

---

## Adaptationism, Deflationism and Anti-Individualism

Tomáš Hříbek

### 1 Introduction

A number of naturalistic philosophers have been trying to integrate theories of mental representation within the domain of evolution and biological function. In particular, these philosophers suggest that adaptationism, which is central to the Darwinian science, can provide a ground for anti-individualist, or externalist, accounts of representational states that have become dominant in philosophy of mind. Adaptationism is usually presented as an *empirical* claim about the causes of phenotypic traits. For starters, we can make do with Elliott Sober's definition:

Natural selection has been the only important cause of most of the phenotypic traits found in most species. (Sober 1996, 72)

Sober is clear in treating adaptationism as an empirical thesis, albeit the one whose truth-value could be determined "only in the long run" (*ibid.*). However, some philosophers as well as scientists have meant something a lot stronger by "adaptationism." But we shall come back to that later. At any rate, it seems beyond doubt that psychological anti-individualism is definitely not a mere empirical claim. It is meant to be a *metaphysical* thesis about the *nature* of representational states. Most of traditional philosophy of mind, in virtue of its assumption that representational mental states can be fully characterized by attending solely to the properties and states internal to the individual's bearer of these states, has been individualistic. By contrast, anti-individualism is a relatively recent theory originated by Tyler Burge. According to his recent definition, anti-individualism is the claim that

the natures of many mental states constitutively depend on relations between a subject matter beyond the individual and the individual that has the mental states, where relevant relations help determine specific natures of those states. (Burge 2010, 61)

Burge means that many mental states, specifically those that are representational, would not be what they are, unless the relevant relations to the external environment were in place. Therefore, the idea of a constitutive dependence of representational states of relations between the individual and her environment is distinct from the idea that mental states causally depend on such external relations. The latter idea is quite acceptable to many individualists. However, while agreeing that thoughts or perceptions are prompted by events in the external environment, individualists go on to claim that those mental states are then fully identifiable in terms of factors internal to the individual. Constitutive dependence is also stronger than metaphysical dependence. Necessities such as that it's true of each mental state that it is not a number or made of cheese are weaker than constitutive dependence because they do not figure in explanation of the nature of mental states.

These are the claims about the character of the relations between mental representations and their environment. The further issue to consider is whether the mind reduces to its external relations; in other words, whether it belongs in the ontological category of relation. Burge resists this conclusion with an example from biology: "It is constitutively necessary that to be a heart, an organ must have the function of pumping blood through a circulatory system" (ibid., 66). However,

the relations to these other entities are not part of the internal structure of the heart. Nor is the heart itself a relation. Thus the nature of the heart is constitutively dependent for being what it is on relations to things beyond it. But the heart itself has a structure that is not made up of those relations. I think that representational mind is like that. (Ibid.)

Now, the idea behind the naturalization project under discussion here is to present anti-individualism as an implication of adaptationism, whether the latter is thoroughly empirical in nature or something else. Thus, Daniel Dennett claims that

Burge's anti-individualistic thesis is then simply a special case of a very familiar observation: functional characterizations are relative not only to the embedding environment, but also to assumptions about optimality of design. (Dennett 1987, 310)

In other words, Dennett suggests that anti-individualism is implied by the fact that the ascription of representational states, which he construes as functional states in the biological sense, presupposes a description of the *selectionist history* of these states. It is precisely this appeal to history that is supposed to supply the required external component—*i.e.*, something outside the individual—of the identity conditions of representational states. Likewise, in her elaborate theory of psychological explanation as a species of biofunctional

explanation, Ruth Millikan argues that the functions in question are to be characterized historically. She says:

I would like to explore implications for the science of psychology of the thesis that the categories of intentional psychology are function categories in the biologist's sense of "function," taking this to be a sense in which function is determined by evolutionary history rather than by current dispositions. (Millikan 1993, 171)

Thus, Dennett and Millikan share a conviction that anti-individualism can be vindicated as a feature of a psychology understood as an offshoot of the adaptationist program in biology. In short, their defence of anti-individualism assumes that psychological explanation is *of a kind* with biological explanation. In addition to the reductionist attitude towards psychology, contemporary naturalists also miss the metaphysical character of the thesis of anti-individualism. With an approving reference to Millikan's work on biological function, Dennett says that it is a particular individual's historical origin that "licenses a certain way of speaking" (Dennett 1987, 292) of her states. It is on the basis of the selectionist history of an organism or the design history of an artefact that we may describe certain of their states as serving the purpose of representing their environment. However, this means that Dennett and other naturalists interpret anti-individualism as a mere *semantical* thesis. That is, anti-individualism turns out to be a claim about how certain internal states of the individual ought to be described—namely, by including the mention of some facts external to the individual, in particular those about her history. For Dennett, this semantical reinterpretation of anti-individualism is part of his rejection of what he calls *original intentionality*. This is the claim that humans possess genuine minds while other candidates (e.g., computers and other artefacts) have minds only by proxy. Dennett argues that there are no minds at the bedrock. All intentionality is derived.

Burge rejects both the reduction of psychology to biology and of the metaphysical thesis to a semantical thesis. As for the former, in his recent book, *Origins of Objectivity*, he asserts:

The explanatory content and goals of theories of perception and belief are not the same as those that underwrite biology. Explaining the way veridical and non-veridical representational states arise, given proximal stimulation, is a different explanatory enterprise from that of explaining any states in terms of their biological functions—their contributions to fitness. So biological explanations cannot reduce explanations whose point is to explain accuracy and inaccuracy of representational states. Since what they explain is different, the former cannot take over the job of the latter. (Burge 2010, 303)

As for the metaphysical character of anti-individualism, recall that Burge makes clear in his definition of the thesis that it concerns the very natures of mental representations. Thus, while no one can claim a monopoly on the usage

of technical philosophical claims such as “anti-individualism,” it appears that naturalists give it a substantially different meaning from Burge.

However, in this chapter, I do not wish so much as adjudicate the dispute between Burge and the Darwinian naturalists. Rather, I plan to analyze the arguments on both sides in order to understand the grounds of the disagreement better. I start with explicating the naturalists’ strategy to derive anti-individualism from adaptationism (section 2). In the process, I note some important differences between Dennett and Millikan, especially over intentional realism. It turns out that Millikan does not share Dennett’s view that all intentionality is merely derived; in this respect her position is closer to Burge’s. And yet, Burge rejects Millikan’s naturalism as well, on the grounds that her construal of representation is as far removed as Dennett’s from the actual practice of psychology (section 4). However, I shall also point out a certain discrepancy in Burge’s argument. We have already seen him appealing to a particular example of the biological organ, the heart, in explicating the character of the environmental dependence of mental representations. Elsewhere he used it to argue for the autonomous character of psychology. Now Burge’s example is repeated almost verbatim by Millikan in her theory of biofunctional explanation. And, curiously, the example implies an individualistic understanding of psychology that Burge officially rejects (section 3).

## 2 Adaptationism and Anti-Individualism

Let us start by looking at the details of anti-individualist arguments. This should be interesting because both Burge and the naturalists appeal to very similar thought experiments. These are the notorious thought experiments featuring physically identical, yet intentionally distinct, individuals, whose intentional difference is explained in terms of a difference between the two individuals’ social or physical environments. Burge summarizes both types of his thought experiment as follows:

Consider a person A who thinks that aluminium is a light metal used in sailboat masts, and a person B who believes that he or she has arthritis in the thigh. We assume that A and B can pick out instances of aluminium and arthritis (respectively) and know many familiar general facts about aluminium and arthritis. A is, however, ignorant of aluminium’s chemical structure and micro-properties. B is ignorant of the fact that arthritis cannot occur outside of joints. Now we can imagine counterfactual cases in which A and B’s bodies have their same histories considered in isolation of their physical environments, but in which there are significant environmental differences from the actual situation. A’s counterfactual environment lacks aluminium and has in its places a similar-looking light metal. B’s counterfactual environment is such that no one has ever

isolated arthritis as a specific disease, or syndrome of diseases. In these cases, A would lack “aluminium thoughts” and B would lack “arthritis thoughts.” Assuming natural developmental patterns, both would have different thoughts. (Burge 2007, 222-223)

Burge notes that the difference between the two individuals cannot be reduced to a difference in the causal origin of two tokens of the same type of a representation. Each token representation is of a different type because each has a different content. And the ascriptions of mental representations are “literal.”

Dennett’s story features artificial devices rather than human protagonists:

Consider a standard soft-drink vending machine, designed and built in the United States, and equipped with a transducer device for accepting and rejecting US quarters. Let’s call such a device a two-bitser. Normally, when a quarter is inserted into a two-bitser, the two-bitser goes into a state, call it Q, which “means”(note the scare-quotes) “I perceive/accept a genuine US quarter now.” Such two-bitserers are quite clever and sophisticated, but hardly foolproof. They do “make mistakes” (more scare-quotes). That is, unmetaphorically, sometimes they go into state Q when a slug or other foreign object is inserted in them, and sometimes they reject perfectly legal quarters—they fail to go into state Q when they are *supposed to*. (Dennett 1987, 290; emphasis in the original)

Now suppose one such vending machine is installed in Panama, where they use quarter-balboas, which are physically indistinguishable from quarter-dollars as far as the machine is concerned. So the two-bitser works correctly when accepting the quarter-balboas in this setting, though this would have counted as a mistake while the machine were located in the US. The question is how to identify the state the machine goes into when accepting balboas in Panama. An individualist would clearly say that no matter where it is located, the two-bitser goes into the same state Q; the only difference is its causal history. An anti-individualist suggests that while the *physical* state that the machine enters remains the same across the two environments, *intentionally* speaking the machine located in Panama goes into a different state—say, QB.

It is important to realize that unlike Burge, Dennett considers the choice between the two alternatives strictly speaking *indeterminate*. We can appeal to the fact that the vending machine is a functional device. It was designed by human engineers to serve certain purposes they had in mind. So there are some historical facts—facts about the origin—due to which the machine may be characterized as a device designed to give out soft drinks in exchange for the US quarters. If so, then also the state of the machine placed in Panama should be characterized in terms of the function it was selected for. With a reference to Millikan’s biological definition of function (the details of which we shall discuss in section 4), Dennett claims that

whether [the two-bitser's] Panamanian debut counts as going into state Q or state QB depends on whether, in its new niche, it was *selected for* its capacity to detect quarter-balboas—literally selected, e.g., by the holder of the Panamanian Pepsi-Cola franchise. If it was so selected, then even though its new proprietors might have forgotten to reset its counter, its first “perceptual” act would count as a correct identification by a q-balber, for that is what it would *now be for*. [...] If, on the other hand, the two-bitser was sent to Panama by mistake, or if it arrived by sheer coincidence, its debut would mean nothing, though its utility might soon—immediately—be recognized and esteemed by the relevant authorities [...], and thereupon its *subsequent* states would count as tokens of QB. (Ibid., 293; emphasis in the original)

Dennett is confident that Burge and other intentional realists would agree that intentional ascription in the case of artefacts is a matter of practical expediency, or perspective, or stance. In the case of persons, however, these realists would insist that there was a *fact of the matter* whether someone meant aluminium, or arthritis, or whatever. But I suggest leaving the controversy over intentional realism for the following section.

Rather, let us now turn to the justification behind Dennett and Millikan's identification of representational states in terms of natural functions or purposes. This natural teleology is justified by that particular interpretation of evolutionary biology that I mentioned in the beginning of this chapter—namely, adaptationism. According to the definition that I quoted, adaptationism is the conviction that natural selection is only a significant source of the observed diversity of living forms. But let us introduce some distinctions here. Sober's definition closely corresponds to what Peter Godfrey-Smith calls *empirical adaptationism*, namely the claim that

[n]atural selection is a powerful and ubiquitous force, and [...] [t]o a large degree, it is possible to predict and explain the outcome of evolutionary processes by attending only to the role played by selection. (Godfrey-Smith 2001, 336)

This should be distinguished from two stronger theses: *explanatory adaptationism*, according to which

[t]he apparent design of organisms, and the relations of adaptedness between organisms and their environments, are the *big questions*, the amazing facts of biology [...] Natural selection is the key to solving these problems; selection is the *big answer* (ibid.; emphasis in the original),

and *methodological adaptationism*, which says that

[t]he best way for scientists to approach biological systems is to look for features of adaptation and good design. Adaptation is a good “organizing concept” for evolutionary research. (Ibid., 337)

Godfrey-Smith argues that Dennett and Dawkins are explanatory adaptationists, sometimes even combining this—as when they marvel at the sheer *amount* of adaptive features in nature—with the empirical claim. I shall leave Dawkins aside, but I think that Dennett in particular actually subscribes to the strongest, i.e. methodological, adaptationism.

Consider such dramatic comments from Dennett’s popular book, *Darwin’s Dangerous Idea* (1995) as:

Adaptationist reasoning is not optional: it is the heart and soul of evolutionary biology. Although it may be supplemented, and its flaws repaired, to think of displacing it from central position in biology is to imagine not just the downfall of Darwinism but the collapse of modern biochemistry and all the life sciences and medicine. (Dennett 1995, 238)

It seems that Dennett claims here that the assumption of good adaptedness is not just a correct answer to the key question of biology, but precisely a “good organizing principle” of all the life sciences, without which they would be unthinkable. I shall elaborate on this in a minute. But first, I need to take note of the fact that precisely the radical challenge to adaptationism that Dennett finds unthinkable arose in the midst of the biological science, in the famous paper by Stephen J. Gould and Richard Lewontin “The Spandrels of San Marco and the Panglossian Paradigm” (1978). In it, Gould and Lewontin deplore the assumption of

the near omnipotence of natural selection in forging the best among possible worlds. This program regards natural selection as so powerful and the constraints upon it so few that direct production of adaptation through its operation becomes the primary cause of nearly all organic form, function, and behaviour. (Gould and Lewontin 1978, 76)

For Gould and Lewontin, many adaptationist explanations are unfalsifiable “just-so stories,” and many alleged adaptations are mere “spandrels”—non-optimal by-products of a variety of constraints on natural selection. Godfrey-Smith argues Gould and Lewontin seek to undermine both empirical and methodological adaptationisms. Or, more precisely, they wish to uproot methodological adaptationism by depriving it of the support it gets from the alleged empirical evidence of good design.

Now back to Dennett’s notion of adaptationism. Compared to Millikan who dismisses Gould’s arguments, especially his rejection of adaptive character of cognitive capacities (see Millikan 1993, 46-47), Dennett is conciliatory. He interprets Gould and Lewontin’s critique as a useful reminder that we should be careful, not hasty, adaptationists. But he rejects their suggestion to supplant adaptationism with the idea of *Baupläne*—the by now largely defunct theory that adaptation was good enough to explain certain superficial features

of the design of organisms, but not their fundamental “body plans” (cf. Gould and Lewontin 1978, 85-89). Drawing on his distinction between “cranes” and “skyhooks” (cf. Dennett 1995, 73), Dennett asks what else than a mysterious skyhook could pull a complete body plan into existence, if the humble crane of natural selection were prohibited? Yet the most important point brought up by Dennett against Gould and Lewontin’s critique is that they misunderstand the nature of adaptationism. The latter is not strictly speaking a theory. Only a theory—a collection of claims—could be either falsified or unfalsifiable. Rather, adaptationism is a stance that biologists are bound to adopt vis-à-vis the process of natural selection, lest they miss certain real patterns in nature.

The concept of a stance is, of course, an import from Dennett’s philosophy of mind. Dennett argues that something is a bearer of representations only from the standpoint of an “intentional stance,” which ascribes these representations under the constraint of an ideal rationality. As he puts it

*all there is to being a true believer is being a system whose behavior is reliably predictable via the intentional strategy, and hence all there is to really and truly believing that  $p$  (for any proposition  $p$ ) is being an intentional system for which  $p$  occurs as a belief in the best (most predictive) interpretation. (Dennett 1987, 29; emphasis in the original)*

Thus, the intentional stance licenses mentalism, while adaptationism licenses a sort of natural teleology. The two stances are analogous in that the former makes an assumption of rationality, whereas the latter that of optimality of design:

*When we adopt the intentional stance toward a person, we use an assumption of rationality or cognitive/conative optimality to structure our interpretation [...] In biology, the adaptationists assume optimality of design in the organisms they study. (Dennett 1990, 187; emphasis in the original)*

Far from being expressions of naïve optimism either with respect to the rationality of agents or the optimality of organisms, the intentional stance and adaptationism are necessary presuppositions for answering certain “why”-questions. In the domain of psychology, we ask why an agent engaged in this or that behaviour; and we proceed by inquiring into what a perfect reasoner would do, given her circumstances; and in due time we are bound to discover that no agent is perfect in her reasoning. In biology, we ask why an organism was designed in a particular way; we hypothesize how it should be optimally designed, given what we know about its environment; and our prediction is going to be falsified (*pace* Gould and Lewontin, an adaptationist story that were too perfect would be useless to biologists, since it would teach them too little). And this is the gist of Dennett’s position concerning adaptationism as a stance, which I think justifies its classification as a methodological, not just explanatory, thesis: “Adaptationism and mentalism [...] are not theories in one traditional sense. They



are stances or strategies that serve to organize data, explain interrelations, and generate questions to ask Nature” (Dennett 1987, 265).

In view of the above, I think that Dennett’s critics—among whom, as we shall see in the following section, we must count Burge—who ascribe to him the idea that stances are adopted or vacated arbitrarily or opportunistically are too quick.

It seems, then, that adaptationism is the basic premise of a Darwinian theory of the mind. As I already mentioned, adaptationism enables a sort of natural teleology. And once there is a room for natural purposes, we are entitled to posit states whose function it is to represent an external environment. Biological functions are identified in terms of a past performance of the traits that are so functionally defined. Hence, representational states, too, are identified in terms of what their predecessors were supposed to do. Finally, anti-individualism follows as a natural upshot of a theory that identifies all traits of an organism, including its representational states, in terms of its historical relations.

This, I take it, is the basic structure of a Darwinian argument justifying a version of anti-individualism shared by Dennett and Millikan. I shall return to some differences between the two philosophers’ specific construal of this argument in a moment. At this point, it is clear that they agree in seeing anti-individualism as an almost trivial outcome of a psychology understood as an integral part of cognitive ethology. Dennett argues that cognitive ethology becomes anti-individualistic once freed from the legacy of *behaviourism*. He cites a particular research on vervet monkeys that have developed, in their natural habitat in the Sub-Saharan Africa, a relatively elaborate communication system involving several types of warning calls signifying the presence of different kinds of predators (see Cheney and Seyfarth 1990). Most of what is interesting about the lives of these animals would be simply missed from the perspective of behaviourism. For example, only from the intentional stance can we recognize a monkey as *deceiving*—i.e., as wishing to be believed as believing something that it does not. Certain fruitful hypotheses can be framed only based on the assumption that the vervets are intentional agents, i.e. having beliefs, intentions, fears and a host of other propositional states in the aetiology of their behaviour. Dennett argues that cognitive ethologists actually adopt the intentional stance towards the animals whose behaviour they study (see Dennett 1987, chap. 7 and Dennett 1998, chap. 20).

Now, the contrast between mentalism and behaviourism within ethology is important for the controversy over psychological anti-individualism in that behaviourism is a paradigmatically individualist doctrine. A behaviourist takes into account only the narrowly conceived inputs and outputs of behaviour, so that two individuals there were behaviourally identical would be psychologically identically as well. It is thus significant that Dennett stresses that, from the

intentional stance, hypotheses about what the observed individuals believe and desire need to be framed “by figuring out what they ought to believe and desire, given their circumstances” (Dennett 1998, 292). Only within the particular circumstances of their natural environment could we figure out what, if anything, vervet monkeys’ calls possibly mean, and we are licensed to attribute to these animals corresponding mental states.

Millikan also construes intentional psychology as part of cognitive ethology, but she sees the latter as haunted by the legacy of *individualism*, rather than behaviourism. She writes:

Will a mature cognitive psychology need to characterize its subjects in ways that make reference to how they are imbedded in their environments? Or will it be “individualistic,” making reference only to what supervenes on the structures of individual bodies and brains? The individualists argue that the behavioral dispositions of a person clearly depend only on that person’s *inner* constitution, and hence that there can be no need to refer to the individual’s relation to the wider environment in order to explain them. The anti-individualists argue that it is impossible even to describe much of the behavior that it is psychology’s job to explain without reference to the environment. (Millikan 1993, 135; emphasis in the original)

As I pointed out in the previous paragraph, behaviourism is naturally interpreted as a type of individualism, but Millikan doesn’t reject behaviourism *tout court*. She thinks individualistic and non-individualistic versions of behaviourism could be distinguished. She goes on to spell out the anti-individualistic notion of behaviour with the help of her concept of “proper function,” to which we shall return in the next section. Here it suffices to say that the proper function of a trait is identified as that function the (by and large) successful performance of which enabled the trait’s ancestors to copy or reproduce themselves in subsequent generations. A proper identification of a trait, including a behavioural trait, thus has a necessarily external element in the form of an *historical* dimension. But Millikan adds that the present, not just past, relations between behaviours and their environments are a necessary condition of a correct identification of behaviours. Millikan goes as far as suggesting that psychology, properly construed as an integral part of ethology, needs to construe its subject matter broadly, as involving both an organism narrowly conceived *and* its natural habitat.

It is a very serious error to think of the subject of the study of psychology and ethology as a system spatially contained within the shell or skin of an organism. What is inside the shell or skin of the organism is only half of a system; the rest, if the organism is lucky, is in the environment. The organismic system, especially (indeed, by definition) the behavioral systems, reach into the environment

and are defined by what constitute proper, or normal, relations and interactions between structures in the organism and in the environment. (Ibid., 158)

Millikan's broad concept of behaviour is attractive and she seems to be careful enough not to extend her thesis into an implausible claim that mental representations themselves, let alone the mind, stretch into the external world. Burge himself argues for a broad construal of behaviour (see Burge 2007, 227), while deploring the tendency, popularized by some idiosyncratic anti-individualists, to "extend the mind" beyond the bounds of the body of an individual (Clark and Chalmers 1998).

Earlier, I announced that, despite a broad agreement between them, there are important differences between Dennett and Millikan's respective versions of a Darwinian theory of the mind. In the remainder of this section, I shall touch on two closely related points of difference.

Dennett seems to assume that the intentional stance is somehow basic in both psychology and biology. We start by making assumptions about what a rational agent would do or what an optimal design should look like before we inquire about the functional architecture of the agent or the evolutionary origin of an organism. By contrast, Millikan argues that the intentional stance needs to be underwritten by what Dennett calls the "design stance."

where one ignores the actual (possibly messy) details of the physical constitution of an object, and, on the assumption that it has a certain design, predicts that it will behave *as it is designed to behave* under various circumstances. (Dennett 1987, 16-17; emphasis in the original)

For Millikan, the fact that something is at all interpretable from the intentional stance is evidence that it was designed: "There is nothing that exhibits apparently rational patterns for any time or in any detail that was not designed to do so, either by natural selection, or by something that natural selection designed" (Millikan 2000, 60). In his response to Millikan, Dennett concedes that the design stance *is* more basic "in the sense [Millikan] defends" (Dennett 2000, 341). That is, anything that is capable of a rich diversity and flexibility in its behavioural and perceptual responses is bound to have been designed either artificially or, ultimately, by means of natural selection. And yet, until a more principled way of distinguishing real intentional systems from merely apparent ones become available, Dennett says he prefers to keep his more "open-ended" approach that licenses the adoption of the intentional stance even toward simple artefacts (thermostats) or primitive organisms (frogs).

This last point ultimately rests on our two philosophers' divergent views of the reality of mental states. Millikan is a realist who believes that "folk psychology postulates inner items (for example, structures or events or states or entities)," and that "folk psychology is probably right" (Millikan 1993, chap. 3).

As for Dennett, he occasionally committed the tactical error of embracing the label “instrumentalist” to describe his theory of representational states (Dennett 1987, chaps. 2 and 3). Beside the fact that instrumentalism proves too difficult to distinguish from fictionalism, or just plain old anti-realism, despite Dennett’s valiant effort (cf. Dennett 1987, 69-81), I think no single label is going to do justice to the complexity of his theory. It is true that some of his claims sound straightforwardly anti-realistic: “Folk psychology is *abstract* in that the beliefs and desires it attributes are not—or need not be—presumed to be intervening distinguishable states of an internal behavior-causing system” (Dennett 1987, 53). In other words, it is unlikely that scientific psychology will discover discrete items in the brain that corresponded to the beliefs and desires postulated by folk psychology. So Dennett suggests that we split folk psychology—which is a sort of a mixed bag, as it is couched in semantic terms, yet also postulating a particular sort of entities—into two new theories. On the one hand, there would be the “intentional systems theory”—i.e., an abstract science of rationality, akin to decision theory or game theory—and the “sub-personal cognitive psychology”—i.e., a concrete science of the neural systems. The former would be dealing in pure semantics, the latter in pure syntax. I take it that the construal of the intentional systems theory as a purely abstract theory is supposed to guarantee the metaphysical sanity of Dennett’s willingness to attribute intentional states to natural selection itself. Millikan probably reads her own intentional realism into Dennett’s theory when she finds the talk of beliefs and reasoning of Mother Nature “otiose in biology” (Millikan 2000, 65). She elaborates that “there is no sense in such talk *because there is nothing in Nature analogous to beliefs and nothing that so much as reminds one of inference*” (ibid., 64; emphasis in the original). I think Dennett’s response to this is ingenious:

As Sherlock Holmes, the patron saint of inference, famously said, once you have eliminated all other possibilities, the one that remains, however improbable, must be the truth. Is that not an inference? Does not Mother Nature eliminate all other possibilities, on a vast (not actually Vast) scale, thereby “inferring” the best design? When Deep Blue eliminates a few billion legal moves and comes to rest on one brilliant continuation, it surely reminds Kasparov of inference! (Dennett 2000, 343)

While I am convinced that Dennett’s attribution of intentions to natural selection is not metaphysically weird for the reasons offered by Millikan, I *do* find a metaphysical difficulty within Dennett’s position. I shall elaborate on it in the following section.

### 3 Deflationism and Realism

Burge rejects Dennett's construal of mental representation. He finds this view not only implausible, but obviously so: "I mention [it] only to lay it aside" (Burge 2010, 293). He dubs it the "deflationary view" of representation:

On this view, treating something as engaging in representation is merely a matter of a "stance," with more or less practical or instrumental value. On such a position, there is no objective kind, *representation*, that can be discovered through normal scientific investigation. On such a position, there is no more *theoretical* reason to treat an individual as having beliefs or perceptions than there is to treat a vending machine, or a planetary system, as representing something. It is all a matter of practical convenience or optional attitude toward the phenomena. (Burge 2010, 293; emphasis in the original)

Dennett returns the favour by dismissing Burge's position as one more example of the traditional belief in the "original intentionality" of human minds, from which all other intentionality—ascrivable, as the case may be, to the artefacts of our own design, or to the inanimate objects of nature—is "derived."

Superficially at least, Burge unites with Millikan against Dennett with respect to the issue of the reality of representational states. Burge and Millikan are intentional realists while Dennett is a sort of anti-realist about the mind. Ultimately, however, Burge is going to classify Millikan's theory, despite its realism, as another variant of the sort of naturalism of which an anti-realist version is Dennett's view. Millikan as well as Dennett turn out to be equally unacceptable to Burge as two models of a basically reductionist view of the representational mind. Yet the arguments on both sides are subtle. I propose, first, to rehearse Dennett's reasons, derived from his understanding of Darwinism, for intentional anti-realism; second, I examine Burge's grounds for rejecting both Dennett and Millikan's naturalistic theories.

Recall that Dennett starts motivating his anti-realism by construing the thought experiment placing a lowly artefact, not a person, in two different environments. Dennett expects that everybody is going to agree that representational states cannot be attributed to mere artefacts literally, so that an uncertainty as to whether to attribute one state rather than another is not disquieting. Next, Dennett needs to demonstrate that a similar indeterminacy befalls intentional description of persons as well. In other words, he needs to show that in the case of persons as well as artefacts, there is no bedrock fact when it comes to possessing representational states, but a mere useful way of speaking. For that purpose, Dennett offers an additional thought experiment.

Suppose someone decided to survive into the twenty-fifth century in a hibernation device of some sort. He would be wise to make that device mobile, so that it can look for the sources of energy. And since these are bound to be scarce,

the mobile hibernation device should be capable of fighting off the machines of other people who—as it might be expected—would build survival machines of their own. The more sophisticated such machines get, the better their chance to deliver their hosts into the future. Hence we might expect the best machines would be robots capable of self-control, of setting their own goals based on their assessment of a current situation, and so on. Now, the intentional realists such as Burge would, according to Dennett, insist that such robots, no matter how sophisticated, have whatever fake intentionality they possess, ultimately derived from our plans and purposes. But here is the clincher: “the conclusion forced upon us is that our own intentionality is exactly like that of the robot, for the science-fiction tale that I have told is not new; it is just a variation of Dawkins’ [*The Selfish Gene*] vision of us [...] as ‘survival machines’ designed to prolong the futures of our selfish genes” (ibid., 298). Where Dennett’s story started with the real meaners as the ultimate source of design, it turns out those meaners mean no more literally than the selfish genes of Dawkins’s story. And yet, Darwinism shows that we can get intelligent design without any real minds: “when natural selection selects, it can ‘choose’ a particular design *for one reason rather than another*, without ever consciously—or unconsciously!—‘representing’ either the choice or the reasons” (ibid., 299, emphasis in the original). Hence we see again that not only can we attribute intentions to Mother Nature, despite the fact that she is no real reasoner, but a mere process of natural selection—but we do the same with persons, despite the fact that, strictly speaking, *they* are no real reasoners, either. There are only *ersatz* thinkers, but anything can be considered as such, if selected properly either artificially or naturally.

This is how it works according to Dennett. We attribute beliefs, desires and other attitudes to each other, but there is no way these folk psychological states—imagined to be both semantic and holistic, as well as concrete and discrete, entities—are going to be recognized by a mature science of psychology. Therefore, Dennett suggests splitting folk psychology into two new theories: “one strictly abstract, idealizing, holistic, instrumentalistic—pure intentional system theory—and the other a concrete, microtheoretical science of the actual realization of those intentional systems—what I will call sub-personal cognitive psychology” (Dennett 1987, 57). At the intentional system level, we are semantic engines; at the microtheoretical level, we are physical, or perhaps syntactic systems. How do these two levels of description relate to each other? In other words, how does the brain, a mere syntactic engine, produce semantics? Dennett answers:

It cannot be designed to do an impossible task, but it could be designed to *approximate* the impossible tasks, to *mimic* the behavior of the impossible object (the semantic engine) by capitalizing on close (close enough) fortuitous correspondences between structural regularities—of the environment and of its own internal states and operations—and semantics types. (Ibid., 61, emphasis in the original)

As far as I can see, Dennett suggests that the brain behaves *as if* it were a semantic engine, in addition to being a syntactic one. We don't have an ability to build brains (yet), but we can build much simpler devices that fulfill some semantically characterizable tasks. For instance, we could build a machine that would catch the telephone communications that are death threats, by picking out words like "... I will kill you..." or "... you... die... unless..." and such (cf. *ibid.*, 62). If so, we would succeed in building a "death-threat interceptor"—that is, a purely syntactic device which is also describable in such semantic terms. The machine would be primitive and unreliable, but we could keep on improving it.

This much could be achieved by artificial design, but what about natural selection? Dennett claims that our brains are dumb syntactic devices that were selected for their ability to mimic semantic engines, and have kept on getting better at this over time: "in the end all one can hope to produce (all natural selection can have produced) are systems that *seem* to discriminate meanings by actually discriminating things (tokens of no doubt wildly disjunctive types) that co-vary reliably with meanings" (*ibid.*, 63; emphasis in the original). So we can interpret each other intentionally owing to a long history of a (more or less) successful coping of our species with its environment. It is due *solely* to the benefit of hindsight afforded by this history that we can *appear* to be reasoners and meaners.

Burge dismisses this whole approach because he disagrees that the sort of responsiveness to stimuli that could be found in nearly all living things captures the kind *representation* employed in psychological explanation. In his critique of Dennett, Burge makes a point he has repeated in polemics with many naturalists over the decades. The point is that these authors understand the relation between science and metaphysics backwards. Their projects are driven by various metaphysical interests, in particular by the interest to make representation and the mind non-mysterious. For example, one worry that seems to have motivated a lot of attempts in the past few decades to naturalize intentionality is epiphenomenalism. Many philosophers assumed that all the causal work is done by the underlying physical processes, while representations *qua* representations are causally inert. But, according to Burge, we should eschew such preconceptions of a (materialistic) metaphysic and instead begin with studying the actual practice of psychological explanation in which intentional idioms figure prominently. Such explanation works in everyday contexts and it is part of a mature scientific psychology as well. Last but not least, it is central to our self-image as agents (see Burge 2007, chap. 16). Hence, there is nothing *prima facie* mysterious about mental representation as it figures in a successful, testable and precise psychological explanation.

When it comes to Dennett, it seems that his project of replacing folk psychological concepts—despite the fact they are commonplace in research pro-

grammes of perceptual psychology and elsewhere—with the two new theories of abstract and sub-personal psychology, respectively, is *stipulative*. It is driven by a metaphysical worry that brains do not possess semantic properties. And Dennett's conclusions are unclear. He doesn't seem to make up his mind as to how seriously should the talk of mental states be taken. On the one hand, when he says that brains “mimic” semantic engines, or that intentional ascription is indeterminate, representations seem merely useful fictions. On the other hand, he also suggests that brains “realize” semantic states (Dennett 1987, 67). The relation of realization, familiar from the functionalist literature, seems more robust than mere “mimicking.” Yet Dennett does not elucidate how similar or different these two relations are.

Burge raises a similar charge of stipulativeness against Millikan's account of representation: “I believe that Millikan's view amounts to a stipulation about how she intends to use ‘representation’” (Burge 2010, 300). He does praise her account for separating representation from mere information-carrying. The latter is straightforwardly causal, so that an organism goes into a particular information-carrying state whenever the appropriate causal prompt is present. There is no room for misrepresentation or mistake. We saw how that room is created by Millikan's appeal to selectionist history. A state of an organism can *misrepresent*, if there is a norm set by the past performance of the ancestors of that state, since then we can say how the state is supposed to function, even though it actually doesn't. There is nothing in this notion that precludes its ascription even to artefacts, as we saw in Dennett's case. The same phenomenon could be described just by using the notion of biological function, normal environmental conditions and sensory discrimination.

#### 4 Historical Function, Systems Function, and Individualism

In this last section, I should like to focus on the concept of function appealed to by Dennett and Millikan. It should become clear that this concept motivates a version of anti-individualism, which is actually incompatible with Burge's original theory. Which is another way of saying that Dennett and Millikan's respective construals of anti-individualism differ from Burge's. It is then surprising to find even Burge, as he does in his defence of the autonomy of psychology, to the biofunctional concept of function, because it results in an inconsistency.

As is well-known, there are two main concepts of function: Millikan's historical theory of function and Robert Cummins's systems theory. I have already explained some elements of Millikan's theory earlier in this chapter. A comparison with Cummins's view might help further to clarify the nature of Millikan's view.

Cummins picks on the complexity of systems of various sorts and ascribes functions to such systems on the basis of the workings on their parts. To be



more precise, an item  $x$  has a function  $\varphi$  within a system  $s$ , assuming a background of an analytic explanation of  $x$ , which appeals to  $x$ 's capacity to  $\varphi$  in  $s$ . Cummins uses an example of the heart to illustrate his proposal: "It is appropriate to say that the heart functions as a pump against the background of an analysis of the circulatory system's capacity to transport food, oxygen, wastes, and so on, which appeals to the fact that a heart is capable of pumping" (Cummins 1975, 64). Although this example is taken from biology, notice that Cummins could apply his approach in assigning functions to inanimate systems as well—thus, our solar system could be regarded as a functional system. Also notice that Cummins refers only to the current properties of a functional item, and that he confines his attention to the internal parts of a system.

Curiously, Millikan also illustrates her alternative theory of function with the example of the heart. Let me quote a relevant passage in its entirety:

A heart, for example, may be large or small (elephant or mouse), three-chambered or four-chambered, etc., and it may *also* be diseased or malformed or excised from the body that once contained it, hence unable to pump blood. It falls in the category *heart*, first, because it was produced by mechanisms that have proliferated during their evolutionary history in part because they were producing items that managed to circulate blood efficiently in the species that contained them, thus aiding the proliferation of that species. It is a *heart*, second, because it was produced by such mechanisms in accordance with an explanation that approximated, to some undefined degree, a Normal explanation for production of such items in that species and bears, as a result, some resemblance to Normal hearts of that species. By a "Normal explanation" I mean the sort of explanation that historically accounts for production of the majority of Normal hearts of that species. And by a "normal heart," I mean a heart that matches, in relevant respects, the majority of hearts that, during the history of that species, managed to pump blood efficiently enough to aid survival and reproduction. (Millikan 1993, 55)

Although Millikan agrees with Cummins in assigning the same function of blood-pumping to the heart, the rationale is importantly different. A particular exemplar of the heart has the function that it does in virtue of its ancestry. There has been a long line of organs that more often than not succeeded in pumping blood in the past and the present exemplar is their descendant. Therefore, even if our particular exemplar fails to pump blood efficiently, or even if it is so defective as to never having pumped any blood, we can still correctly identify it as the kind *heart* in virtue of its relation to the line of ancestral hearts that have enabled the survival up to now. In other words, we can assign a function properly in virtue of a background of normality. (That is why Millikan speaks of "normal explanation").

We can see now that the historical concept of function is narrower than the systems view, in that the former is restricted to items that are products of design

of one sort or another, whereas the latter was applicable to any complex system, whatever its origin. In another sense, however, Millikan's concept of function is obviously broader than Cummins's concept. The historical function is identified as such in relation to things outside of the system of which it is a part of, including things in a distant past. By contrast, the Cummins function disregards relations between a functional system and its surroundings or its origin.

We have already seen in section 2 that the historical account seems well suited for the purposes of establishing anti-individualism, since it enables us to run the familiar thought experiments. Dennett's thought experiment featuring a vending machine whose powers to detect currency are affected by an environment directly draws on Millikan's construal of functional ascription. In section 3, I explained that Burge does not like Dennett's conclusion that functional, hence semantic, ascription remains forever indeterminate. But Burge does not like Millikan's account of representation, despite its intentional realism, either. I showed that Burge sees these naturalist theories as two versions of "deflationism," namely a tendency to stipulate various minimal detection capacities in place of the robust concept of representation, which is at home in everyday life and scientific psychology. Recall that for Burge, psychology types its kinds anti-individualistically, yet there is no need to reduce it to some lower, presumably individualistic, level of discourse.

In view of this critique of deflationism, it is then surprising to find Burge supporting his view of psychology by means of a twin story featuring the biological item that we have already seen in both Cummins and Millikan—the heart. Burge invites us to imagine a physical replica of the human heart placed in an alien body:

Something is a heart because its organic function is to pump blood in a circulatory system that extends beyond the surfaces of the heart. One can imagine an organ in a different sort of body with a totally different function (it might pump waste for example). The causal powers attributed to such an organ by biology would be different from those attributed to a heart. Such an organ would not be a heart, but it might be chemically and structurally homologous to a heart. (Burge 2007, 323)

Like Millikan, Burge is explicit that in order to categorize properly the physiological kind *heart*, we must attend to something external to its instantiations—namely, their selectionist history:

To be a heart, an entity has to have the normal, *evolved* function of pumping blood in a body's circulatory system. One can conceive of a physically homologous organ whose function is to pump waste—or even a physically homologous entity that came together accidentally and lacks a function. Such entities would not be hearts. (Ibid., 326; emphasis added)

Thus, in his claim that the heart and the alien waste-pump differ in terms of their divergent histories, Burge assumes the historical notion of function.

The point of presenting physiology as another special science that types its kinds anti-individualistically is to suggest that intentional psychology is no worse. If physiology enjoys respectability at least comparable to that of physics, then psychology should not be looked down upon, either. In accordance with his methodological decision to prioritize an actual scientific practice over metaphysics, Burge seems to be suggesting that the claims of naturalism are satisfied by taking a successful explanatory practice in psychology and other special sciences at its face value. There is no call for trying to force that practice into a straightjacket of some reductive metaphysics.

However, I wonder whether the above defence of the respectability of intentional psychology is entirely consistent with Burge's anti-individualism. It is true that Millikan's historical concept of function, exploited by both Burge and Dennett, appears to be anti-individualistic in character compare to Cummins's theory. For Cummins, who takes into account only the internal parameters of a functional item, there is no way to distinguish between a blood-pump and a waste-pump. In a parallel case, an individualist in psychology has no way of distinguishing between aluminium-thought and twin aluminium-thought as long as he restricts his attention to internal parameters of a thinker alternating between two environments. Millikan provides a resource for drawing the required distinction. A heart is distinct from a waste-pump in terms of its divergent evolutionary origin.

Yet recall Burge's recent definition of anti-individualism that I quoted at the outset. It is meant as a theory of the very *nature* of representational mental states. That means it is not merely a theory of *description* of these states. It does not merely say that in speaking of representations, we should mention their relations to an environment. Dennett could be easily critiqued as misinterpreting anti-individualism as a descriptive, rather than metaphysical, theory. In his thought experiment about the vending machine, we saw him making an explicit assumption that the machine goes into a particular state that could be described differently relative to an environment. If so, the nature of such a state should be identifiable independently of its various environmental descriptions. In fact, Dennett in at least one text admits as much. For evidence of Dennett's betrayal of externalism, see his response to Frank Jackson: "So let me confirm Jackson's surmise that I am a behaviorist; I unhesitatingly endorse the claim that 'necessarily, if two organisms are behaviorally alike, they are psychologically exactly alike'" (Dennett 1993, 923). Vending machines certainly behave the same: accepting coins and churning out bottles of soft drink. So some narrow description of what they represent must in principle be available, too. The description of machines which takes into account such facts as their location in the US or Panama, respectively, is something extraneous. And this might perhaps help to find a correlate of environmentally identified state at the sub-personal, or syntactic, level.

Now, the above critique of Dennett as a closet individualist, or perhaps someone who just misunderstood anti-individualism, can unfortunately be extended to Burge's own theorizing about the autonomy of intentional psychology. It is crucial to realize that there is a narrow, individualistic way of identifying both the heart and its alien counterpart: they are both a kind of *pump*. If so, an environmental description of this organ is strictly speaking optional, since we can descend to a lower, individualistic level. It is true that Burge is speaking of biology (or perhaps physiology), but the point of his example clearly is that psychology is analogous to biology. In section 1, I quoted Burge putting the heart example to a somewhat different use—namely, arguing that this organ does not consist of its external relations—but even here, he said: “I think that representational mind is like that” (Burge 2010, 66). So, Burge appears to take an analogy between psychology and biology very seriously. And yet, elsewhere he opposed attempts to dilute anti-individualism to a mere theory of description in terms of environmental relations. He said that in lower-level science, we often do have alternative ways of identifying the instances of explanatory kinds. In psychology, this is not available. “We have no such ways of identifying states of the body that (putatively) are beliefs, independently of assumptions about the beliefs” (Burge 2007, 353). Accordingly, though both the heart and an alien organ belong to the kind *pump*, there is no way to identify a thought individualistically.

I conclude that both naturalistic attempts to found representation ultimately in evolutionary biology and Burge's nonreductive attempt to preserve the autonomy of intentional psychology are ridden with problems. Burge may be right that what Dennett or Millikan succeeded in deriving from biology is not representation, as it is understood in everyday life and psychology, but something too minimalistic. On the other hand, Burge, despite his claim that a widespread fear of reductive metaphysics is simply a prejudice, still seems wishing to connect psychology with biology, which entangles his account in an inconsistency.

## References

- Burge, T. (2007). *Foundations of Mind*. Oxford: Oxford University Press.  
 \_\_\_\_\_ (2010). *Origins of Objectivity*. Oxford: Oxford University Press.  
 Cheney, D. L. and Seyfarth, R. M. (1990). *How Monkeys See the World*. Chicago and London: The University of Chicago Press.  
 Clark, A. and Chalmers, D. (1998). The Extended Mind. *Analysis* 58, (1): 10–23  
 Cummins, R. (1975). Functional Analysis. In Sober 1994: 49–69.  
 Dennett, D. (1987). *The Intentional Stance*. Cambridge, Mass.: The MIT Press.  
 \_\_\_\_\_ (1990). The Interpretation of Texts, People and Other Artifacts. *Philosophy and Phenomenological Research* 50, Supplement, Fall, 177–194.

- \_\_\_\_\_ (1993). The Message Is: There Is No Medium. *Philosophy and Phenomenological Research* 58 (4): 919-931.
- \_\_\_\_\_ (1995). *Darwin's Dangerous Idea*. New York: Simon & Schuster.
- \_\_\_\_\_ (1998). *Brainchildren*. Cambridge, Mass.: The MIT Press.
- \_\_\_\_\_ (2000). With a Little Help from My Friends. In Ross, Brook and Thompson 2000: 327-388.
- Godfrey-Smith, P. (2001). Three Kinds of Adaptationism. In *Adaptationism and Optimality*. S. H. Orzack and E. Sober (eds.). Cambridge: Cambridge University Press, 335-357.
- Gould, S. J and R. C. Lewontin (1978). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. In Sober (1994): 73-90.
- Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: The MIT Press.
- \_\_\_\_\_ (2000). Reading Mother Nature's Mind. In Ross, Brook and Thompson (2000), 55-75.
- Ross, D., A. Brook and Thompson, D. (eds.) (2000). *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, Mass.: The MIT Press.
- Sober, E. (ed.) (1994). *Conceptual Issues in Evolutionary Biology*. 2<sup>nd</sup> Ed. Cambridge, Mass.: The MIT Press.
- \_\_\_\_\_ (1996). Six Sayings about Adaptationism. In *The Philosophy of Biology*, D. L. Hull and M. Ruse (eds.) Oxford: Oxford University Press 1998, 72-86.

