“Two” Many Optimalities

Òscar Vilarroya (24678ovo@comb.es)
Centre de Recerca en Ciència Cognitiva
Gran de Gràcia, 127, 2-1; 08012 Barcelona, Spain

Abstract

In evolutionary biology a trait is said to be optimal if it maximizes the fitness of the organism, that is, if the trait allows the organism to survive and reproduce better than any other competing trait. In engineering, a design is said to be optimal if it complies with its functional requirements as best as possible. Cognitive science is both a biological and engineering discipline and hence it uses both notions of optimality. Unfortunately, the lack of a clear methodological stance on this issue has made it common for researchers to conflate these two kinds of optimality. In this paper I argue that a strict distinction must be kept in order to avoid inaccurate assumptions.

Cognitive Explanations

Contemporary biological explanations are teleonomic explanations. Teleonomic explanations are those explanations that treat biological traits as adaptations. Adaptations are innovations that make a difference between alternative biological designs embodied in different individuals. Such designs are “chosen” from among alternative designs on the basis of how well they perform in a given environment.

The phenotype of an organism is therefore understood as a collection of functionalities that were added and maintained (i.e., copied over generations) in the design features of a given species *because* these functionalities had the consequence of solving problems that promoted, in some way or other, survival and reproduction (Millikan, 1984). A complete teleonomic explanation would then consist of a step-by-step process in which the scientist must:

(A) Teleonomic Explanation

1. Identify the trait that is likely to be under selection.
2. Identify the adaptive problem that the trait is supposed to solve.
3. Show:
   (a) that the trait is specialized for solving the adaptive problem;
   (b) that it is unlikely to have arisen by chance alone; and
   (c) that it is not better explained as the by-product of mechanisms designed to solve some alternative adaptive problem.
4. Establish the fitness value of the trait in the population.[1]

This sort of story could be the general strategy for *any* discipline which required teleonomic explanations. However, cognitive science has certain particularities that change this methodology. In explaining the cognitive mechanisms of biological organisms, the cognitive scientist may attempt to identify the adaptive problem that the brain is supposed to solve. In reality, however, this turns out to be quite difficult, because the trait—that is, the specialized device within the brain that is responsible for solving the adaptive problem—is not as self-evident as, say, an eye, a liver or a wing. Specialized cognitive mechanisms lie within the circuitry of the brain in such a way that makes them not as obvious as one would like (Barkow, Cosmides & Tooby, 1992).

In order to overcome this problem, cognitive scientists usually invert the first and second steps of the algorithm above:

(B) Cognitive Explanation

1. Identify the adaptive problem that the organism is supposed to solve.
2. Presuppose the trait that is likely to be under selection.
3. Show:
   (a) that the trait is specialized for solving the adaptive problem;
   (b) that it is unlikely to have arisen by chance alone; and
   (c) that it is not better explained as the by-product of mechanisms designed to solve some alternative adaptive problem.
4. Establish the fitness value of the trait in the population.

Inverting the two first steps of algorithm (A) cannot be without consequences. The most obvious is the fact that identifying the function of a trait, such as a wing

[1] In other words, the theorist must establish that the distribution of the trait in the population contributes to the evolutionary notion of fitness, which for present purposes means simply the capacity to survive and reproduce.

or an eye, is much easier when the trait has been identified rather than presupposing the trait when we only know its function. In the first case we can simply observe the trait at work or determine if it is rarely used, etc., or perhaps we might test it under all imaginable circumstances just to see what it does. Cognitive scientists face a much more difficult enterprise, however, since we have to "imagine" the design features of the trait. This is where the reverse-engineering strategy comes in.

Optimality in Reverse-Engineering

Dennett (1995) is perhaps one of the most adamant about defending the reverse engineering strategy in cognitive science. This strategy can be defined as the interpretation of an already existing intelligent artifact or system through an analysis of the design considerations that must have governed its creation. Logically, this overidealizes the design problem, because it presupposes that the trait is *optimally* executed by the cognitive machinery. Thus, reverse-engineering takes cognitive systems to be systems that are designed to solve the problem identified by the theorist; otherwise, the analysis could not get off the ground. As Dennett observes, if cognitive scientists cannot assume that there is a good rationale for the features they observe in cognizers, they cannot even begin their task. Optimality must be the default assumption in cognitive explanations.

A standard way of advancing the reverse-engineering strategy is to resort to Cummins' notion (1983) of functional analysis. Basically, a functional analysis amounts to:

1. System S performs F
2. F can be broken down into $f_1, f_2, f_3 \ldots f_n$
3. S implements $f_1, f_2, f_3 \ldots f_n$

The first step is therefore to establish F, that is, what the system does. The usual way in which one characterizes F is to call upon the computational theory proposed by Marr (1982). In this framework, the theorist must provide an abstract formulation of the information-processing task that defines a given cognitive ability. Peacocke (1986), for example, describes such formulations as characterizations of the information state that the system draws upon.

Now, the notion of "information drawn upon" can be spelled out as follows:

> A state draws upon some information whenever such state carries the information which is causally influential in the operation of the algorithm or mechanism. (Peacocke, 1989, p.102).

Given this definition, explanations based on Peacocke's notion can be seen as a fully causal. For instance, facts about the meaning, syntactic structure, and phonetic form of linguistic expressions are causally explained by facts about the information drawn upon by algorithms or mechanisms in the language-user (ibid., p.113). Thus, if a system draws upon the correct information, then the explanation of system is correct regardless of the detailed algorithm that the system uses.

Be this as it may, the computational characterization of a problem determines the specifications by which the reverse-engineering must proceed. These specifications are taken at face value and are considered sufficient to establish the design features of the mechanism. Cummins (1983) further argues that such characterizations should proceed by decomposing them into a set of simpler capacities that are to be explained by subsumption. The overall capacity is thus explained in terms of the contributing capacities of parts of the system, and the function of a given item is its contribution to the overall capacity. Such design theories of functions define the function of a mechanism or process in terms of the roles they might play, that is, in terms of their contribution to some capacity of the system to which the process or mechanism belongs. In short, design theories relativize functions to capacities of containing systems.

Optimality in Evolutionary Biology

The previous sense of optimality must be distinguished from the notion of optimality that is used in evolutionary biology. To begin with, the biological notion of an optimum does not imply an optimal design, as it does in engineering. Rather, it refers to a solution to a given adaptive problem that maximizes the fitness of the organism in the adaptive situation.

A biological optimum can be said to be the point at which the difference between costs and benefits of environmental and genetic variables (e.g., amount of food, energy requirements, distribution of the trait, alternative phenotypes, etc.) is maximized (see Figure 1). Thus, the role of a trait as an adaptation must be established by considering the manner in which such a trait contributes to the optimum. This makes the notion of optimality in biology and engineering orthogonal to one another:

Biological optimality: Natural selection favors the trait that maximizes the organism's *fitness*.

Reverse-engineering optimality: A mechanism is *designed* to comply with its function.

The fact of the matter is that a trait need not be optimally designed to be adaptive: what is optimal is the fitness value of the trait, not its design characteristics. Evolutionary solutions must, on this view, only be "selectively efficient," that is, they need only to comply with adaptive requirements. It follows, then, that the notion of optimal design should be detached from the notion of fitness value.

That natural selection is only susceptible to the fitness-value maximizations is anything but surprising: evolution by natural selection only requires that biological systems be minimally effective (stay alive and leave offspring) with respect to their of adaptive problems. Hence, rather than actively designing and building organisms that are well-adapted to the world, nature eliminates

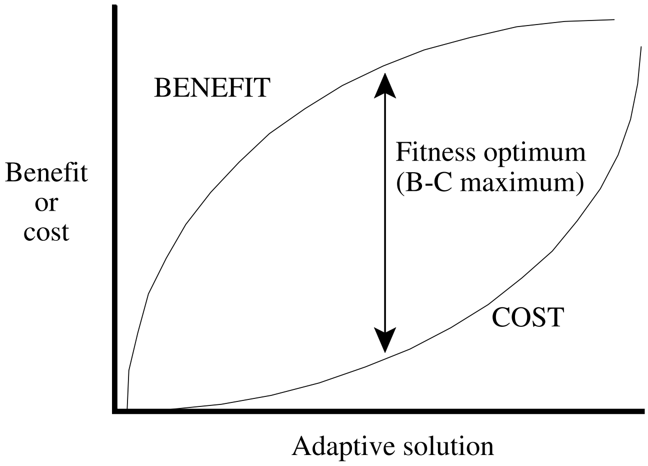

Figure 1: The optimum for a given adaptive solution is the point in which the benefit/cost relationship is maximized.

those that are too ill-suited for survival and reproduction. In other words, biological systems are simply not designed by engineers. Design and evolution are different precisely because they have different strategies open to them. An engineer may build a system out of an analysis of the problem, and thus may go from the problem to the solution. This is not a possibility for biological systems: biological systems are blind to the solution until they have stumbled upon it.

The bottom line, then, is that we should explain how cognitive systems are selected for and maintained by taking into account not only the adaptive problem itself, but also their resources and the environment in which they are evolving.

Conflating the Two Optimalities

The previous discussion makes it advisable to maintain the engineering and the biological optimalities separate. However, some cognitive scientists (e.g., Barkow, Cosmides & Tooby, 1992) seem to conflate both notions. According to these theorists, what must guide the design specifications of cognitive mechanisms is a computational characterization (Marr, 1982) of the adaptive problems that these mechanisms were meant to solve. These specifications are considered sufficient to establish the design features of the mechanism:

> In effect, knowledge of the adaptive problems humans faced, described in explicitly computational terms, can function as a kind of Rosetta Stone: It allows the bewildering array of contents effects that cognitive psychologists routinely encounter—and usually disregard—to be translated into meaningful statements about the structure of the mind. (Cosmides & Tooby, 1992, p.221)

In other words, these theorists take the brain to be the seat of specialized mechanisms that are optimally designed, in an engineering sense, to solve specific adaptive problems.

It is of course possible for engineering and biological characterizations to coincide for a given trait and a given organism. Nonetheless, counterexamples abound (e.g., Ullman, 1996; Steels, 1994; Dehaene, 1997; Cooper & Munger, 1993;). Consider the well-known example of the sight-strike-feed mechanism of the frog (Gilman, 1996). Frogs catch flies by way of a strike with their tongue. It is assumed that mediating between the environmental presence of a fly and the motor response of the tongue strike there is some sort of mechanism that registers the fly's presence in the vicinity of the frog. That is, the presence of the fly causes the relevant mechanism to go into state *S*, and its being in state *S* causes the tongue to strike.

This story goes on to assume that the information drawn upon by state *S* is that of "fly", "fly, there," or "edible bug, there," since this information can be derived from that the fact that the function of the frog's sight-strike-feed mechanism mechanism is to detect the presence of flies. Yet, an analysis of the frog's cognitive system indicates that the best account of the system's function is in fact detecting "little ambient black things." Specifically, the function of the mechanism is to mediate between little ambient black things and the frog's tongue strike.

This means that the frog's mechanism is functioning optimally even when the frog strikes at a little ambient black thing that is not a fly but a BB-gun pellet that happens to be in the vicinity. To be sure, from a reverse-engineering point of view, the system is not optimally designed to catch flies. However, from a biological point of view, it might be the optimal system. The reason is quite simple. The cost of "fly-detecting" mechanism may outweigh the cost of eating, say, lead pellets. The guarantee that a frog with such an less-than-perfect mechanism could have survived and reproduced is provided by the contingent fact that, during natural selection, a sufficient number of little ambient black things in the frog's environment were flies (or edible bugs). The combination of the benefits (which should include adequate feeding) with the costs (which should include design-building costs) shows that a better design need not mean better fitness, which is what would be predicted if only design were considered. Accordingly, in some situations a better design can actually mean a drop in fitness.

It might be objected that the fact that frogs also flick their tongues out at little black things that happen not to be flies is an empirical discovery and, hence, either sense of optimality could have been wrong about what frogs would do when confronted with BB-gun pellets. For example, an evolutionary biologist might predict that natural selection would favor the trait if the trait were specifically tailored to fly catching, since this would maximize fitness. Yet this prediction would have been wrong. A reverse-engineering perspective, by contrast, might well have made the correct prediction: striking at little black things that are not flies might be seen as an acceptable amount of noise, and not necessarily an unoptimally designed fly-catching mechanism. Such being the case, the problem, it might be argued, does not really have anything to do with the two different notions of optimality, but rather with the claims one is making about a particular trait and what sort of evidence should be used in evaluating those claims.

It seems to me that such an objection would miss the point of the argument, which is to uncover two different methodological strategies, and not competence in hypothesis formation. I will illustrate the problem with another example. Peacocke (1993) has argued that the use of particular kinds of physical principles is constitutive of the capacity of normal mature subjects to reason about and predict object motions. Such a constitutive basis is held to underlie the remarkable precision of our perceptual systems in extracting and using the motion of objects in space. Examples of this capacity include our ability to anticipate the trajectories of objects in order to intercept, follow, or avoid them.

As is well known, two general types of information are used in classical physics to describe the behavior of moving objects. On the one hand, kinematic information describes the pure motion of bodies without regard to mass (i.e., the position, velocity and acceleration of an object). On the other, dynamics describes the forces causing movement or acting on objects with mass. According to Peacocke (ibid.), in order to qualify as being able to reason about objects, we must attribute to humans the capacity to reason according to dynamic principles. This would correspond to what I have described as the task characterization, which is a normative description: it is what the system must do in order for its behavior to be selectively efficient (e.g., avoid falling stones).

If we employ a reverse-engineering strategy, the task characterization of reasoning and predicting object motions (*qua* dynamic computation) will be all that we need to analyze the system that accounts for such a capacity. If, on the other hand, we employ an evolutionary strategy, then we will have to develop a model of adaptation (see, for example, Parker & Maynard Smith, 1990). Such a model will have to consider competing alternatives that exist in an adaptive scenario. For instance, we can assume that we should evaluate the performance of a system that predicts object motion according to kinematic variables, and another according to dynamic variables. This comparison should establish the performance of each system, not in isolation but as a part of the whole organism-environment interaction. Once this is done we will be able to consider the costs of either system, in terms of design and computation.

It is very conceivable that this model might yield an outcome that is very different from the reverse-engineering analysis. It could, for example, provide the hypothesis that the kinematic system is the most adaptive solution because it satisfies the task of predicting object trajectories in a way that outweighs the cost of a much more complex, yet more optimally designed, computational system that computes dynamic variables. Among other things, the errors induced by a kinematic system may not be unacceptably gross and may be easily compensated by the continuous activation of the perceptual system. This would be congruent with empirical research such as Cooper and Munger (1993).

The point, then, is this: if we had relied the reverse-engineering strategy we would not have reached the correct analysis. This is not because we would have assumed an incorrect claim but because we simply would have employed the wrong optimality strategy.

Having said this, it is no doubt true that the distinction between the engineering and biological optimalities might not be an easy matter, at least not at first blush. On the one hand, the functionality of the system (e.g., detecting flies) is amenable to both reverse-engineering and ecological analyses; on the other, it is not always clear how to establish the parameters of the fitness-maximization process that constrains adaptive cognitive traits. The latter might not be impossible to establish in cognitive science (Vilarroya, 2001). The former requires changes in algorithm (B).

Cognitive Explanation Revisited

In my opinion the explanatory strategy of cognitive science cannot be simply an inversion of the first steps of the teleonomic explanation. It is not enough to identify the adaptive problem and then infer the mechanism. Rather, we need to complement the assumption about a trait's design with a characterization of how the adaptation might have appeared over evolutionary time. Fortunately, we have the elements to proceed to the different steps necessary to complete a cognitive explanation. In order to do that, we should divide the first step of algorithm (B) into two substeps, namely:

(B′) Cognitive Explanation

1. Characterize:

(a) the adaptive problem that the organism is supposed to solve; and
(b) the fitness-maximization process.

2. Presuppose the trait that is likely to be under selection.
3. Show:

(a) that the trait is specialized for solving the adaptive problem;
(b) that it is unlikely to have arisen by chance alone; and
(c) that it is not better explained as the by-product of mechanisms designed to solve some alternative adaptive problem.

4. Establish the fitness value of the trait in the population.

The explanation should proceed as follows. The first sub-step should yield the informational-theoretic characterization (Marr, 1982) of the functional requirements needed to satisfy the adaptive problem. Specifically, we need to indicate the requirement imposed by the adaptive problem that the system should satisfy in an idealized situation.

Once this characterization has been established, then the theorist must proceed to characterize the fitness-maximization process (including the adaptive requirements that are to be satisfied by the organism in such a process). Then, the theorist should verify whether the computational characterization of the adaptive problem is compatible with the optimum established in the fitness-maximization process. If both draw upon the same information, then the characterization of the adaptive problem can be used in conjunction with reverse-engineering methodology. This should yield the assumed design specifications for the trait. If, on the other hand, the the adaptive problem's computational characterization is not compatible with the fitness-maximization optimum, then the functional requirements of the fitness-maximization process should guide the design assumptions. As the case of the frog has shown, the fitness-maximization account allowed the assumption that the trait should be designed to detect "little ambient black things," rather the one offered by the adaptive problem which would have been "fly-there."

The characterization of the fitness-maximization process in cognitive science is, unfortunately, not a straightforward operation, as I have shown elsewhere (Vilarroya, in press). It is actually a complex process because, among other things, there is an essentially open-ended set of factors that influence just where the cost-benefit curve reaches its maximum. What allows an individual, or a group of individuals, to survive and leave offspring depends precisely on their biological constitution and the exact characteristics of the surrounding environment with which they interact.

Nonetheless, the elements in this characterization are objective. Therefore one can hope to make them explicit, and thus provide an adaptive characterization of the trade-off between the costs as well as the benefits of available solutions. This will (or should) eventually yield a description of the functional account of the cognitive system, if the analyst takes into account: (a) the nature of the adaptive problem itself, (b) the analysis of the system's resources, (c) the environment and interaction with competitors, as well as (d) the way in which all these elements interact.

How can we apply this characterization in the case of the frog? I believe that there is a way to account for the paradox that the cognitive system of the frog accords with a characterization of the adaptive problem, even though it is not the characterization of the mechanisms that accounts for the adaptive capacity. For one thing, we already have an account of the adaptive problem that the system has to solve: identify flies. This is a normative description; it is what the system must do in order for its behavior to be selectively efficient. For another, we have the functional requirements that the fitness-maximization process imposes: identify "little ambient black things." The evolutionary rationale behind this solution is consistent with the assumption that the extra computational cost of taking only flies into account (over and above other small dark ambient things) arguably outweighs the small increase in accuracy that would be gained from doing so.

In sum, a system may seem to accord with a certain functionality (e.g., identifying flies) that is, in actuality, different from the description of the structure of the cognitive system (e.g., identifying "little ambient black things"). Accordingly, such a system may in fact take advantage of mechanisms that are not specifically designed to deal the problem at hand. This does not mean that the solution is somehow deficient. The fact that the fly-catching mechanism of the frog is not sensitive only to flies does not mean that it cannot identify flies. It can and it does.

Conclusion

Cognitive science uses two distinct notions of optimality: engineering optimality and biological optimality. In engineering, optimality refers to design ideals whereas, in biology, optimality refers to fitness maximization. While conflation of the two concepts is understandable, neglecting the distinction entails incurring risk of arriving at a mistaken conclusion. Fitness value and designs are therefore best analyzed separately.

Acknowledgments

I would like to thank the following friends and colleagues for their help and advice: Antoni Gomila, David Casacuberta, Joseph Hilferty, Joan Carles Soliva, Javier Valenzuela, Maria Verdaguer, and Agustín Vicente.

References

Barkow, J., Cosmides, L. & Tooby, J., (Eds.) (1992). *The Adapted Mind: Evolutionary Psychology and the*

Generation of Culture. New York: Oxford University Press.

Cooper, L.A. & Munger, M.P. (1993). Extrapolating and Remembering Positions along Cognitive Trajectories: Uses and Limitations of Analogies to Physical Motion. In Eilan, N., R. McCarthy & Brewer, B. (Eds.), *Spatial Representation*. Oxford: Basil Blackwell.

Cosmides, L. & Tooby, J. (1992). Cognitive Adaptations for Social Exchange. In Barkow, J., L. Cosmides, & Tooby, J. (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.

Dennett, D.C. (1995). Cognitive Science as Reverse Engineering: Several Meanings of "Top Down" and "Bottom Up." In Prawitz, D., Skyrms, B. & Westerstahl, D. (Eds.), *Proceedings of the 9th International Congress of Logic, Methodology and Philosophy of Science*. Dordecht: North Holland.

Franks, B. (1995). On Explanation in the Cognitive Sciences: Competence, Idealization, and the Failure of the Classical Cascade. *British Journal of Philosophy of Science 46*, 475-502.

Gilman, D. (1996). Optimization and Simplicity: Computational Vision and Biological Explanation. *Synthese*, *107*, 293-323.

Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.

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

Parker, GA, & Maynard Smith. (1990). Optimality theory in evolutionary biology. *Nature* 348, 27-33.

Peacocke, C. (1986). Explanation in Computational Psychology: Language, Perception and Level 1.5. *Mind and Language 1*, 101-123.

Peacocke, C. (1989). When is a Grammar Psychologically Real? In Alexander, G. (Ed.), *Reflections on Chomsky*. Oxford: Basil Blackwell.

Peacocke, C. (1993). Intuitive Mechanics, Psychological Reality and the Idea of a Material Object. In Eilan, N., McCarthy, R. & Brewer, B. (Eds.), *Spatial Representation*. Oxford: Basil Blackwell.

Steels, L. (1994). The Artificial Life Roots of Artificial Intelligence. *Artificial Life*, *1*, 75-110.

Ullman, S. (1996). *High-Level Vision*. Cambridge, MA: MIT Press.

Vilarroya, Ò. (2001). From Functional "Mess" to Bounded Functionality. *Minds and Machines* 11, *in press*.