

Symmetry between the intentionality of minds and machines? The biological plausibility of Dennett's position

Bence Nanay

Published online: 19 May 2006
© Springer Science+Business Media B.V. 2006

Abstract One of the most influential arguments against the claim that computers can think is that while our intentionality is intrinsic, that of computers is derived: it is parasitic on the intentionality of the programmer who designed the computer-program. Daniel Dennett chose a surprising strategy for arguing against this asymmetry: instead of denying that the intentionality of computers is derived, he endeavours to argue that human intentionality is derived too. I intend to examine that biological plausibility of Dennett's suggestion and show that Dennett's argument for the claim that human intentionality is derived because it was designed by natural selection is based on the misunderstanding of how natural selection works.

Keywords Adaptation–explanation · Computer · Dennett · Design · Intentionality · Natural selection · Trait type individuation

Dennett on the intentionality of computers

One of the most influential arguments against the claim that computers can think is that while our intentionality is intrinsic, that of computers is derived: it is parasitic on the intentionality of the programmer who designed the computer-program (Searle, 1980; 1983; 1992). If one aims to argue that computers can think, one needs to address this supposed asymmetry between the intentionality of minds and machines.

The standard way of explaining away this asymmetry is to argue that the intentionality of computers is not derived at all. This strategy seems more and more

B. Nanay (✉)
Department of Philosophy, University of California, Berkeley, CA 94720, USA
e-mail: nanay@berkeley.edu

promising, since some of the more successful artificial intelligence systems of the last two decades are capable of acquiring very impressive capacities without drawing on any built-in routines.¹

Daniel Dennett chose a different and rather surprising strategy for arguing against the asymmetry between human and computer intentionality. Instead of denying that the intentionality of computers is derived, he endeavors to argue that human intentionality is derived too. He writes:

As a late and specialized product, a triumph of Mother Nature's high tech, our intentionality is highly derived, and in just the same way that the intentionality of our robots (and even our maps and books) is derived. A shopping list in the head has no more intrinsic intentionality than a shopping list on a piece of paper. What the items mean (if anything) is fixed by the role they play in the larger scheme of purposes. We may call our own intentionality real, but we must recognize that it is derived from the intentionality of natural selection, which is just as real- but just less easily discerned because of the vast difference in time scale and size. (Dennett, 1987, p. 318, see also Dennett, 1990, p. 62).

The score is simple: computers are designed by programmers, we are designed by natural selection (see also Dennett, 1995, pp. 401–427). The intentionality of computers is not intrinsic, since computers are designed by humans. Similarly, human intentionality is not intrinsic either, since humans are designed by natural selection.

We now have an answer to the question of where we got our intentionality. We are artefacts, in effect, designed over the aeons as survival machines for genes that cannot act swiftly and informedly in their own interests.... So our intentionality is derived from the intentionality of our "selfish" genes. They are the Unmeant Meaners, and not us, and in so far as some theorist can interpret an event or structure in us as being about something or other (as, say, our inner-ear signals are about our orientation with regard to gravity and our acceleration), it is only because of the informative role that such signaling plays within the artifact, and the way it contributes to its self-preservation.... There is no need or room for a more absolute, "intrinsic", and "original" intentionality than this (Dennett, 1990, p. 59).

Thus, the symmetry between human and machine intentionality is regained: both are derived.

Dennett's solution has the additional advantage that it is reached on the basis of conceptual analysis: in order to make this argument, one does not need to know what actual computers can or cannot do. Thus, if Dennett's solution is sound, then we do not even need to go into the confusing details of computer science to figure out how computer intelligence works.

Dennett's suggestion has been criticized for various reasons (see Newton, 1992; Davies, 1995; Beisecker, 2002, for example). Also, several authors pointed out that Dennett's solution presupposes a form of adaptationism.² In this paper, I intend to examine the

¹ See for example McClelland & Rumelhart (1986), Rumelhart & McClelland (1986), Smolensky, Legendre, & Miyata (1992), Shastri & Ajjanagadde (1993). See also Dennett's criticism of some of these, Dennett (1993).

² On the question of adaptationism pro and contra see: Dawkins (1976), Lewontin (1978), Gould & Lewontin (1979), Gould & Vrba (1982), Dennett (1995), just to mention the most well-known titles in the very extended literature.

biological plausibility of Dennett's suggestion. I will point out that the claim that human intentionality is derived because it was designed by natural selection is based on the misunderstanding of how natural selection works.

What does 'derived intentionality' mean?

Dennett claims that human intentionality is derived. What does it mean? The intentionality of a computer is derived because it is explained by the way its programmer designed it. By analogy, human intentionality is derived because it is explained by the way it was designed, that is, as Dennett emphasizes, by natural selection. Thus, in order for human intentionality to be derived it must, at first approximation, be explained by something external to it: by what Dennett calls (in the quote above) "the intentionality of natural selection". If a system has derived intentionality, then its intentionality is (at least partially) explained by something external to it. Applying this to human intentionality, what we get is that if human intentionality is derived, then it is (at least partially) explained by natural selection. This is the claim I will examine in the next two sections.

Note that the condition that human intentionality should be (at least partially) explained by natural selection may be a necessary, but certainly not a sufficient condition for derived intentionality. If human intentionality is not partially explained by natural selection, then we cannot call it derived, but if it is partially explained by natural selection, it does not guarantee that it is derived. I will take up this question in more detail in the second half of the paper.

Now, of course, a computer programmer and natural selection differ in various respects. A computer programmer has a plan or design that exists before the computer program comes into existence. In the case of natural selection there is no such pre-existing plan or design. Also, it is questionable whether we can really talk about "the intentionality of natural selection". Most biologists would be scandalized even upon hearing this term.

These differences between natural selection and a human designer are undeniable, but Dennett does not need to worry about them. As long as human intentionality is explained by natural selection in the same way as the intentionality of computers is explained by the programmer's design, he can maintain his claim that human intentionality is as derived as the intentionality of computers.

Unfortunately, it is unlikely that he can do just this. One of the most fashionable topics in philosophy of biology in the last decade or so has been the question whether selection can explain adaptation. In other words, whether selection can explain why an organism has the traits it has. In order for Dennett's argument to work, this question needs to be answered in the positive. And there are important considerations that can be raised against this position.

First objection: what can natural selection explain?

Why do we have two eyes? Why is it the female mosquito that bites? Why do cats have sharp teeth? The answer to these questions is supposed to be provided by what is usually called adaptation-explanations. Adaptation-explanations aim to explain the supposed or real teleology of the world. As Robert Brandon says:

Adaptation–explanations [are] answers to what-for questions. Questions concerning putative adaptations, an anteater’s tongue, the structure of the human eye, or the waggle-dance of honeybees—are naturally formulated using what-for (one might also ask the same questions using why or how-come. The distinction is not a simple syntactic one). In contrast, we balk at using what-for in formulating other evolutionary questions, such as Why is hydrogen more abundant in the universe than uranium? (Brandon, 1985, 86–87; cf. Brandon, 1996, 30–45).

In other words, in adaptation–explanations the explanandum is why an organism has the trait it has. But what is the explanans? The obvious suggestion is that the explanans is a selection process. Adaptations should be explained with reference to selection processes that shaped the traits to be explained. Thus, we can explain why certain creatures have the traits they have by referring to why these traits were selected in the course of evolution. This is exactly the step that is needed to complete Dennett’s argument. Unfortunately, it has been argued repeatedly that contrary to our hopes to use selection to answer the questions like the ones quoted above, selection cannot play a role in explaining adaptation.

The aim of this section is to introduce the recent debate about whether such explanations are possible. This debate has been quite severe in the last decade or so. The view that selection can play a role in explaining adaptation has been defended mainly by Karen Neander (1995a, b, see also Millikan, 1990; Nanay, 2002; 2005). On the other side of the trench the central figure is Elliott Sober (1984; 1995, see also Cummins, 1975; Dretske, 1988; 1990 and Walsh, 1998). After Sober (1995) and Walsh’s (1998) arguments, the position of Neander’s side appears rather shaky.³ In order for Dennett’s suggestion to work, he would need to provide munitions to this camp.

Before turning to the actual argument, it is important to clarify the framework of the debate. First, it is important to draw a distinction between the scope of this debate and that of adaptationism. The adaptationist’s claim is that if an organism has a trait, then there is (or at least tends to be) a selection process that explains this organism’s having this trait. In other words, if organism *x* has trait *A*, then there is (or at least tends to be) a selection process that explains why *x* has trait *A*.

Contrast this with the central claim of the debate about the explanatory power of selection in explaining adaptations: if *x* has trait *A* and if *x* is in a population where trait *A* has been selected over trait *B*, then this selection process is explanatorily relevant to why *x* has trait *A*. Both are concerned with the explanation of adaptation with the help of selection. But they are very different indeed: the second claim does not assume that all or even most traits have adaptation–explanations.

Dennett has been repeatedly criticized for his adaptationism. He is aware of this consequence of his theory and provided powerful arguments defending his position. Without assessing these arguments or judging whether Dennett’s adaptationism is tenable, it needs to be pointed out that the question whether selection explains adaptation is orthogonal to the question of adaptationism. Even if Dennett can give a convincing answer to the former question, the latter may still raise problems for him.

Hence, I will assume in what follows that the question under scrutiny is the following: if in a population a trait has been selected over other traits, can this explain (at least partially) why organisms in that population have this trait? In other words, I take it for granted that we know the selection process, and we want to tell why certain organisms have the traits they have.

³ Matthen (1999) criticized Sober’s position, arguing that Sober’s argument does not work if the selection process is sexual selection, but as Lewins (2001) pointed out, Matthen’s argument is not conclusive.

I would like to narrow down the question even further and focus on whether *cumulative* selection can play a role in explaining adaptation or not.⁴ There are non-cumulative selection processes that cannot play any role in explaining adaptation: the ones whereby the replicators do not change from generation to generation. The most successful replicator may spread and make all the others extinct, but by doing so it will not change. An example could be the clay crystal that grows faster than the other crystals in the same pool (cf. Bedau, 1991, 650–654; Walsh, 2000, 142–143). After a certain time, the fastest growing crystal will be the only one in the pool, but its structure will not change in the selection process.⁵ This is an example of a non-cumulative selection process that does not play a role in explaining any adaptation.

Leaving the non-cumulative case aside, I will focus on the question whether cumulative selection can play a role in explaining adaptation or not. Also, it needs to be emphasized that the question is whether cumulative selection can *play a role in explaining* adaptation and not whether it can fully explain adaptation.

Two strong arguments have been given in favour of the claim that no selection process can play a role in explaining adaptations. The first one is that selection is a negative force, it eliminates, but it does not create, hence it cannot play a role in explaining adaptation. According to the second, selection cannot play a role in explaining adaptation, since selection is a population-level phenomenon, whereas adaptation occurs on the individual level. These arguments are often provided together, but I take them to be logically independent.

I will not say much about the second argument here, because even if the argument is conclusive, it may be possible that the presence of a trait in a population is explained by the function of this trait. In other words, selection cannot explain why *I* have two eyes, why *this* cat has sharp teeth, etc., but it can explain why these traits are present in the population. Thus, even if Sober is right and selection cannot explain why individual organisms have certain traits (Sober, 1984, p. 150; Sober, 1995, p. 384), it can still explain the frequencies of traits in populations. Note that Dennett does not need to talk about the individual level at all in order to make the claim that natural selection plays a role in explaining our intentionality. Sober's second argument about the levels of explanation may show that natural selection cannot explain *my* or *your* intentionality, but it does not say anything about whether it can explain the presence of intentionality in the human population. And the latter of these claims is sufficient to complete Dennett's argument.

The first argument about selection as a negative force, however, does challenge the assumption that natural selection explains what human minds in general are supposed to do.

Selection as a negative force

Sober claims that selection is a negative force: it does not create, it only destroys (Sober, 1984, chapter 5). The upshot is that random mutations create a variety of individuals (or genetic plans) and selection eliminates some of these, but the explanation of the traits of

⁴ On the question of adaptationism pro and contra see: Dawkins (1976), Lewontin (1978), Gould & Lewontin (1979), Gould & Vrba (1982) Dennett (1995), just to mention the most well-known titles in the very extended literature.

⁵ Cumulative selection is a selection process whereby the changes of the replicators accumulate: they are transmitted to the next generation. Natural selection, for example, is cumulative.

one of these individuals is provided by random mutation and inheritance (and, of course, some developmental factors), not by the elimination process. Selection can explain why certain individuals were eliminated, but it cannot say anything about the traits of the ones that were *not* eliminated.

Karen Neander analyses this argument, which she calls the argument for the Negative View of selection, in great detail (Neander, 1995a). She argues that selection does play a role in explaining why an individual has the traits it has, but only a certain kind of selection: cumulative selection. But she admits that Sober's criticism is valid for non-cumulative selection processes.

First, it is important to make some comments on the terminology Neander and Sober use. More precisely, it is crucial to examine whether the opponents and the advocates of this argument mean the same thing when they talk about selection. Sober analyses mutation as something distinct from selection: he claims that the former notion plays an important role in explaining adaptation, whereas the latter does not. The question is how consistent this is with the generally accepted notion of selection.

According to David Hull, selection consists of "repeated cycles of replication and environmental interaction so structured that environmental interaction causes replication to be differential".⁶ He analyses selection, conceived of traditionally as "heritable variation in fitness", as cycles of a copying process (replication) and the interaction with the environment.

If we accept this concept of selection, it is hard to see how Sober could maintain that it is not selection, but mutation that plays a role in explaining adaptation, since in Hull's picture, mutation (replication with variation) is one of the two steps of the selection process. If selection consists of repeated cycles of replication and interaction, then replication is obviously part of the selection process. And this replication process must be differential; hence, replication with variation, i.e. mutation, is part of the selection process.⁷ According to Hull, selection is replication with variation *followed by* interaction. In the light of this, it is a surprising claim that selection cannot play a causal role in explaining adaptations, while mutation can, if mutation is part of the selection process.

This, however, would be a too easy way to oppose Sober's argument. He obviously means something else by selection. It is reasonable to say that what he means is what Hull means by interaction.⁸ Interaction is indeed a negative force: all it does is to eliminate some of the interactors.

And here we run into another confusion. Hull's notion of selection is a notion of cumulative selection. If selection is "the repeated cycles of replication and environmental interaction so structured that environmental *interaction causes replication [that is, replication in the next generation] to be differential*" (my emphasis), then the changes of the replicators must be transmitted to the next generation. If Sober takes Hull's notion of interaction to be selection, then it is difficult to see how he could allow for cumulative selection.⁹ Neander, on the other hand, explicitly talks about cumulative selection, when she talks about selection; she even admits that non-cumulative selection cannot play a role in explaining adaptation, but cumulative selection can. Could it be that

⁶ The clay crystals, described in Bedau, 1991 behave this way, and not the way Bedau himself explained.

⁷ Hull, Langman, & Glenn, 2001: 53; cf. Hull, 1981: 40–41.

⁸ This is also true under different widespread definitions of selection, such as Darden and Cain's. Note that Darden and Cain's notion of selection is not necessarily a cumulative one, whereas Hull's is. See also Nanay, 2001.

⁹ The same is true of Vrba's concept of selection: "Selection is the interaction between heritable, emergent character variation and the environment [...]" Vrba (1984: 319).

Neander uses a notion of selection that is similar to Hull's, whereas Sober uses a notion of selection that is more similar to Hull's notion of interaction? Is it possible that the whole debate is terminological?

I think not. It seems that Neander accepts the way Sober refers to selection when she argues against his position. In Hull's terminology, both are concerned with the question whether interaction is causally relevant to the explanation of adaptation or not. This, of course, leaves open the possibility to argue that *selection* in Hull's original sense does play a causal role in the explanation of adaptation, but this is a possibility I cannot pursue here. Instead, I would like to examine further the Sober–Neander debate that we managed to localize as the question whether interaction is causally relevant to the explanation of adaptation or not.

Sober's argument is that, in Hull's terms, it is replication that explains why an individual has a certain trait. Environmental interaction (of the previous generation) does not play any role in such explanation. The gist of his argument is the following. Let us take an organism that has two offspring, one of which has a certain trait A, whereas the other does not. Since trait A is advantageous to these organisms in the given environment, the second offspring dies, whereas the first will have offspring, one of which, call her individual x, also has trait A. The question is of course, what explains that x has trait A. Sober's answer is that it is the mutation as a result of which A appeared in x's mother and inheritance, as a result of which A was transmitted to x. The explanation is simple: A appeared as a result of a random mutation in x's mother and then x inherited it from her mother. What explains the presence of the trait is, hence, mutation and inheritance. Selection is irrelevant, since the fact that x's uncle died or not does not have any causal influence on whether x has trait A or not (Sober, 1984. See also Cummins, 1975, 750–751; Sober, 1995, 393).

Some think that this argument is flawed (Neander, 1995a, p. 76; Mathen, 1999; Nanay, 2005).¹⁰ Without going into the details of the arguments pro and con, it is important to point out that Dennett's strategy about human and computer intentionality works only if he assumes what appears to be a very controversial claim in philosophy of biology.

Second objection: the structure of Dennett's explanatory scheme

Suppose, however, that Sober is wrong and the argument outlined above is not conclusive. Suppose that at least one of the three counterarguments that have been given against it (Neander, 1995a; Matthen, 1999; Nanay, 2005) works. Would this mean that Dennett's strategy is unproblematic from a biological point of view? Alas, it would not. There is another argument against Dennett's assumption that is logically independent from the worry I raised above.

The worry I raised in the last sections was that it is questionable whether natural selection can play any role in explaining why a certain organism has a certain trait. Now suppose that this claim is true: natural selection can indeed play some role in explaining why a certain organism has a certain trait. Would this be enough for Dennett's argument to work? I will argue that it would not.

Remember that the condition that human intentionality should be at least partially explained by natural selection was given as a probably necessary, but certainly not sufficient condition for derived intentionality. If human intentionality is not partially

¹⁰ In fact, the notion of cumulative selection he endorses in footnote 10 of Sober (1995) definitely would not qualify as cumulative selection under Hull's or Neander's definition.

explained by natural selection, then we cannot call it derived, but if it is partially explained by natural selection, the question is still open whether it is derived or not.

Just because a capacity or a trait is partially explained by something external to it, it does not follow that this capacity or trait is derived. Virtually anything can play a role in explaining an *explanandum* (see Van Fraassen, 1980). The fact that my shirt is white is partly explained by my successful attempt to avoid dripping coffee on it this morning. Still, the color of my shirt is not somehow parasitic on this event, nor would we be justified to call the color of this shirt “derived”. When the intentionality of computers is described as derived, something stronger is implied.

If the intentionality of computers is derived intentionality, then it is the programmer who is responsible for the intentionality of the computer. In other words, the programmer’s work is *sufficient* for explaining the intentionality of the computer.

The structure of the claim is the following. If the computer was designed to do F, then this fact is sufficient for explaining that the computer (in suitable circumstances) does F. Similarly, if Dennett wants to maintain the symmetry between human and machine intentionality, then he needs to show that the fact that the human mind was selected to do F by natural selection is sufficient for explaining that the human mind (in suitable circumstances) does F.

But what does the term ‘in suitable circumstances’ mean? It would be obviously wrong to say that x’s having been designed to do F is sufficient for explaining x’s doing F. If a corkscrew is designed to open wine bottles, this does not fully explain that it does in fact open wine bottles. Lots of other factors will figure in this explanation, for example that I know how to use a corkscrew. Hence, the criterion that x’s having been designed to do F is sufficient for explaining x’s doing F is a too strong. It sounds plausible to say, however, that if x was designed to do F, then x is supposed to do F. More precisely, x’s having been designed to do F is sufficient for explaining x’s having the *function* to do F. The fact that the corkscrew was designed to open wine bottles is sufficient for explaining that it has the function of opening wine bottles. Similarly, the fact that a computer was designed by a programmer to do a certain calculation is sufficient for explaining that it has the function of doing this calculation. It is important that the fact that a computer was designed to do F is not sufficient for explaining what the computer actually does—various other factors play important role in such an exhaustive explanation (for example that the computer’s electricity supply works properly). However, the fact that a computer was designed to do F is indeed sufficient for explaining that the *function* of the computer is to do F. Similarly, the fact that our mind was designed by natural selection to represent veridically is sufficient for explaining that it has the *function* to represent veridically. Thus, the claim we now need to examine is whether x’s having been designed to do F is sufficient for explaining x’s having the function to do F.¹¹

If we say that the function of a corkscrew is to open wine bottles because it is designed to open wine bottles, then the *explanans* and the *explanandum* are claims about the same token object. A token object, a corkscrew was designed to open wine bottles and, as a result, the same (token-identical) corkscrew has the function to open wine bottles. Thus, the structure of the claim is that x was designed to do F, therefore, x has the function to do F.

¹¹ Interestingly enough the three authors who criticize Sober’s argument do so for very different reasons. Neander claims that cumulative selection can indeed explain adaptation, Mathen argues that sexual selection can, whereas Nanay insists that cumulative selection can explain adaptation if the environmental resources are limited.

Of course this model applies only to artefacts that are not mass-produced, but most of the artefacts that surround us (including corkscrews) are in fact mass-produced. The structure of the claim in the mass-production case is that tokens of type X were designed to do F, therefore, x (a token of type X) has the function to do F. The *explanandum* and the *explanans* are claims about tokens of the same type. Both of these explanatory schemes seem unproblematic.

How about the design by natural selection? Dennett's answer is that design by natural selection works in an analogous way as designed by a designer. The function of my heart is to pump blood because the heart was selected for pumping blood. What is the logical structure of this claim? Tokens of type X (my ancestors' hearts) were selected for doing F, therefore, x^* (my token heart) has the function to do F. This explanatory scheme would be analogous to the artefact case (and it would be unproblematic) only if it is argued that x^* is a token of type X. In other words, only if it is argued that a trait of an organism is of the same type as the traits that were selected for in some earlier generation. Only in this case would the *explanandum* and the *explanans* be claims about tokens of the same type. But this assumption is far from being obvious.

So, how can we individuate trait types? What makes hearts different from non-hearts? I will argue that there is no coherent non-circular way of individuating trait types that Dennett could use to complete his argument.

The individuation of trait types

The crucial question that trait type individuation raises for Dennett's argument is: how are trait types individuated?; e.g. what makes hearts different from non-hearts? I will consider three possible options in answer to this question and point out that none of them can be used in Dennett's argument for the derived intentionality of human minds.

Functional criteria

The most widely accepted account of trait type individuation holds that tokens of a certain trait type all have the same function. A token object belongs to trait type T if and only if it has certain functional properties: if it has the function to do F. Those entities are hearts that have the function of pumping blood. Those entities that do not have this function are not hearts.

As Karen Neander puts it: "Most biological categories are only definable in functional terms" (Neander, 1991a, 180; see also Beckner, 1959, 112; Lewens, 2004, 99). This account of individuating trait types is widely (though not universally) accepted as a general suggestion both in philosophy of biology and in philosophy in general. Tyler Burge, for example, writes that "to be a heart, an entity has to have the normal evolved function of pumping blood in a body's circulatory system" (Burge, 1989, 312).

Everything looks promising. It is important to note, however, that Dennett cannot help himself to this way of individuating trait types when making the claim that natural selection explains the function of our minds without running into circularity.

As we have seen, this claim presupposes an account of trait type individuation. Now, if we want to avoid circularity, we certainly cannot use the notion of function in order to explain trait type individuation. When we are explaining function, the claim that x^* is a token of type X is part of the *explanans*, so we cannot then use the *explanandum* (function)

to explain the *explanans* (why x^* is a token of type X).¹² Thus, if we want to talk about trait types in the explanation of function, we need some other way of individuating them.

Morphological criteria

A simpler suggestion is that we can use morphological criteria for individuating trait types. The proposal is that a token object belongs to trait type T if and only if it has certain morphological properties. An entity is a heart if, for example, it has a certain shape, size and color, and it is not a heart otherwise.

This suggestion was attacked by Karen Neander. She writes:

“[H]eart” cannot be defined except by reference to the function of hearts because no description purely in terms of morphological criteria could demarcate hearts from non-hearts. Biologists need a category that ranges over different species, and hearts are morphologically diverse: fish have a single pump with only one auricle, but amphibians and most other reptiles have the single heart with two auricles, and while many reptiles have the ventricle partly partitioned, only crocodiles, birds and mammals have the two separate ventricles (Neander, 1991a: 180).

Thus, trait types cannot be individuated morphologically, because they need to range over different species.

But even if we only want to individuate a trait type within a certain species, we still cannot use morphological criteria. A malformed heart that does not have the morphological properties hearts have is a heart all the same. What keeps hearts and non-hearts apart cannot be a set of morphological properties.

A possible suggestion would be to say that hearts are those entities that play a certain causal role and those entities that do not play this causal role are not hearts. This proposal may be seen as a version of the morphological account of trait type individuation, but it could also be interpreted as a functional account (if we conceive of function in a way Cummins (1975; 2002) does). This proposal, however, would not work either in the malformed heart case: a malformed heart does not have the causal role hearts have, but it is a heart nonetheless. What matters is not what the heart does (or how it looks), but what it is supposed to do. In other words, the suggestion that morphological criteria could be found for individuating trait types does not work.¹³ This leaves us with the third alternative.

Homological criteria

The third possible answer is the following. One could argue that what guarantees that two traits are tokens of the same type is that they are homologues: they are members of the same “reproductively established family”, or, to put it very simply, they are both coded by

¹² I leave aside the entire function literature as well as the question of the similarities and differences between biological and artefact function (see Wright, 1973; Cummins, 1975; Boorse, 1976; Wright, 1976; Millikan, 1984; Neander, 1991a, b; 1993; Griffiths, 1993. See also Mills & Beatty, 1979; Bigelow & Pargetter, 1987; Godfrey-Smith, 1994; Walsh, 1996).

¹³ See Neander, 2002, 403; see also Griffiths, 1993; Davies, 2000; 2001.

the same gene.^{14,15} My trait in the present generation was coded by my gene sequence G. The traits that were selected for in some earlier generations were coded by some other genes and my gene sequence G is a copy of these genes. It does not matter that my trait and the ancestors' traits do not look alike. The fact that the gene that coded my trait is a replicate of the gene that coded my ancestors' is enough to make these traits tokens of the same type (see for example, Amundson & Lauder, 1994; Lewens, 2004, pp. 99–100). Thus, the suggestion is that two token objects belong to the same trait type if and only if they are coded by genes that are replicates (copies) of one another.

A possible worry about this suggestion is that a trait is not the expression of a gene. The way a trait turns out depends partly on the gene, but it also depends on the intra- and extra-cellular environment during the developmental process. Assuming that the genetic factor can be singled out from this complex causal network is biologically very implausible.

One of the more influential theories of the last decade in philosophy of biology is the developmental systems theory (Griffiths and Gray, 1994: esp. 298ff; Oyama, Griffiths, & Gray, 2001; see also Sterelny, Smith, & Dickison, 1996; Godfrey-Smith, 2000; Nanay, 2002). This theory emphasizes that the genetic and the developmental parts of evolutionary explanations cannot be separated from one another. More precisely, they argue that it is not possible to make a sharp distinction between “genetic” and “environmental” developmental causes (see esp. Griffiths and Gray, 1994: 304). It is easy to see that the notion of reproductively established family, and the homological account of trait type individuation, presupposes a sharp distinction between genetic and environmental causes.

Thus, if one takes the insights of developmental systems theory seriously, then one needs to be suspicious about the concept of a reproductively established family; one needs to reject the view that a trait is of a certain type in virtue of the fact that it was coded by a certain gene. Therefore, one also needs to reject the view that two traits are of the same type in virtue of the fact that they are coded by genes that are replicates.

But this conclusion is too quick. Perhaps the notion of homology can be reconciled with the insights of developmental systems theory. V. Louise Roth, for example, writes that “A necessary component of homology is *the sharing of a common developmental pathway*” (Roth, 1984, 17, original emphasis). Günther Wagner also developed an account of homology where developmental processes play a crucial role in establishing homology (Wagner, 1994). Thus, the suggestion is that whether a trait belongs to a homologically established trait type depends not only on what gene this trait was coded by, but also by the developmental process. If this is plausible, then we do not need to worry about the possible

¹⁴ The morphological account of trait type individuation may be supplemented with a homological one (Amundson & Lauder, 1994; Lauder, 1994). As we will see, however, this homological account raises serious worries.

¹⁵ See for example Millikan, 1984. The term “reproductively established families” was introduced by her (ibid, p. 23ff). To be fair, Millikan’s notion is not committed to the claim that members of a reproductively established family are coded by replicate *genes*. Her notion of reproductively established family is a very general one: it does not presuppose that the replicates that code our traits must be genes. Any replicator can play this role in the definition of a reproductively established family. This distinction is important in the light of recent research about epigenetic inheritance systems. It has been suggested that mechanisms other than the copying of the DNA molecule can also play a role in transmitting information from one generation to the other (Jablonka & Lamb, 1995: 79–108, esp. 79–85). Thus, the DNA is only one component of what Jablonka and Lamb calls “the multiple inheritance system” (Jablonka & Lamb, 1995: 226). Thus, the replicator that codes our trait is unlikely to be identical to the gene. The notion of reproductively established family can accommodate these findings. Replacing the concept of gene with that of replicator helps here, but it cannot help when addressing the problems I will raise about the notion of reproductively established family and homology in general. For simplicity, I will use the concept of gene (and not that of replicator) in characterizing the homological account.

conflict with developmental systems theory. The homological way of individuating trait types does seem to be a promising candidate.

However, there is a stronger reason for giving up the homological account of trait type individuation. Take the following example. The forelimbs of vertebrates, such as the wing of a bird, and the foreleg of an ancient amphibian, are homologous: the wing of an eagle is coded by genes that are replicates of the genes that coded the foreleg of some ancient amphibian. Thus, they are part of the same reproductively established family. According to the suggestion for individuating trait types we are considering here, they must be of the same trait type. But the wing of the eagle and the foreleg of an ancient amphibian are clearly not tokens of the same trait type. One of them is a wing and the other is a foreleg. They belong to very different trait types indeed.

Perhaps one could argue that such old members of a reproductively established family (for example, the traits of our ancient amphibian) just do not count. A possible suggestion would be to say that what makes a trait token belong to a trait type is that it is coded by genes that are copies of the genes that coded other *recent* members of the reproductively established family. In other words, two token objects belong to the same trait type if and only if they are coded by genes that are *recent* replicates (copies) of one another.

The problem with this suggestion is that there is no non-circular way of cashing out what is meant by ‘recent’ here. We would be happy to say that the eyes of the eagle and the eyes of the ancient amphibian are tokens of the same type. Then why cannot we do the same with forelimbs? What is so different in the two cases that makes us sort the two token traits under the same type in the latter case but not in the former one?

The only thing that differentiates the eye-case from the forelimb-case is that the selection pressure changed in the latter case, but it did not change in the former. Forelimbs have been selected for doing something *different* in the bird population and in the ancient amphibian population. Eyes, on the other hand, have been selected for doing *the same thing* in the bird population and in the ancient amphibian population.

In other words, if we want to make sense of the suggestion that what makes a trait token a member of a trait type is that it is coded by genes that are copies of the genes that coded other *recent* members of the reproductively established family, then we will have difficulties defining what is meant by the term ‘recent’. In defining the eagle’s trait types, ‘recent’ includes the ancient amphibian population when we are analyzing the eye example, but it certainly does not include the ancient amphibian population when it comes to forelimbs. The bottom line is that what ‘recent’ amounts to depends on what the trait in question has been selected for; what ‘recent’ amounts to depends on the (etiological) function of the trait. The suggestion boils down to the claim that what makes a trait token belong to a trait type is that it is coded by genes that are copies of genes that coded other trait tokens that were selected for doing the same thing as this token (or, in other words, that had the same function as this token).

Thus, this way of individuating trait types collapses into the functional account of trait type individuation. But we have seen above that the functional account of trait type individuation cannot be used when explaining that design fixes function without running into circularity.

But if this is true, then Dennett cannot rely on any of the three ways of individuating trait types. Since the claim that natural selection explains the function of our minds requires an unproblematic way of individuating trait types, he is not entitled to make this claim.

To sum up, in the design by natural selection case, the function of a trait does not seem to follow as nicely from what it was designed to do as in the artefact case. In other words,

whereas the fact that a computer was designed by its programmer is sufficient for explaining the intentionality of this computer, the fact that our mind was designed by natural selection is not sufficient for explaining our intentionality. In other words, the intentionality of the computer is derived in a sense in which our intentionality is not. Dennett's argument for the symmetry between human intentionality and the intentionality of computers does not work.

Conclusion

It is important to emphasize that the argument I presented in this paper does not show that computers cannot think or even that their intentionality is different from ours. All I aimed to show was that Dennett's strategy to argue that human intentionality is as derived as the intentionality of computers does not work. Thus, if we want to argue that the intentionality of computers is similar to that of humans, we cannot spare going through the complicated details of the current findings and models of contemporary computer science.

References

- Amundson, R., & Lauder, G. V. (1994). Function without purpose: The uses of causal role function in evolutionary biology. *Biology and Philosophy*, 9, 443–469.
- Beckner, M. (1959). *The biological way of thought*. New York: Columbia University Press.
- Bedau, M. (1991). Can biological teleology be naturalized? *Journal of Philosophy*, 88, 647–657.
- Beisecker, D. (2002). Dennett and the quest for real meaning: In defense of a 'Myth'. *Philosophy in the Contemporary World*, 9(1), 11–18 (Spring-Summer 2002).
- Bigelow, J., & Pargetter, R. (1987). Functions. *Journal of Philosophy*, 84, 181–197.
- Boorse, C. (1976). Wright on functions. *Philosophical Review*, 85, 70–86.
- Brandon, R. N. (1985). Adaptation explanations: Are adaptations for the good of replicators or interactors? In D. J. Depew, & B. H. Weber (Eds.), *Evolution at a crossroads: The new biology and the new philosophy of science* (pp. 81–96). Cambridge, MA: The MIT Press.
- Brandon, R. N. (1996). *Concepts and methods in evolutionary biology*. Cambridge: Cambridge University Press.
- Burge, T. (1989). Individuation and causation in psychology. *Pacific Philosophical Quarterly*, 70, 303–322.
- Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, 72, 741–765.
- Cummins, R. (2002). Neo-Teleology. In A. Ariew, R. Cummins, & M. Perlman, (Eds.), *Functions* (pp. 157–173). Oxford: Oxford University Press.
- Davies, D. (1995). Dennett's stance on intentional realism. *Southern Journal of Philosophy*, 33, 299–312.
- Davies, P. S. (2000). Malfunctions. *Biology and philosophy*, 15, 19–28.
- Davies, P. S. (2001). *Norms of nature: Naturalism and the nature of functions*. Cambridge, MA: The MIT Press.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press, A Bradford Book.
- Dennett, D. C. (1990). The myth of original intentionality. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale, & K. V. Wilkes (Eds.), *Modelling the mind* (pp. 43–62). Oxford: Clarendon Press.
- Dennett, D. C. (1993). Labeling and learning. Commentary on Clark and Karmiloff-Smith. *Mind and Language*, 8, 540–548.
- Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Touchstone.
- Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: The MIT Press.
- Dretske, F. (1990). Reply to reviewers. *Philosophy and Phenomenological Research*, 50, 819–839.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, 28, 344–362.
- Godfrey-Smith, P. (2000). The replicator in retrospect. *Biology and Philosophy*, 15, 403–423.
- Gould, S. J., & Lewontin R. (1979). The spondrels of San Marco and the Panglossian paradigm. *Proceedings of the Royal Society*, B205, 581–598.

- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8, 4–15.
- Griffiths, P. E. (1993). Functional analysis and proper functions. *British Journal for the Philosophy of Science*, 44, 409–422.
- Griffiths, P., & Gray, R. (1994). Developmental systems and evolutionary explanation. *Journal of Philosophy*, 91, 277–304.
- Hull, D. L., Langman, R. E., & Glenn S. S. (2001). A general account of selection: Biology, immunology and behavior. *Behavioral and Brain Sciences*, 24, 511–528. Reprinted in Hull, David L. (2001).
- Hull, D. L. (1981). Units of evolution: A metaphysical essay. In U. J. Jensen, & R. Harré (Eds.), *The philosophy of evolution* (pp. 23–44). Brighton: Harvester Press.
- Jablonska, E., & Lamb, M. J. (1995). *Epigenetic inheritance and evolution*. Oxford: Oxford University Press.
- Lauder, G. V. (1994). Homology, form, and function. In B. K. Hall (Ed.), *Homology: The hierarchical basis of comparative biology* (pp. 151–196). London: Academic Press.
- Lewins, T. (2001). Sex and selection: Reply to Matthen. *British Journal for the Philosophy of Science*, 52, 589–598.
- Lewins, T. (2004). *Organisms and artifacts*. Cambridge, MA: The MIT Press.
- Lewontin R. (1978). Adaptation. *Scientific American*, 273, 213–228.
- Matthen, M. (1999). Evolution, Wisconsin style: Selection and the explanation of individual traits. *British Journal for the Philosophy of Science*, 50, 143–150.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing* (Vol. 2). Cambridge, MA: MIT Press.
- Millikan, R. G. (1984). *Language, Thought and other biological categories*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1990). Seismograph readings for, explaining behavior. *Philosophy and Phenomenological Research*, 50, 819–839.
- Millikan, R. G. (1993). *White queen psychology and other tales for Alice*. Cambridge, MA: MIT Press.
- Mills S., & Beatty J. (1979). The propensity interpretation of fitness. *Philosophy of Science*, 46, 263–286.
- Nanay, B. (2001). A more pluralist typology of selection processes. *Behavioral and Brain Sciences*, 24, 547–548.
- Nanay, B. (2002). The return of the replicator: What is philosophically significant in a general account of replication and selection? *Biology and Philosophy*, 17, 109–121.
- Nanay, B. (2005). Can cumulative selection explain adaptation? *Philosophy of Science*, 72, in press.
- Neander, K. (1991a). Functions as selected effects. *Philosophy of Science*, 58, 168–184.
- Neander, K. (1991b). The teleological notion of ‘function’. *Australasian Journal of Philosophy*, 69, 454–468.
- Neander, K. (1995a). Pruning the tree of life. *British Journal for the Philosophy of Science*, 46, 59–80.
- Neander, K. (1995b). Explaining complex adaptations. A reply to Sober’s ‘Reply to Neander’. *British Journal for the Philosophy of Science*, 46, 583–587.
- Neander, K. (2002). Types of traits. Function, structure, and homology in the classification of traits. In A. Ariew, R. Cummins, & M. Perlman (Eds.), *Functions* (pp. 402–422). Oxford: Oxford University Press.
- Newton N. (1992). Dennett on intrinsic intentionality. *Analysis*, 52, 18–23.
- Oyama S., Griffiths P. E., & Gray R. (Eds.). (2001). *Cycles of contingency: Developmental systems and evolution*. Cambridge, MA: MIT Press.
- Roth V. L. (1984). On homology. *Biological Journal of the Linnean Society*, 22:13–39.
- Rumelhart D., & McClelland J. L. (Eds.). (1986). *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. I). Cambridge, MA: MIT Press.
- Searle, J. (1980). Intrinsic intentionality: Reply to criticisms of minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 450–456.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: The MIT Press.
- Shastri L., & Ajjanagadde V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16, 417–494.
- Smolensky, P., Legendre G., & Miyata Y. (1992). *Principles for an integrated Connectionist/symbolic theory of higher cognition*. Report 92–08. Institute of Computer Science, University of Colorado.
- Sober, E. (1984). *The nature of selection*. Cambridge, MA: The MIT Press.
- Sober, E. (1995). Natural selection and distributive explanation. *British Journal for the Philosophy of Science*, 46, 384–397.
- Sterelny K., Smith K., & Dickison M. (1996). The extended replicator. *Biology and Philosophy*, 11, 377–403.
- Van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.

- Vrba, E. (1984). What is species selection? *Systematic Zoology*, 33, 263–292.
- Wagner, G. P. (1994). Homology and the mechanisms of development. In B. K. Hall (Ed.), *Homology: The hierarchical basis of comparative biology* (pp. 273–299). London: Academic Press.
- Walsh, D. M. (1996). Fitness and function. *British Journal for the Philosophy of Science*, 47, 553–574.
- Walsh, D. M. (1998). The scope of selection: Sober and Neander on what natural selection explains. *Australasian Journal of Philosophy*, 76, 250–264.
- Walsh, D. M. (2000). Chasing shadows: Natural selection and adaptation. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 31, 135–153.
- Wright, L. (1973). Functions. *Philosophical Review*, 82, 139–168.
- Wright, L. (1976). *Teleological explanations*, Berkeley: University of California Press.