Burge on representation and biological function

Agustín Vicente

Ikerbasque: Basque Foundation for Science & University of the Basque Country UPV/EHU

Abstract

In *Origins of Objectivity*, Burge presents three arguments against what he calls 'deflationism': the project of explaining the representational function in terms of the notion of biological function. I evaluate these arguments and argue that they are not convincing.

Introduction

In *Origins of Objectivity*, Burge defends that the representational function, i.e., the function of representing veridically, or accurately, is not reducible to any other function; in particular, it is not reducible to any biological function[1]. He nonetheless claims that the representational function *is* a natural function or implements natural norms, in the sense that its conditions of good and bad functioning are "[independent] of any individual's having a positive or negative attitude toward the function or the norm" (p. 311)[2]. These two claims, namely, (a) that the representational function is a natural function, and (b) that the representational function is not a biological function, partly capture the contrast between Burge's own distinctive position and his two targets thorough the book: Individual representationalism (IR), and deflationism. IR is characterized by holding that the subject cannot have representations of the physical environment if he is unable to represent some *preconditions* which are constitutively required for such representations (e.g. the distinction between *seems* and *is*). Burge takes IR to exemplify a pernicious tendency in philosophy to hyper-intellectualized analysis. According to him, that subjects have representational capacities is independent of their personal-level capacities, knowledge or attitudes: representational capacities are *autonomous*. Deflationism, in turn, is characterized by the denial that the notion of representation is primitive. Deflationists attempt at explaining the notion in terms of

[1] Burge is only concerned with *perceptual* representation. So will I.

[2] The quote continues: "Specifically, the applicability of natural norms is independent of any individual's setting or acceding to them –accepting them as applicable".

other, more primitive, notions, their favourite candidate being the notion of biological function. Burge, however, holds that the notion of representation is irreducible. In what follows, I will examine Burge's arguments against deflationism and will try to show that they are not good.

Burge's arguments

In ch. 8 of *Origins* Burge presents three arguments, or considerations, against deflationism. The first alludes to the motivations of deflationism. Deflationism is the result of the so-called 'naturalization program' in the philosophy of mind, which attempts to explicate all mental notions in terms of "kosher" notions drawn from the natural sciences. The standpoint and main intuition of the naturalization program is that mental properties are mysterious. Burge rejects this presupposition: according to him, there is no reason to think that psychological properties are any more mysterious than the properties of any other successful science. We can call this argument 'the argument from motivation'.

The second argument that Burge presents is that the representational function is not a biological function because it cannot be accounted for in terms of adaptability: an accurate, or veridical, representation is not *ipso facto* an adaptive representation. Accurate representations do not have to maximize fitness in order to be accurate, i.e., correct. Though representations usually have a biological function, e.g. to trigger a fleeing behaviour, the proper, individuating, function of a representational state is to represent accurately, not to cause the kinds of behaviours it reliably causes. This argument, thus, tells us that the representational function is in fact irreducible to the notion of biological function. I propose to call it 'the argument from irreducibility'.

Burge's third argument is that deflationism fails to draw a boundary where there is one. There is a distinction between *merely* sensory systems and representational systems. IR tried to honour this distinction, but it drew it at the wrong place, with the result that only individuals that have some quite powerful cognitive capacities turn out to be able to represent environmental objects and attributes. However, deflationism treats sensory discrimination and representation as essentially the same kind of thing, and they blur a distinction that has to be honoured. If to represent *F* is, e.g., to have the function of carrying the information that *F*, then, given that merely sensory systems are also

information-carriers, there is no principled distinction between representation and sensory discrimination. I will call this 'the argument from blurring'.

The argument from motivation

These three arguments are not of equal strength. The first one recaps Burge's attitude towards the naturalist/physicalist enterprise (see, e.g. Burge, 1993). Burge thinks that our metaphysics should be dictated by our scientific explanatory practices. So, if psychology, as a successful science, makes essential use of the notion of representation, there is no reason even to wonder whether representation is or is not something else. However, this kind of consideration will not convince those who still think that there is something *prima facie* mysterious in mental properties. Mental properties, it is usually claimed, seem to be different from the rest of the properties postulated by the different sciences in their aboutness or their phenomenal character. The difference lies in that the properties invoked by other sciences seem to admit explanations in terms of the properties of physics, while the essential features of mental properties induce to think that physics does not have the resources to account for them. The philosopher who feels the urge to explain mental notions in terms of other notions may be wrong about the relationships between the other natural sciences and physics, but if one parts from the presumption that all the natural sciences exhibit some kind of unity, it is justified to try and legitimize psychology by showing that its properties are also amenable to unification.

Apart from this, the naturalistic philosopher does not need to talk about alleged mysteries to justify her project. It is an open question in the philosophy of science whether, in what sense, and to what extent, the different sciences are unified or dis-unified. It seems that the right methodology in order to approach these questions is by first testing whether properties at one level are nothing but properties at a lower level. Naturalistic philosophers can be said to be in the business of testing whether representational properties are nothing but biological properties.

The argument from blurring

Before I turn to examine the second argument, which will be the focus of the rest of the discussion, let us briefly consider Burge's third argument. Burge claims that there is a distinction between purely sensory systems and representational systems that has to be

honoured. Then he holds that deflationists blur that distinction. Now, a number of questions arise concerning the two claims. Let's begin with the first, i.e., the claim that there is a distinction between sensory and representational systems: Burge holds that the distinction is intuitive *and* that it is a distinction at the level of *natural kinds*. However, it is possible to grant that there is an intuitive distinction, but that it is not intuitive that the distinction is a distinction of kinds. The intuition can be honoured even if it is drawn at the levels of degree, of complexity, or whatever. And it is also possible to grant that the distinction is a distinction of kinds, but, even so, to defend that the kind *representation* is a biological kind, that is, that it is explainable in terms of the notion of biological function.

There is indeed a great difference between sensory detection and perceptual representation, especially if Burge's remarkable insights about the notion of representation are conceded. The difference between purely sensory states and representational states is that representational states are the output of computations – formation laws- that aim to "reach" distal causes. Representation is linked to objectification, i.e., the ability to reach the mind-independent, physical world. It is customary in cognitive science to speak about representations in another sense, in such a way that a system is considered representational if it carries out complex computations. The idea here is that computations require units that encode information and that it is right to call these units representations. Burge's point is that the information encoded by these units is not necessarily representational: it may be purely sensory information. Computational complexity is not sufficient for representation. Representation requires objectification.

In turn, objectification is linked to other two special and distinctive properties of representational states: that they have accuracy or veridicality conditions and that they exhibit perceptual constancies. Purely sensory states, according to Burge, do not really have accuracy conditions. They do not have accuracy conditions because the only thing they do is detect, or react to, proximal stimuli, whereas accuracy conditions involve a relation of correspondence between states and the world of distal causes. Ascribing accuracy conditions to a merely sensory state is, in fact, an abuse. Representational states, on the other hand, do have accuracy conditions, and accuracy conditions are linked to perceptual/representational constancies. Perceptual constancies are capacities to perceive a given environmental property under many types of stimulation. For

instance, colour constancy involves a subject's continuing to see something as blue even if the intensity of the light changes. Sensory systems do not exhibit perceptual constancies, even if they can mimic them, i.e., it may look as if they do exhibit them. Again, ascribing constancies to merely sensory systems is an abuse.

However, the claim that representational states, thus characterized, form a natural, irreducible, kind requires further argumentation. The difference between merely sensory systems and representational systems is, in essence, that representational systems operate according to formation laws, a kind of computation that is absent in merely sensory systems. Formation laws take as input sensory stimuli and give as output information about the likely distal causes which produced the stimuli. A system which implements formation laws has accuracy conditions, exhibits perceptual constancies, etc. Now, why should we think that the states that result from the computation of a sensory system which implements formation laws *are not* biological functions? The point about the "big difference" reveals that representational states and merely sensory states cannot be dealt with in the same way. But it does not show that, if sensory states are biological functions –as Burge concedes-, then representational states are not biological functions.

The adaptability of accurate representations: the argument from irreducibility

Burge's second argument against deflationism is that the function of representing veridically or accurately is not a biological function because it cannot be accounted for in terms of adaptability. Burge adopts the etiological selected-effects account of functions, according to which a given entity is functional if it "is there" because its ancestors produced effects that were fitness-enhancing. He then claims that representations do have functions: representations typically produce effects –like triggering fleeing behaviour- that are beneficial for the individual, and it is these effects which were the object of selection. So, representations have the (biological) function of producing certain effects. However, this generic, forward-looking, function that representations have is not a representational function, because the representational function consists in representing accurately. Thus, if it is conceded that biological functions are selected effects, given that the representational function is not, *per se*,

related to effects, we have to conclude that the representational function is not biological. Putting this in a more explicit form, the argument claims:

(1) Doing *F* is a biological function of *R* if and only if *R*'s doing *F* is a selected effect;
(2) *R*'s representing accurately is not a selected effect, whereas *R*'s causing *p* (e.g., fleeing behaviour) is a selected effect;

———

(3) *R*'s representing accurately is not a biological function.

There is usually a certain correlation between the representational function's good performance and adaptability: if the representational state represents accurately, the agent will display successful, adaptive, behaviours, *ceteris paribus*. However, this only holds *ceteris paribus*: it is possible that the representational state malfunctions as such, and yet that its effects are adaptive. We can easily think of an animal which escapes every time a sound is produced in its vicinity: the animal is wrongly representing a danger, but its behaviour may turn out to be correct from the point of view of its survival, reproduction rate, etc. This shows that representational states are not recruited by how accurately they represent, i.e., by how well they perform their representational function, but by the effects they have.

I think that this argument ultimately fails, but that it also reveals something very interesting about the particularity of some biological functions. However, let me first explain where I think it does not fail. It is easy to be misled by Burge's use of the example of the escaping animal. Burge uses this example to prove his point about the separability of the representational function and the representational state's forward-looking function. The example shows that these two functions can go in different, opposite, directions: one functioning badly and the other correctly. And the key is that natural selection does not bother about the representational function: it only "looks" at the forward-looking function. However, the example is controversial, and it may confuse rather than clarify. Burge tells us that the animal is misrepresenting danger. Now, why do we have to believe that the content of its representation is danger? It seems that we can just as well maintain that the animal represents *sound in my vicinity*. And if that is the case, it may be possible to hold that the good or bad functioning of a representation entails increases or decreases in adaptability.

Burge thinks that the content of the representation is *danger* because, in several places thorough the book, he tells us that perceptual systems represent attributes which are linked to the individual's good performance in its environment. Thus, he insists that the contents of perception can include attributes such as prey, danger, edible thing, etc. Burge even argues that ethological and zoological facts, i.e. facts discovered by the sciences which aim to tell us what biological function representational states fulfil, play a constitutive role in the determination of representational contents. This claim about constituency is, as far as I can see, unmotivated: it is difficult to see what role ethological and zoological facts have, apart from the epistemic role of providing some hints about what formation laws a certain species may implement (roughly, the ones necessary to cope with environmental demands). The contents of perceptions are the outputs of formation laws. An animal will probably be able to represent entities of the world that serve its goal of surviving, reproducing, etc., but whether it does or does not actually represent some particular entities is not a question that ethology or zoology can answer: what entities an animal represents depends on what computations its perceptual systems realize. However, Burge gives no example of formation laws whose output is, e.g., prey or danger. This means that we can suspect that there are no perceptual systems that represent prey or danger as such.

However, even if Burge does not succeed in motivating his assumption that the escaping animal is representing danger, it seems that the response that the animal may be representing *sound in my vicinity* will not succeed either. First of all: An animal that escapes whenever there is some sound in its vicinity, will also escape whenever it senses a quick shadow crossing, whenever the distribution of colours in front of it suddenly changes, etc. There are lots of events that can make it escape and seek refuge. So it makes sense to think that, even if its perceptual systems do not represent an attribute such as danger, the animal is representing, as the effect of cross-modal integration, some higher-order attribute, such as *sudden changes in the environment*. It is also possible to hold that the animal only represents *sounds around*, *shadows around*, etc., and that all these representations cause the same kind of behaviour. The drawback is that by going this latter way we seem to renounce to explain what is common to the animal's behaviours. Things get worse if we think that the animal does not escape upon hearing just any old sound, but only upon hearing certain specific sounds which may be very different one from the other. We would have to have an enormous list of distal

stimuli that generate the same kind of behaviour. Anyway, whether we go for this list or for the higher-level attribute, it seems that Burge's point stands: it is always possible that it is adaptive for the animal to generate false positives concerning those specific stimuli or that higher-level attribute. That is, the representational function may be failing while the behaviours it brings about are adaptive.

So, I think Burge's argument does not err in its claim that there may be failures in the representational function that do not result in failures in adaptability. I think that, in effect, the two functions are separable. And Burge's point, in any case, is that we can individuate the representational function as such without mentioning its effects. This makes the representational function peculiar, and distinguishes it from other "correlating" functions. For instance, the chameleon skin has the function of correlating with the colour of the surface it is on. This correlating function, however, is the direct object of natural selection: the correlation itself is an effect which has been selected, as the chameleon's skin's properties have been selected because they correlated with colours of surfaces. In this case, we cannot explain the correlating function without mentioning its impact on adaptability –it is a function because it is a selected effect. In contrast, representing accurately is not a selected effect: the effects which have been selected are the effects brought about by the states which have the representational function. This is the key point of the argument: representational states represent, on the one hand, and put the body in motion, on the other. Natural selection only tracks the impact of the representational state on action.

So, the problem of the argument lies elsewhere. The problem is that the fact that the two functions are separable does not mean that the representational function is not biological. (Consequently: the fact that representational failures do not align with, and are separable from, behavioural failures does not imply that representational failures are not failures in a biological function). Let me explain:

Consider Dretske's (1988) theory. According to Dretske, representational states have the following origin: we have a structure *C* which becomes activated in the presence of *F*. At that stage, *C* merely carries the information of, or indicates, *F*. As it happens, *C* gets linked to some motor response *M*, e.g., to fleeing behaviour, which turns out to be adaptive. As a result of this, *C* is recruited by some mechanism of selection, say, by natural selection. The interesting thing is that the recruitment process generates a double

functionality: *C* has the function of causing *M*, but it also acquires the function of indicating *F*. That is, there are two (manifested) dispositions that are converted into functions at the same time: a forward-looking disposition (*C*'s causing *M*) and a backward-looking disposition (*C*'s being caused by *F*). Note again that before *C* was recruited to cause *M*, *C* did not have the function of indicating *F*: it merely carried the information that *F*. *C* is recruited to indicate *F* (and so *C* represents *F*) only when the effects of *C*'s indicating *F* are selected.

Applying the theory to the case of the escaping animal above, we obtain the following picture: an animal has some internal structure *C* which co-varies with dangers to an interesting degree. The activation of this structure generates at some point fleeing behaviour. As it is adaptive to escape whenever there is some sudden change around -let's leave "danger" aside-, *C* gets selected. It is selected for its contribution to fitness, but selection does two things at the same time: it converts the effects of *C* into a function (that is, *C*, from now on, in the animal and in its progeny, has the function of making the animal escape), and it converts the indicative properties of *C* into another function (that is, from now on, in the animal and in its progeny, *C* has the function of indicating the presence of sudden changes). This should not sound strange: a natural detector acquires the function of indicating what it indicates when we put it to some use, that is, when it also acquires the function of doing something. For instance, a temperature detector acquires the function of indicating changes in temperature when it is put to use as forming part of a thermostat.

Now, let's grant Dretske's account for the time being and with it the idea that a single natural selection process may have as a result a double functionality –forward-looking and backward-looking. Two consequences seem to follow: first, that both functions are equally biological. Second, that it is possible for the first function to work well while the second one works badly and vice versa. A representational state has two functions: to cause *M* and to indicate *F*. It may fail to cause *M* while it correctly indicates *F*, or it may fail to indicate *F* while it correctly causes *M*. This is what seems to be going on in Burge example: the state causes what it has to cause –fleeing behaviour-, even though it does not co-vary with what it has to co-vary with –e.g., sudden changes-.

A possible response to this way of treating Burge's case is to hold that the animal could well be incorrectly representing sudden changes from the very beginning, but that there

was selection for the effects of its representational state anyway. That is, it seems as though the Dretskean picture requires that there was a moment of "purity" were the animal reacted only to sudden changes, i.e., where the representation was always correct. Malfunctions came only later, after the representation was recruited. However, this is not correct. The Dretskean story requires that there was a satisfactory degree of co-variation between sudden changes and the structure or state that is to be turned into a representation, and that the state is not regularly triggered by other kind of events. At that point, it is wrong to describe the activations of that state as correct or incorrect: the state has no function as yet. That description is only available *ex post facto*. Still, it can be claimed that it is a natural detector of sudden changes, because it reliably co-varies only with sudden changes in the environment and does not regularly co-vary with other, different, events –even if it is occasionally triggered by other events. So, there may well be what we would call, in a later stage, malfunctions of the detector. When the detector is selected for the effects it brings about, the detector acquires the representational function, and those reactions can be properly called malfunctions. What is true is that, from that moment on, it begins to be possible that the representation as such functions worse and worse while its effects keep on being adaptive.

If this sort of response does not sound convincing, it may be advisable to turn our eyes to the probably most popular teleosemantic approach, Millikan's. For adopting Millikan's standpoint may make things easier at this point. Under this view, the animal's state may co-vary with a number of states of affairs. But when it acquires a function, it acquires the function of co-varying only with those states of affairs that its consumers need in order to perform their function. That is, the state may be "malfunctioning" from the very beginning. When natural selection enters the picture, it also selects what contents the state has to represent. In our case, natural selection determines that the content be such that it is connected to what fleeing behaviour is for.

Now, however we treat the particularities of the case, the claim seems to stand: the representational function may perfectly be a biological function, that is, a function that arises from the action of natural selection. Natural selection "looks at" effects, that is true, but by selecting states that look forward and also backward, it converts into functions not only the forward looking aspect of the state but also its backward looking aspect. Burge's argument reads: "There is, however, a *root* mismatch between representational error and failure of biological function... Biological functions are

functions that ultimately have to do with contributing to fitness for evolutionary success... Explanations that appeal to biological function are explanations of the practical (fitness) value of a trait or system. But accuracy is not *in itself* a practical value" (p.301, Burge's italics). My claim is: even if accuracy is not in itself a practical value, this does not mean that representational error is not a failure in a biological function.

Let me close with two important points that follow from the position here adopted. First of all: note that it has been uncovered that there is a special kind of functions which are not selected effects, but are biological functions nonetheless, as they are the result of a process of natural selection. This may take us to reconsider the selected effects account of functions. Next: note also that there is nothing special to representations if this approach is taken, for all that has been said applies also to merely sensory systems –in fact, I have been making no distinction between merely sensory states and representational states thorough. There can be the same kind of mismatches between the backward-looking functions and the forward-looking functions of sensory systems as there are between the backward-looking functions and the forward-looking functions of representational systems. That means that if the representational function is not a biological function because it has no direct effect on adaptability, the indicative function is not biological either; the indicative function is, as such, separable from the forward-looking function. This is a problem for Burge because he tells us that merely sensory systems can be fully explained in terms of causal co-variation and biological function, that is, in naturalistic terms.

So, all things considered, I think that Burge has failed to show that the representational function has to be a primitive. His arguments, first, that the naturalistic enterprise is unmotivated; second, that the representational function is not a biological function because nature only selects for effects; and third, that deflationists necessarily blur a distinction that has to be respected, have been found wanting. If we take the three arguments together, and allow one to intrude the next, we get an overall rather convincing picture. However, taken in isolation one by one, things are much less convincing. Or so I have tried to show.

Acknowledgements

The ideas of this paper developed as a result of a reading group on *Origins* held at the University of the Basque Country. I thank Alvaro Moreno, Arantza Etxeberria, Alba Amilburu and Agustín Arrieta for helpful discussions. I am particularly grateful to Ignacio Vicario and Marc Artiga, who helped me to improve the paper significantly. Thanks also to the editors of *Thought: A Journal of Philosophy*. Research leading to this work has been supported by the Spanish
Government, research projects FFI2010-15717 and FFI2011-30074-C02-02.

References

Burge, T. (1993) "Mind-Body Causation and Explanatory Practice", in Heil, J. and Mele, A., eds., *Mental Causation*. Oxford: Clarendon Press.

Burge, T. (2010) *Origins of Objectivity*. Oxford University Press.

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*. MIT Press.