, the mechanism does not produce an effect unless it has ensured that that effect will produce yet another. When a mechanism exhibits this feature, in producing one effect it is *controlling for* the production of another effect. An even greater level of control is exhibited by a mechanism that actively *creates*, prior to producing an effect, the conditions necessary for the propagation of that effect. We can thus define what a mechanism controls for as follows. Let M be a mechanism that produces some proximal output O, which is the first in a series of effects of which E is an arbitrary member produced by O only through the mediation of a set of conditions C in M's environment. Then: M *controls for* the production of E *only if, prior to* producing O, M either creates C or checks that C

obtains. When one of these conditions is satisfied, *M*'s production of *O* ensures the production of *E*.

Note that *C*, here, is *not* a set of conditions necessary for the *operation* of *M*, but is a set of conditions necessary for one effect of *M*'s to produce another. The VCR's rewind mechanism will not operate if the VCR is not plugged in; so, if *C* were taken to include the conditions necessary for the operation of the rewind mechanism, the conclusion would be that the mechanism does not control for *any* effect at all, since it neither plugs in the VCR nor checks to make sure it is plugged in. And, since the overwhelming majority of mechanisms do not create or check for the conditions of their own operation, the conclusion would be that the overwhelming majority of mechanisms control for no effect whatsoever. In the case of the VCR, *C* would include instead the tape's being in the carriage, since that is the condition necessary for the spindle's rotation to produce the cartridge sprocket's rotation. The reason, of course, for restricting *C* in this way is that we want to know which of a mechanism's many effects it controls for *in its operations*—that is, when it actually produces *some* effect.

This notion of what a mechanism controls for explains many of our intuitive judgments about the function of a mechanism. For example, why don't we see it as the function of the carburetor to rotate the car's wheels? Because the carburetor does not "control for" the rotation of the car's wheels; that is, it neither checks that the car is in gear nor puts it in gear prior to vaporizing the gasoline. Why don't we see it as the function of the light switch to illuminate the room? Because the light switch neither checks to make sure a bulb is in the socket nor puts one in it prior to diverting current. Since issues of control are never explicit in our reasoning about functions, however, it is not appreciated that what a mechanism controls for is evidence of its function. To illustrate, consider again Millikan's claim that the function of the chameleon's color changing mechanism is predation avoidance. Since the color changing mechanism does not check for the presence of a predator prior to effecting a color change the mechanism does not control for predation avoidance. In addition, given the results when the chameleon's skin is covered, it is clear that the color changing mechanism does not even check for the conditions beyond the surface of the chameleon's skin prior to effecting a color change. What the mechanism actually controls for, then, is just the proximal distribution of pigments in the chameleon's skin. And this indicates that the function of the mechanism is only to produce that proximal output, since that is all it controls for.

This conclusion works together with that of the last section to support a wholly individualistic individuation of the color changing mech-

anism, since they (independently) support viewing its function as the production of a proximal color change. Of course, under Normal conditions, the color change produced by the mechanism is correlated with the coloring of the chameleon's background and this (sometimes) effects predation avoidance. So, under this interpretation, the chameleon's color-changing *behavior* itself performs a non-individualistic function (as part of a functional series). Recall the discussion of the non-transitivity of functionality. The case of the chameleon is one in which the color changing behavior is specifiable both functionally and non-functionally; the mechanism functions to produce that behavior under a non-functional description, while the behavior functions to produce a further effect only under its functional specification. The non-functional description of the color-changing behavior happens also to be individualistic, while the functional specification of that behavior happens also to be non-individualistic. This interpretation, however, is still fully consistent with an adaptationist account of the chameleon's color-changing mechanism. Under this interpretation, the mechanism was selected for because of the individualistic function it performs; for when it performs the function of changing the chameleon's color, the changed color benefits the chameleon by (sometimes) further effecting predation avoidance.

3.3. Behavioral Strategies and Fitness. There may, however, be a sense that the above parenthetical "sometimes" conceals the real issue. This sense would derive from the fact that the function of a proximate mechanism is the production of that effect because of which the mechanism was selected, together with Millikan's idea that the only effect of the chameleon's proximate mechanism that confers a benefit on the chameleon is that of *actually avoiding predators*. But Millikan's way of calculating benefits is not mandatory. Implicit in Millikan's theory is what I will call a *distributive method* of calculating benefits. This involves grouping all the occasions on which a mechanism M produces a behavior B and then *individually* examining the *Bs*. When the *Bs* are examined individually, it is noticed that some confer no actual benefit while others do, where the difference is that (respectively) between not producing and producing a beneficial effect E. Since M would be selected for only if it confers a benefit on its possessor, it is concluded that it must be the function of M to produce E, rather than B alone, because it is only when B produces E that any benefit is actually conferred on the possessor of M.

But there is another way to calculate benefits, which I will call the *collective method*. Rather than comparing the distributive benefits associated (or not) with individual behaviors, this method compares the

benefit levels of competing *behavioral strategies*. So rather than comparing the benefits associated with individual *B* productions, it looks at the *overall benefit* that accrues to a *strategy* of *B-production*. Where the distributive method asks "When does an individual that produces *B* thereby benefit?", the collective method asks "Does an individual that employs a strategy of producing *B* enjoy an overall greater benefit than an individual that employs an alternative strategy?" If the answer is yes, the collective method predicts that the mechanism responsible for producing *B* will be selected for.

To clarify the difference between these methods of calculating benefits, consider investment strategies. Suppose I instruct my accountant to follow a strategy of investing in everything Donald Trump invests in, but to spare me details about where individual investments are made or whether they pay off; and suppose my net worth consequently increases. The distributive method would look at my financial records and say, "Here your strategy benefited you and here it did not." The point, however, is not each individual investment and whether it paid off, but that my *strategy* of investment increased my net worth (read: financial fitness) and did so *even though I had no information regarding the individual investments*. Indeed, if I am a wise investor, I will adopt whatever investment strategy will maximize my net worth. And, in choosing among competing strategies, I can ignore the success or failure rate of the individual investments under each strategy, and focus *only* on the impact of each strategy on my net worth.

Returning to the chameleon's proximate mechanism, the collective method would say that the behavioral strategy of producing a color change has a greater overall benefit associated with it than competing strategies, regardless of whether (or how often) it effects predation avoidance on individual occasions. So what actually benefits the chameleon is the *strategy of changing color*. The collective method thus renders the same judgment as the distributive method about the contribution of the proximate mechanism to the chameleon's overall fitness. But the collective method enjoys two advantages over the distributive method. First, its benefit calculations are made at a level of behavioral description that is consistent with *independent* evidence (from Sections 3.1 and 3.2) of the functions of proximate mechanisms, whereas the distributive method *infers* those functions *from* benefits and is inconsistent with the evidence of Sections 3.1 and 3.2. Second, only the collective method accords with the focus in evolutionary game theory on modeling behavioral strategies, rather than individual behaviors and their effects (or lack thereof) *in situ* (see, e.g., Dawkins 1989, Maynard Smith 1982).

3.4. Functions and Explanatory Asymmetry. There is, perhaps, *still* a sense that one must go distributive in order to account for why a strategy is beneficial—that the chameleon's strategy of producing a color change, for example, is beneficial only because on certain individual occasions it effects predation avoidance. This sense would derive from the idea that the function of a proximate mechanism cannot be understood in the absence of information about the function of *the behavior* that the mechanism produces. But now we should ask: What *explanatory* function is served by talk of "the function of a behavior"?

Consider a hypothetical population of chameleons that are anatomically and physiologically identical to actual chameleons and, consequently, also change color to match their backgrounds. These hypothetical chameleons fortunately have no predators in their environment; but, unfortunately, the creatures on which they prey are a scarce and cautious lot, alighting nearby only when a chameleon is camouflaged. So the color-changing behavior of these hypothetical chameleons aids in luring prey, rather than avoiding predators. The considerations I have urged so far entail that the proximate mechanisms in both types of chameleon have *the same single function* of producing color-changing behavior, the difference between them concerning only the function of the *behavior* produced by the mechanisms.

Focusing now on the functions of the color-changing behavior in both types of chameleon, we can see that the explanatory role of reference to the function of the behavior is actually that of explaining *why* that behavior is beneficial and that, in turn, explains *why* the proximate mechanism producing *that* behavior was selected. Thus, saying that the function of the color-changing behavior is predation avoidance is equivalent to saying that Nature favored chameleons possessing a proximate mechanism *with the function of producing a color change* by allowing fewer of them to be eaten. Similarly, saying that the (hypothetical) function of that behavior is prey luring is equivalent to saying that (hypothetical) Nature favored chameleons possessing a proximate mechanism with that function by allowing fewer of them to starve. Talk of the function of the behavior produced by a proximate mechanism, then, is equivalent to talk of why a mechanism that functions to produce that behavior was selected for. But talk of a mechanism's function (of producing some behavior) is not equivalent to talk of why some other property of the organism was selected for. Consequently, there is an *explanatory asymmetry* between talk of the function of a proximate mechanism and talk of the function of a behavior: The latter is equivalent to talk of why the former was selected for. Given this asymmetry, there is no reason to expect that "the function" of a behavior

would "reflect back into" the function of the proximate mechanism that produces it.

Note that this is *not* claiming that there is a single level at which it is legitimate to explain the *selection* of a proximate mechanism. Consequently, it is compatible with Goode and Griffiths' demonstration that "there are several levels of theory in evolutionary biology," since "[s]election processes can be described at more or less abstract theoretical levels" (1995, 107), where these theoretical levels correspond to the levels in the series of effects produced by a proximate mechanism. Goode and Griffiths, however, infer from this that all such levels "generate genuine, complementary descriptions of etiological function" (p. 107). This inference goes through *only* given the supposition, for which Goode and Griffiths do not argue, that every level at which a selection process can be described as acting on a proximate mechanism corresponds to a level of functional description *of that mechanism*. The reason they do not argue for this claim is that it appears simply to be the etiological theory of functions—that the function of a mechanism is to produce an effect that it was selected for producing. But the preceding arguments show that, while the etiological theory captures the core of a theory of the functions of biological items, by itself it is insufficient; rather, a complete theory of functions must conjoin the etiological theory with the considerations of the last four sections.

3.5. "Animal Magnetism" (or: The Philosopher's Lodestone). Consider now a focused application of the preceding arguments to a much-discussed case. A typical anaerobic aquatic bacterium has in its cell a chain of approximately twenty particles of magnetite, called "magnetosomes." This magnetosome chain is in fixed position within the cell and aligned with the flagellum on the cell's exterior. The chain behaves as a compass needle that, in an aquatic environment, moves freely in both horizontal and vertical planes. In bacteria in the Northern Hemisphere the magnetosome chain is north-seeking, pointing north and downward (with the angle of inclination increasing as it moves closer to geomagnetic North); the flagellum thus propels the cell in the direction of geomagnetic North. This mechanism is an obvious adaptation; for, in the Northern Hemisphere, aquatic movement toward geomagnetic North is also downward movement toward deoxygenated water. Thus, such "compass needles" have been selected for because they steer anaerobic bacteria to deoxygenated water. This hypothesis is also supported by the fact that aquatic bacteria in the Southern Hemisphere contain reversed polarity, south-seeking "compass needles," which orient them toward geomagnetic South and, hence, downward toward deoxygenated water (see Blakemore & Frankel 1981).

Dretske claims that the magnetosome chain functions *to indicate*, but that it is indeterminate as to whether it indicates deoxygenated water, geomagnetic North, or merely magnetic North (1988, 63 n. 6). Millikan similarly views the magnetosome chain as representational, but claims that what it represents "is univocal; it represents only the direction of oxygen-free water," since it is only movement toward deoxygenated water that benefits the bacterium, not simply movement toward (geo)magnetic North (1993, 93). I will bypass issues about what the magnetosome chain indicates or represents (if anything) and focus only on its function in the behavioral economy of the bacterium. For, since both Dretske and Millikan fix the representational content of the chain from its function in the bacterium's behavioral economy, a specification of that function is prior to a specification of what it represents; so it is possible to focus on that function independently of issues about representation. The positions of Dretske and Millikan are interesting here because they correspond to viewing the chain as functionally indeterminate (producing orientation toward magnetic North, geomagnetic North, or deoxygenated water) or as determinately functioning to orient the bacterium toward deoxygenated water. But neither of these positions is correct.

To see why, we should focus not only on cases in which everything goes swimmingly, but also on cases of breakdown of the widest system embedding the mechanism. When we do, we find a couple notable phenomena. First, hemispheric displacement of a bacterium has fatal results. For example, a Northern bacterium transplanted into the Southern Hemisphere swims upward into oxygenated water and dies, since in the Southern Hemisphere its north-seeking magnetosome chain orients it upward. Second, even in its home hemisphere, a bacterium can be made to swim toward oxygenated water simply by passing a bar magnet overhead.

When breakdowns like these occur, we get the widest system up and running again simply by restoring a broken connection between the magnetosome chain and the wider environment. Since we do not repair the magnetosome chain itself, the magnetosome chain is performing its function even when that does not effect orientation toward deoxygenated water (first breakdown) or geomagnetic North (second breakdown). Indeed, the only breakdown in which we would repair the magnetosome chain itself is one in which it has become depolarized; and the effect that it produces in all cases but this is that of orienting the bacterium toward magnetic attraction. In addition, as both types of breakdown show, the magnetosome chain does not control for orientation toward either deoxygenated water or geomagnetic North; the only output the chain actually controls for is orientation toward mag-

netic attraction. These facts conjointly imply that the function of the magnetosome chain is *only* to orient the bacterium toward magnetic attraction. But even this is too general. For there may be more than one source of magnetic attraction in the environment. How would we specify the relevant attraction? We could attempt to do so by describing the field gradient at some arbitrary (yet small) distance *D* from the surface of the bacterium's cell. But this will pinpoint the relevant attraction only if the field gradient *at* the surface of the cell is the same—only if there is no disturbance between *D* and the cell's surface. So the relevant attraction is actually the one that impinges most strongly on the surface of the cell. Thus, the magnetosome chain has the single, determinate function of orienting the bacterium toward the strongest magnetic attraction impinging on the surface of its cell. Since this is a proximal output, we have arrived at a wholly individualistic functional individuation of the mechanism. This individualistic mechanism does, however, serve to increase the fitness of its possessors by functioning to initiate a series of effects that (usually) culminates in movement toward deoxygenated water. And information about the function *of its output* explains *why* the mechanism with *that* individualistic function was selected for.

4. Conclusion. If these arguments are right, then evolution by natural selection "designs" proximate mechanisms in accordance with a principle of cognitive economy that Clark calls "the 007 principle." As Clark puts it:

> evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it [instead]. . . . That is, know only as much as you need to know to get the job done. (1989, 64)

It is easy to see why natural selection would have employed this principle. In the case of the chameleon, controlling for changing color if (and only if) a predator is present would be no more effective than simply changing color without first checking for the presence of a predator. Indeed, in general, a mechanism that controlled for distal conditions of the environment would be costlier to "build" than one that controls for only proximal conditions. And, since Normal conditions provide correlations between the distal conditions of the environment and the proximal conditions of a proximate mechanism, the costs involved in "building" such a mechanism would outweigh the payoffs. So natural selection finds it most cost effective to "design" mechanisms that control for ("know about") only proximal inputs and outputs.

The mechanisms need do no more than that, for Normal conditions do the rest.

These conclusions are drawn, however, only about mechanisms that have become fixed as a result of evolution by *natural selection*. Some theories (e.g., the "neural Darwinism" of Edelman 1992) maintain that many brain mechanisms become fixed as a result of neural selection during an *individual's history*. These types of mechanism, for all I have argued, may always require non-individualistic functional individuation. My arguments have shown to be functionally individualistic only those mechanisms present in individuals of a current population because they conferred adaptive advantage at some point in the evolutionary *history of a species*. These are the only mechanisms of which evolutionary psychology provides explanations. Thus, insofar as there are behavior control mechanisms that are functionally non-individualistic, explaining their functioning devolves to other research programs in psychology.

REFERENCES

Alcock, J. (1993), *Animal Behavior: An Evolutionary Approach* (5th ed.). Sunderland, MA: Sinauer.

Amundson, R. (1994), "Two Concepts of Constraint: Adaptationism and the Challenge from Developmental Biology", *Philosophy of Science* 61: 556–578.

Blakemore, R. P. and R. B. Frankel (1981), "Magnetic Navigation in Bacteria", *Scientific American* 245(6): 58–65.

Clark, A. (1989), *Microcognition*. Cambridge, MA: MIT Press.

Cosmides, L. & J. Tooby (1987), "From Evolution to Behavior: Evolutionary Psychology as the Missing Link", in J. Dupré (ed.), *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, MA: MIT Press, pp. 277–306.

Cummins, R. (1975), "Functional Analysis", *Journal of Philosophy* 72: 741–765.

Dawkins, M. S. (1995), *Unravelling Animal Behaviour* (2nd ed.). Essex, UK: Longman Scientific & Technical.

Dawkins, R. (1989), *The Selfish Gene* (2nd ed.). Oxford: Oxford University Press.

Dretske, F. (1988), *Explaining Behavior*. Cambridge, MA: MIT Press.

Edelman, G. M. (1992), *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books.

Endler, J. A. (1986), *Natural Selection in the Wild*. Princeton: Princeton University Press.

Goode, R. and P. E. Griffiths (1995), "The Misuse of Sober's Selection for/Selection of Distinction", *Biology and Philosophy* 10: 99–108.

Griffiths, P. E. & R. D. Gray (1994), "Developmental Systems and Evolutionary Explanation", *Journal of Philosophy* 91: 277–304.

Maynard Smith, J. (1982), *Evolution and the Theory of Games*. New York: Cambridge University Press.

———. (1993), *The Theory of Evolution* (Canto ed.). Cambridge: Cambridge University Press.

Mayr, E. (1988), "An Analysis of the Concept of Natural Selection", in *Toward a New Philosophy of Biology*. Cambridge, MA: Harvard University Press, pp. 95–115.

Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

———. (1993), *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Neander, K. (1995a), "Misrepresenting and Malfunctioning", *Philosophical Studies* 79: 109–141.
———. (1995b), "Pruning the Tree of Life", *British Journal for the Philosophy of Science* 46: 59–80.
Sober, E. (1984), *The Nature of Selection*. Cambridge, MA: MIT Press.
———. (1995), "Natural Selection and Distributive Explanation: A Reply to Neander", *British Journal for the Philosophy of Science* 46: 384–398.
Tooby, J. (1988), "The Emergence of Evolutionary Psychology", in D. Pines (ed.), *Emerging Syntheses in Science*. Santa Fe: Santa Fe Institute, pp. 67–76.
Wilson, R. A. (1994), "Causal Depth, Theoretical Appropriateness, and Individualism in Psychology", *Philosophy of Science 61*: 55–75.
Young, J. Z. (1981), *The Life of Vertebrates* (3rd ed.). Oxford: Oxford University Press.