Perverse Engineering*

Chris Haufe†‡

Evolutionary psychologists, among others, have used a method called "reverse engineering" to uncover (*a*) whether a trait was selected for, and (*b*) if so, why that trait was selected for. In this paper I argue that reverse engineering cannot deliver on either (*a*) or (*b*), and tends to pervert, rather than enhance, our knowledge of natural history. In particular, I expose as false a fundamental assumption of reverse engineering—namely, that all traits selected for a particular function will share some nontrivial properties.

1. Introduction. Reverse engineering (RE) is a method for studying evolutionary history in which we infer the selective importance of some environmental factor from the existence of an organismal feature that seems ideally suited for the purposes of dealing with that aspect of the environment. The essence of RE is to think about what a trait must be like if it is to be good for doing a particular task—what "design criteria" it must fulfill—and then gauge whether the trait in question does in fact meet those criteria.

In this paper I argue that RE, particularly as it is used in the Darwinian approach to human nature known as *evolutionary psychology*, is apt to direct us away from knowledge of the selective history of an organism rather than to lead us toward it. The research tenets of RE have perverted evolutionary psychologists' (and others') sense of what kind of evidence is probative when considering the plausibility of a hypothesis which suggests that some property of an organism is an adaptation.

My argument will proceed as follows. First, I call into question the ability of RE to differentiate the effects of one selection pressure from those of another, as well as its ability to discriminate between selective

*Received March 2006; revised June 2008.

†To contact the author, please write to: Department of Philosophy, 229 Major Williams Hall (0126), Virginia Tech, Blacksburg, VA 24061; e-mail: haufe@vt.edu.

‡Many thanks to Philip Kitcher, Patricia Kitcher, and Alison Wylie for insightful recommendations. Special thanks also to Maysan Haydar for her substantive and editorial contributions.

Philosophy of Science, 75 (October 2008) pp. 437–446. 0031-8248/2008/7504-0003$10.00
Copyright 2008 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/595837 Published online by Cambridge University Press

and nonselective influences. These objections notwithstanding, I move on to attack the very notion of "design criteria," and argue that it may not be applicable to many categories of selection pressures. I close with a defense of current reproductive success as evidence for adaptation.

2. The Method. How do we uncover an organism's selective history? In evolutionary psychologist Randy Thornhill's words, we need to "hypothesize about the selection that made an adaptation, and then test predictions derived from the hypothesis. The predictions are about what must be true of the design of an adaptation if the hypothetical historical selective force was, in fact, causal" (1997, 6). This passage is excerpted from the clearest explication of RE of which I am aware, "The Concept of an Evolved Adaption" (Thornhill 1997). In this essay Thornhill defends the view that RE is the *only* way of knowing which adaptive problems were faced by our ancestors, which he believes follows from the idea that "adaptations are the biologist's sole source of information about the forces of selection that were actually effective in designing organisms during the evolutionary history of life" (1997, 5).

The argument behind Thornhill's view that adaptations are our only source of information about the effects of selection is fairly straightforward. First, the necessary and sufficient conditions for some factor F to be a force of selection are that F causes type A organisms to be better equipped to reproduce than type B organisms, and that A's actually do outreproduce B's. Whichever property P differentiates A's from B's with respect to F is an adaptation. Notice that whether F qualifies as a force of selection depends on whether A's outreproduce B's, *not* the other way around (i.e., rather than A's outreproducing B's depending on whether F qualifies as a force of selection). Now, because P is the only existing record that A's once outreproduced B's, P is the only source of information we have for knowing whether some factor once qualified as force of selection.

RE capitalizes on the specialness of P as our link to the past by devising ways of uncovering what kind of a force would have caused P to spread through the population; in essence, trying to discover an evolutionary question to which P is the answer. The general approach is, as Thornhill says, to try to imagine what *must* be true of a hypothetical adaptation in order for it to have been a successful way of dealing with some hypothetical selective force, and then determine whether that which must be true of a hypothetical adaptation *is* true of some organismal feature. If we find a property which meets these criteria, we can infer that the property evolved to deal with the hypothetical selective force, which of course entails that the hypothetical selective force was actualized during the relevant period (Tooby and Cosmides 1992, 2005; Williams 1992, 1996; Thornhill 2007).

https://doi.org/10.1086/595837 Published online by Cambridge University Press

3. Reversing RE. There are a number of objections we can raise against the idea that RE is likely to lead to reliable conclusions about evolutionary history. First, the focus of RE on the final product of an evolutionary process as the best source of information about that process runs a grave risk of being undercut by the problem of multiple realizations—specifically, the possibility that selection for different types of functions could give rise to the same form. For, if what must be true of an adaptation to *F* (which ex hypothesi humans encountered) might also be true of an adaptation to *G* (which ex hypothesi humans did not encounter), then the fulfillment of those criteria does not by itself allow us to discriminate between traits that are adaptations to *F* and those that are adaptations to *G*. Since RE does not incorporate any other means of discriminating between rival hypotheses regarding historical selection pressures, the success of RE as a strategy for reconstructing human evolutionary history is entirely dependent upon our ability to correctly guess which historical environmental factors had a selective influence on our ancestors.

How likely is it that we will be able to do this? We do not have the luxury of inferring our selective history by direct comparison with other taxa without already having lots of relevant historical information (which we do not have). Nor can we appeal to selection pressures which *all* possible organisms will face, because there is nothing to suggest that there are any such nontrivial selection pressures. We could try a slightly watered-down version of this appeal, though. We could argue, for example, that our chances of correctly guessing which selection pressures were faced by our ancestors are pretty good because all possible organisms will face certain *categories* of selection pressures. Thus, we do not need any specific information about, say, the early hominin selective environment. We can simply apply the "principles that shape species" to our own, and this will furnish us with knowledge of which selection pressures produced our adaptations.

There is a part of this suggestion which, I think, is basically right. All possible organisms *will* encounter many of the same categories of selection pressures, and the more narrowly one restricts his attention to the tree of life, the more likely organisms are to share categories. Survival and reproduction are obvious ones, and there are even subcategories which we can identify for lots of families of organisms. Many organisms will face foraging problems, or mating problems, for example. Where the appeal to shared categories of problems breaks down, however, is when we try to extend that principle to the level of individual problems within a category. The specific problems for a particular category encountered by different populations are likely to be both novel and highly idiosyncratic in many cases, such that they may never have arisen before and may never arise again.

https://doi.org/10.1086/595837 Published online by Cambridge University Press

Consider the role of sensory bias in mate choice. The selective effects of sensory bias have serious potential to be both pervasive and very powerful. Work on sensory bias has revealed is that the biases of different taxa do not follow any sort of predictable pattern, that they will tend not to resemble each other, and that they do not bear any lawful or principled relation to the environment in which they arise. These are all properties one would expect from the accidental effects of neural organization, which is where sensory biases reside. Moreover, the ubiquity of adaptations of one sex to the sensory biases of the other suggests that this specter is not a remote logical possibility, but is in fact very common across taxa. Nor are the adaptations to sensory biases trivial aesthetic modifications of the organism. Often they have harsh consequences for survival, as is suggested by the backlash effect found in instances of Fisherian runaway. So it won't due to reply that adaptations to sensory biases are marginal in terms of their impact on the evolution of the species; they are not. The prospect of idiosyncratic selection pressures like sensory bias presents a fundamental obstacle to our hopes of guessing which selection pressures early hominins were likely to have faced.

Research on the genus *Xiophophorus* represents a telling example of how the application of so-called "principles that shape species" can mislead. One "principle" to which evolutionary psychologists have been partial is the idea that females will be selected to prefer the fittest males. The alleged principle would have been of no use (or worse) in determining the origin of the female swordtail (*X. helleri*) preference for males with longer swords (a part of the tail). Whereas the application of "fundamental evolutionary laws" (Tooby and DeVore 1987, 189) might have easily led to the idea that female swordtails evolved to prefer longer swords because longer swords indicate "good genes," it turns out that the female preference for longer swords evolved *prior* to swords themselves. This was determined by demonstrating that females from evolutionarily prior taxa in which swords did not exist *also* prefer males with longer swords. The RE approach suggests that we could have discovered the evolutionary origin of female swordtails' mate preferences by simply reflecting on the ecological reasons for why preferences might evolve. But the principles of mating ecology cannot tell us anything about the reasons for the female swordtails' preference, because those principles were not, so it appears, interestingly involved in determining that feature of female swordtails (Sinervo and Basolo 1996, 174–175).

The fact that sensory biases are both idiosyncratic and important means that there is a least one category of adaptive problems for which our ability to correctly guess the specific instances across large swaths of evolutionary time is basically nonexistent. Unlike the common response to adaptationism that appeals to idiosyncratic *solutions* to problems (see

https://doi.org/10.1086/595837 Published online by Cambridge University Press

below), there is no refuge for the adaptationist in the face of novel selective *problems*. As we said above, the success of RE is entirely dependent on our ability to guess which selective problems were faced by our ancestors. The existence of novel, short-lived, and idiosyncratic problems suggests that this ability is woefully deficient[1]—certainly deficient enough to warrant abandoning any research strategy which depends upon it.[2]

A related difficulty for RE concerns its ability to discriminate between properties that are adaptations per se and properties that are not. In the same way that what *must* be true of any adaptation in order for it to successfully deal with *F* may also be true of an adaptation designed to deal with *G*, it is also plausible to think that what must be true of adaptations to *F* may also be true of properties of organisms that are not adaptations at all. The failure of RE to discriminate between adaptations and nonadaptations is rooted in the same feature which is responsible for its failure to discriminate between different kinds of adaptations—i.e., the fact that something could fulfill the design criteria for adaptations to *F* and yet fail to be an adaptation to *F*.

Tim Lewens (2002) recognizes the inability of the question, "what *must* be true of any adaptation in order for it to successfully deal with *F*" to pick out *only* adaptations to *F*. He states that "for the inference to be watertight" would require that some adaptive solution *S* "is adopted when and only when" *F* is a genuine selection pressure (2002, note 10). Lewens' suggestion implies that the failure of this strategy to pick out *only* adaptations to *F* is due to the fact that it leaves the door open for adaptations which satisfy the design criteria delineated for *F* but that are in fact adaptations to some other selection pressure *G*. Although this is true, it is too permissive. The problem with Lewens' recommendation is that, while it shores up the ability of RE to differentiate adaptations to *F* from those that are not, it still will not allow us to distinguish between adaptations which match the necessary design criteria from suites of properties which also match the design criteria but which are not themselves adaptations.

This is not a minor quibble. Indeed, the problem I attribute to Lewens' formulation lies at the very heart of complaints surrounding adaptation-

1. There is much to recommend the appeal to novel adaptive *problems* as a general worry for adaptationism. Anti-adaptationists would do well to focus on developing arguments geared towards categories in which it is known that adaptive problems tend to take on very unique forms.

2. Interestingly, the father of modern RE, George C. Williams, has commented that "every organism will show a long list of characters that make no adaptive sense but record past adaptations" (1992, 76). This is essentially the point I have been making, and it is rather difficult to see how this fact can be reconciled with the alleged reliability of RE.

https://doi.org/10.1086/595837 Published online by Cambridge University Press

ism. RE allegedly has the ability to simultaneously demonstrate both the *presence of* and *reasons for* selection. Lewens' comment suggests that the problem with RE is that it has the potential only to confuse the *reasons* for selection, while leaving intact the ability of RE to correctly identify the *presence* of selection. Lurking behind his formulation is that idea that looking for conformity to a priori design criteria is a good way to identify adaptations per se, it's just not a very good way to know *why* a population "adopted" a particular "solution." But in its canonical form, RE is no better at picking out the *presence* of selection than it is the *reasons* for selection. In order to be "watertight," the inference from the presence of properties which conform to a priori design criteria to the conclusion that the organism has an adaptation to *F* requires not merely that a population *adopt* solution *S* when and only when faced with *F*, but that properties which conform to a priori design criteria *appear* when and only when a population is faced with *F*. Where the latter biconditional holds, embodying design criteria could, in principle, sanction inferences concerning the presence of and reasons for selection.

4. Are There Design Criteria? A third, more fundamental objection to RE is to raise questions concerning the very idea of design *criteria*. Some authors have criticized RE on the basis that the same adaptive problem "can be solved in several different ways" (Griffiths 1996, 517–519). Although Griffiths is lamentably unclear on how this suggests problems for RE, one might guess from his example (alligators and anacondas are very differently adapted fresh water predators) that he imagines that the often very different solutions to the same adaptive problem may lead researchers to mistakenly rule out some trait as an adaptation to selection pressure *F* when in fact it *is*. But while it is undoubtedly important to point out that there are many ways to solve an adaptive problem, as a response to RE this is somewhat wide of the mark. For, evolutionary psychologists (and any other adaptationists) can simply reply (and, indeed, have replied) that different local solutions do not negate the fact that "at some level" these seemingly different solutions can be characterized using the same functional description, and it is at that level that we should be looking (e.g., Tooby and Cosmides 1992).

Here is an alternate approach. Stated another way, the question "what must be true of an adaptation designed to deal with selective force *F*" asks, "what do all possible adaptations caused by selective force *F* have in common?" The appropriate response here seems to me to be that it is possible that the answer to this question is "Nothing." RE assumes that, no matter *which* way an organism adapts to *F,* and no matter what its phylogenetic history and constraints are, there will be some nontrivial level at which the organism's adaptation to *F* is similar to all other ad-

https://doi.org/10.1086/595837 Published online by Cambridge University Press

aptations to *F*, whether these organisms evolved on our planet or in some other spacetime dimension. But what reason do we have to believe this? Of course, it *might* be true. Whether it is or not, if RE is going to be able recommend itself as a research strategy, it *has* to give a positive reason for why we should expect that all possible adaptations designed to solve adaptive problem *F* will have *something*—let alone *many* things—in common. The mere presence of functional similarities among distantly related organisms living in similar habitats (aka "convergence") is not sufficient to infer that, for any given selection pressure, all possible organisms will have some design features in common.

We do not as of yet have any reason to think that all possible adaptations caused by a certain environmental factor will ipso facto share a suite of properties. Thus, the identification of conformity to certain design criteria is, for all we know, not *necessary* for identifying adaptations for that feature of the environment. Furthermore, it is plausible to suppose that adaptations for different types of problems could all call for the same criteria to be fulfilled, so identification of conformity to design criteria is not *sufficient* for identifying adaptations caused by a particular environmental factor. Since RE is neither necessary nor sufficient for demonstrating adaptation, it is a fortiori incapable of revealing a trait's selective history.

5. Design Criteria vs. Reproductive Success. The final blow to RE is delivered in the context of its proponents' assertions that conformity to design criteria is stronger evidence for selection than reproductive success differentials. Thornhill (1997, 2007) argues for the proposition that "the only way you could be sure that you had identified an EEA feature that mattered in terms of generating reproductive success differentials is to find evidence for its existence in the design of the adaptation" (Thornhill 1997, 15). In this regard he explicitly juxtaposes the significance of conformity to design criteria with the *insignificance* of measures of the contribution of a trait to current reproductive success as evidence for selection, a view in which he is joined by the majority of evolutionary psychologists (Cosmides and Tooby 1987; Tooby and Cosmides 1990; Williams 1992, 40).

There are two serious problems for the juxtaposition of the usefulness of conformity to design criteria with the uselessness of current contributions to reproductive success. First, it may be true that whether an historical environmental force *F* was selective depends solely on whether there are (or were) adaptations for it. However, it is quite another thing to say that the design of extant adaptations (or, more accurately, properties which conform to design criteria) are the "only evidence that support . . . hypotheses" concerning *F*'s selective efficacy (Thornhill 1997, 15). Second,

https://doi.org/10.1086/595837 Published online by Cambridge University Press

the problems which Thornhill and others have raised for measures of current reproductive success apply equally to perceptions of conformity to design criteria.

Lots of biologists take demonstrations of selection in the form of current differential contributions to reproductive success as very good evidence for selection, and I presume Thornhill would agree with them (see Endler 1986). But the value of these kinds of demonstrations can go beyond what they tell us about current reproductive success. There is also very good evidence for what *would* happen in an historical population where the relevant parameters were instantiated. It is easy to interpret John Endler's (1980) study on guppies within this framework. Endler demonstrated experimentally via reproductive success differentials that one color of guppies will outreproduce another under certain predatory conditions (specifically, conditions similar to those of their native habitat). This experiment is rightly interpreted as supportive of the hypothesis that an adaptive relation between coloring and predation explains the native trait frequency distributions, because it shows us what *would* happen given native conditions and thus, what *might plausibly* have produced the native distributions. Scores of similar experiments attest to the widely held view that these kinds of tests are good for establishing rules of inference, and the reasoning linking the structure and results of these experiments to hypothetical historical states of affairs suggests that these kinds of tests actually *are* good at underwriting certain rules of inference.[3]

Evolutionary psychologists have attacked the use of—or insistence that they themselves provide (Thornhill 2007)—data on current reproductive success as being insignificant to whether some organismal property P is an adaptation to an historical environmental factor F because of the fact that P could be an adaptation to F without P contributing positively to current reproductive success. It is undoubtedly correct that current reproductive success is conceptually distinct from the historical facts about selection, due primarily to the possibility of environmental differences between the past and present. But it is *equally* true that whether something conforms to a priori design specifications is conceptually distinct from the historical facts about selection. If, as Thornhill (2007) says, "a scientific prediction is one that must be true if the hypothesis generating it is true," then a hypothesis of conformity to a priori design criteria, "improbable functionality," "complexity," or any other heuristic in the evolutionary psychological arsenal, is no more "scientific" than hypotheses concerning

3. And it doesn't stop there. There are many other categories of evidence which can be and are routinely brought to bear on selectionist hypotheses. We can, for instance, have independent evidence that genetic drift or pleiotropy played a minor or insignificant role in the fixation of a trait (Abrams 2001, 292).

https://doi.org/10.1086/595837 Published online by Cambridge University Press

current contributions to reproductive success. Arguably, data on current contributions to reproductive success are a far better source of evidence about selection—past or present—than conformity to design criteria, if for no other reason than the fact that the former are not reliant upon the researcher's intuitions to the extent of the latter (Williams 1992, 41). Furthermore, we have a rich empirical and theoretical tradition informing our use of reproductive success to learn about selection. There are no corresponding credentials for a priori design criteria.

6. Conclusion. The preceding discussion has been aimed at establishing the unreliability of RE as a method for learning about the historical selection pressures faced by our ancestors. The reliability of RE consists of the conjunction of our ability to accurately specify a priori what must be true of an adaptation if it evolved in response to a particular selective challenge, combined with our ability to correctly guess which selective challenges may have been faced by a species in the past. The first conjunct is undermined by the fact that there may be no properties which are shared by all possible solutions to a particular adaptive problem. The second is undermined by the fact that our ability to correctly guess historical adaptive problems may be irremediably poor, owing to the often random, idiosyncratic nature of these problems.

REFERENCES

Abrams, Peter (2001), "Adaptationism, Optimality Models, and Tests of Adaptive Scenarios", in Steven Orzack and Elliot Sober (eds.), *Adaptationism and Optimality*. New York: Cambridge University Press, 273–302.

Cosmides, Leda, and John Tooby (1987), "From Evolution to Behavior: Evolutionary Psychology as the Missing Link", in John Dupré (ed.), *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, MA: MIT Press, 276–306.

Endler, John A. (1980), "Natural Selection on Color Patterns in *Poecilia reticulata*", *Evolution* 34: 76–91.

——— (1986), *Natural Selection in the Wild*. Princeton, NJ: Princeton University Press

Griffiths, Paul (1996), "The Historical Turn in the Study of Adaptation", *British Journal for the Philosophy of Science* 47: 511–532.

Lewens, Timothy (2002), "Adaptationism and Engineering", *Biology and Philosophy* 17: 1–31.

Sinervo, Barry, and Alexandra Basolo (1996), "Testing Adaptation Using Phenotypic Manipulations", in Michael Rose and George Lauder (eds.), *Adaptation*. San Diego: Academic Press, 149–185.

Thornhill, Randy (1997), "The Concept of an Evolved Adaptation", in Gregory Bock and Gail Cardew (eds.), *Characterizing Human Psychological Adaptations*. London: Ciba Foundation, 4–22.

——— (2007), "Comprehensive Knowledge of Human Evolutionary History Requires Both Adaptationism and Phylogenetics", in Steven Gangestad and Jeffry Simpson (eds.), *The Evolution of Mind.* New York: Guilford, 31–37.

Tooby, John, and Leda Cosmides (1990), "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments", *Ethology and Sociobiology* 11: 375–424.

https://doi.org/10.1086/595837 Published online by Cambridge University Press

——— (1992), "The Psychological Foundations of Culture", in Jerome Barkow, Leda Cosmides, and John Tooby (eds.), *The Adapted Mind*. New York: Oxford University Press, 19–136.
——— (2005), "Conceptual Foundations of Evolutionary Psychology", in David Buss (ed.), *Handbook of Evolutionary Psychology*. Hoboken, NJ: Wiley, 5–67.
Tooby, John, and Irven DeVore (1987), "The Reconstruction of Hominid Behavioral Evolution through Strategic Modeling", in Warren Kinzey (ed,), *The Evolution of Human Behavior: Primate Models.* Albany, NY: SUNY Press, 183–237.
Williams, George C. (1992), *Natural Selection: Domains, Levels, and Challenges*. New York: Oxford University Press.
——— (1996), *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.

https://doi.org/10.1086/595837 Published online by Cambridge University Press