

WILLEM A. DEVRIES

EXPERIENCE AND THE SWAMP CREATURE

(Received in revised form 24 August 1994)

Exactly what is the relation between the state of one's body (regardless of the relations it may bear to things external to it) and one's mental life? Individualism - the doctrine that for every thinking individual there is a non-intentional description of it that mentions nothing external to that individual, whether physical, social, or psychological, and is sufficiently detailed to determine fully its intentional states - draws a very tight relation between our narrow physical states and our mental states. But individualism has come under sharp criticism from philosophers like Tyler Burge (1979, 1982) and Lynne Rudder Baker (1987), and it now seems unlikely that our narrow physical states determine all our intentional states. There have been several attempts to limit the significance of anti-individualist critiques by carving off some important region or aspect of our mental lives for which individualism still holds. Burge, Baker, and others have answered attempts to demonstrate the existence of "narrow" intentional states. In this paper I seek to cut off another attempt to save individualism - the notion that at least our sensory experience is individualistic.

While the primary topic of concern will be individualism, there are other matters of philosophical moment tied to this issue. In particular, the viability of the Cartesian (or neoCartesian) view of the mind as an in-principle autonomous realm/entity distinct from the material or natural realm or entities seems to be tied to our ability to cordon off some area in which our access to or knowledge of the mental is distinctly different from and even independent of our access to or our knowledge of natural or material entities, properties, or relations. One goal of this paper will be to argue that the naturalistic anti-individualist can have a perfectly reasonable theory of sensations, which are standardly taken to pose anti-Cartesianism its greatest difficulty.

Philosophical Studies 82: 55-80, 1996.

© 1996 KluwerAcademic Publishers. Printed in the Netherlands.

Arguments about individualism usually make use of thought experiments involving thinking subjects that are molecular duplicates (or quantum-particle duplicates) of each other. In these thought experiments the existence of such duplicates is usually ascribed to a grand cosmic accident, but there are still different ways we can imagine such an accident occurring, and it makes a difference how one fills in the story. We could imagine a set of fairly normal causal processes in the world that happen to generate in the normal way a being who just happens to be a molecular duplicate of, well, let's call him A1. This being would have parents, just like the rest of us, he would have matured and developed just like the rest of us, and the cosmic accident would be simply that somehow this being ended up a molecular duplicate of A1. This whole process would have occurred *naturally*, not just in the sense that it violated no law of nature, but in the Aristotelian sense that courses of events unfolded in their *natural* and customary way. These kinds of cases are familiar to everyone who traffics in anti-individualist arguments, but they are *not* the cases I want to consider today.

The other way to imagine a duplicate is even more freakish: somehow a collection of molecules or quantum particles, in the midst of otherwise random motions, gathers into a collection that turns out to be an exact duplicate of A1. No Aristotelian natural processes here at all, just one huge throw of the dice and there he stands, Sal.

What should we say about the mental life of such an unnatural duplicate? Confronted with such a scenario, many people have apparently strong intuitions that *of course* Sal has a rich subjective experience, just as A1 does. The differences between A1 and Sal are many, admittedly. They differ certainly in their spatio-temporal location and history, the veridicality of their apparent memories, perhaps even in the very intentionality of their subjectivity. But, the intuitions are, surely both A1 and Sal are similar in being experiencers, for whether something has an experience is an occurrent and individualistic fact if anything is. I want to argue here that we have good reason to say that in the first few moments of its existence Sal the swamp creature has no experiences.

Burge (1986) does briefly consider what to say about these grand quantum accidents in footnote 7 in "Cartesian Error and the Objectivity of Perception." After pointing out that anti-individualist arguments

utilize in a crucial fashion the objectivity of thought and perception, Burge remarks that,

Perhaps one should see in this light the fanciful examples of beings that are bodily identical to us over a period of time, but that are extended quantum accidents with no regular relations to their physical environment. There is perhaps enough in the mental events of such beings to count them thinkers. (Actually, this seems to me problematic: there are problems about dispositions; but let them pass.) Whatever thoughts they entertain have no determinate *objective* reference. Their mental goings-on and their physical movements are compatible with successful adaptation to and regular causation by any one of an infinity of possible environments. Such intentionality as their phenomenological states have should perhaps be seen as making reference to qualitative phenomenological types, not to objects that are in principle independent of the individual's thought and perception. For such a being, individualism is perhaps true. But its truth would be bought at the price of interpreting the thinker as a solipsist unaware. I leave the ultimate coherence of such a description an open question. (pp. 135-36)

My attempt here will be to close the question: The weight of reasons legislates against attributing to the nascent swamp accident any mental states, conceptual, perceptual, or even sensuous.¹ The case I am considering may not be exactly the case Burge had in mind, for his note seems to imply that his accidental being never engages the world in the kind of causal interactions that are typically human, whereas I assume that when the swamp creature emerges from the ooze this I or he or it (the thing) which thinks(?) would proceed to behave just like its duplicate would in similar circumstances.

There are three answers to the question "Does the swamp creature have mental states?" current in the literature. 1) The swamp creature has mental states; 2) the swamp creature has no mental states; and 3) there simply is no fact of matter. While I have some sympathies with the latter answer, I think it is a position to occupy only as a last resort, but not because I have strong metaphysical prejudices against the indeterminacy of important categories. The possession of mentality is a very complex affair, and the easy claim that there is "no fact of the matter" can deter one from uncovering some of the structure underpinning our attributions of mentality. There is at least a regulative injunction to push the yea and nay positions to the hilt, and to fall back upon a factless metaphysics only when neither yea nor nay survives.

I will begin by looking at Ned Block's interpretation of the swamp creature case. His story occurs as part of an attempt to distinguish intelligence and intentionality, and while there is clearly an important difference, Block suggests - though he shies away from endorsing - a complete independence between the two. Block cites naturalistic indicator-based theories of intentionality - theories like Dretske's that ground intentionality in a natural, causal relation between the representing and the represented - as giving possible grounds for thinking intentionality possible without intelligence, but avows skepticism that any such theory could succeed.²

The swamp creature appears when Block turns to argue that intelligence without intentionality is possible. Such a swamp creature, Block claims, would arguably be intelligent, for "it will have all the capacities (behavioral capacities) that [A1 has], and they will be produced by the same sort of physiological processes as occur in IA1" [Block (1990), 263]. And since "what makes a system intelligent is what it can do," the swamp creature is intelligent.

But it has no intentional states, argues Block, and for the right reasons. Though it may produce an utterance acoustically identical with A1's utterance "Perot will run in '96 if the deficit is still climbing," its utterance is not semantically type-identical with A1's, nor are the internal states involved in producing its utterance semantically the same as A1's. The creature has had no causal contact with Perot, nor has it had any opportunity to acquire concepts of deficits, elections, or 1996; how could we tie its utterance or its internal states to such things? ³ Well, we can't, as Block quite correctly notes. So the swamp creature is not thinking anything when it first emerges from the ooze, even though what is going on inside it is, on some description, identical with what would be going on inside A1, who most manifestly would be thinking something. This part of Block's argument is in perfect order, but we should question Block's reasons for calling the swamp creature intelligent.

According to Block, "what makes a system intelligent is what it can do." What can the swamp creature do? According to Block, anything A1 can do. So if A1 is intelligent (and we're not doubting that), it's

intelligent. But wait; Block devoted a large portion of the earlier part of his paper to arguing against a behaviorist theory of intelligence. He argued that it might be in theory possible to build a machine that can do everything I can do, but that does these things by brute force methods. Such a machine would not be intelligent, Block claims. His later claim that "what makes a system intelligent is what it can do," however, is susceptible to a behaviorist interpretation. If Block is to be consistent with his earlier argument, he must mean that what makes a system intelligent is what it can do in the richer sense that includes under what the system "does" the process generating the system's overt behavior as well as its result. So if the swamp creature is intelligent because it can do everything A1 can do, it must also be utilizing the same processes as A1. But which of the many processes going on in A1 are the relevant ones here? My guess is that most people would answer that A1 is intelligent because A1 can and does think. (And this seems the right answer to me as well.)⁴ A1 is intelligent because he knows a good deal about the world and can utilize that knowledge in figuring out how to achieve his goals, including the acquisition of still more knowledge to help him cope with the world in the future. The relevant processes accounting for intelligence in our case are in fact processes involving intentional states. So if A1 is intelligent because of what A1 can do, and the swamp creature is supposed to be intelligent because it can do everything A1 can, it must also utilize intentionally described processes. This train of reasoning, however, leaves Block with a dilemma: either the swamp creature is intelligent like us and has intentionality, or it has no intentionality, and therefore cannot do everything we can do, and its intelligence is still very much in doubt.

Block never even mentions the fact that our behavioral capacities are produced, at least in part, by processes involving intentional states. He simply asserts that the swamp creature's capacities are the same as A1's and "will be produced by the same sort of physiological processes as occur in [A1]." Processes involving intentional states as such are simply left out of account here, and it is just assumed that the relevant processes are physiological processes. Intelligence is not a merely output-sensitive notion; as Block's criticisms of behaviorism make clear, it is process-sensitive as well. Standardly, I claim, the relevant processes are

intentionally, semantically, or informationally characterized and thus subject to all the anti-individualist arguments the literature has made so familiar. Block recognizes that the swamp creature is non-intentional because intentionality is not individualistic; Block fails to see, however, that intelligence is not individualistic either, and therefore holds on the idea that the swamp creature is intelligent.

It should, of course, come as no surprise that intelligence is not individualistic. Many people have pointed out the context-relativity of intelligence. But we must be very careful here, for Block claims that there is a context-free notion of intelligence, and that I have committed a fallacy of equivocation in my criticism. 'Intelligent' sometimes means *smart* ("in a sense that includes adaptation to context" (Block (1992), p. 2), and the *smart* thing to do is always context-dependent. But there is another use of 'intelligent' that does not imply *smart*, is the sense that Block intended, and is, Block claims, context-free. "Intelligence' as I used it applies to a system in virtue of the information processing that goes on *inside the* system, independently of context" [Block (1992), p. 2]. What Block has in mind is that the correct way to explain the intelligence of an agent (or machine) is to perform a *functional analysis*, showing how the intelligence of the whole arises from a complex orchestration of (levels of) less intelligent agents, analyzing each new layer in turn until the analysis bottoms out in perfectly unintelligent, mechanical processors. "The methodology implicit in this type of functional explanation is individualistic," Block claims [Block (1992), p. 3].

Granted, there is an important distinction between the general quality of intelligence and smarts. But I am far from willing to grant that intelligence is capable of either context-free or individualistic definition. If intelligence is context-free in this sense, then it is capable of a thoroughly formal definition, and anything and everything that satisfies that formal definition is intelligent. But then there are far too many intelligent entities in the world. The atoms in a bucket of water, the particles in a galactic nebula, and who knows what other weird conglomerations of matter may well instantiate the context-free formal structure(s) that supposedly characterizes intelligence. An abstract, context-free conception

of intelligence falls to the same kinds of counterexamples as abstract (or machine) functionalist conceptions of the mental.⁵

We explain intelligence by analysis in terms of information processing functions, and information is not an individualistic, context-free notion. Functional analysis may eventually unveil non-informationally characterized mechanisms that implement the information processing, but it needs more argument to show that the method of functional explanation is individualistic because one sooner or later reaches an individualistically characterizable set of mechanisms. Indeed, I shall argue in section II that Block is quite wrong about functional analysis -as applied in the natural sciences, it is an essentially non-individualistic methodology.⁶

Thus, I don't think that distinguishing smarts from intelligence will protect Block against the reminder that intelligence is a context-dependent, nonindividualistic notion. The thrust of Block's arguments was to save for the swamp creature at least one property with strong mentalistic associations - intelligence. My arguments here are intended to show that they don't work, and we have no good reason to call the swamp creature intelligent. This is still far from establishing that there are no good reasons for attributing any kind of mentality to the swamp creature, but if we push further along the line I have begun here, we can begin to pick more holes in the idea that the apparent similarity between A1 and the swamp creature must mean that the swamp creature is like A1 mentally as well.

2

The next step in the argument is to question Block's assumption that the swamp creature's capacities are "produced by the same sort of physiological processes as occur in you." It seems unproblematic at first glance, but I think it is simply false. My argument is fairly straightforward and also fairly basic. The sorts of physiological processes that go on in me are functionally individuated. There are the digestive processes, the reproductive processes, the sensory-motor processes, etc. Of

course, these are very high-level descriptions that subsume any number of specific sub-processes that may in turn have still further subprocesses as parts. And it is certainly true that there is a point where specific processes are no longer described or individuated by their functional role in the organism's design, but by the chemical event they instantiate. The chemical interactions in the swamp creature do not instantiate the same physiological processes as their counterparts in AI, however, for functional individuation, at least of natural objects and processes, is rarely individualistic itself, but involves aetiology essentially. Assertions about the identity of processes are, in fact, relativized to a level of description, and two chemically identical processes may not be physiologically identical, just as they may not be psychologically identical. So one of my aims in this section is to argue that Block, although he has abandoned mental-content individualism, is still plagued with physiological individualism. My other aim here is to argue that functional individuation itself, at least as used in the biological sciences, is inherently non-individualistic. In the third section of this paper I will then draw some further conclusions for the philosophy of mind from this material.

Why is functional individuation rarely individualistic? Individuating an object or event in terms of its function is only as good as our ability to individuate the function itself. Functions are usually specified (and non-accidentally so) in terms of some goal. In fact, there is usually a hierarchy of goals involved. Our organic functions may generally serve the goal of enabling our genes to survive, but this requires nutrition, reproduction, resistance to enemies and Lord knows what else. And each of these ends, of course, is accomplished via further subgoals. The overall goal and a good many of the higher-level subgoals usually involve the environment in some way (I will discuss the relevant ways shortly), and thus the function itself is nonindividualistically specified. Furthermore, the environmental involvement is inherited even by those subprocesses that can apparently be specified independently of the environment. The mechanisms of internal temperature control in mammals *are temperature* control mechanisms and not electrolyte control mechanisms because it is as temperature control that they fit into the natural goal structure of the organism. Only when we have gotten down to some

level where we stop talking about functions and start talking about strict, law-governed behavior do we leave implicit environmental ties behind. But this is usually accomplished in quite a different vocabulary, even a different discipline (not that that means much).

There are at least two ways in which a goal can be tied to the environment. First, it can, of course, simply be a goal that is "out there." Nutrition, for example, is often a goal of an organism, and it is usually found in the environment. If the goal of animal behavior were some kind of internal equilibrium state (as some have theorized - the biological equivalent of Fodorian Methodological Solipsism, I guess), the animal organism and its processes might be individualistically individuated, but if the goal of animal behavior in general is some kind of survival, and particularly if the main goal is not personal survival but species or genotype survival, then the goal itself can be described only by referring to states beyond the individual and thus importing non-individualistic considerations. Now, there is surely no *requirement* that all function-characterizing goals be "out there" - many biological goals are internal. And some goals standardly characterized by reference to external states may be redefinable individualistically. But I think it is very probable that *some* goals of natural organisms must be irreducibly environmentally tied. My guess is that environmentally-tied goals are as necessary to a functioning organism as the presence of some *de re* attitudes is to a cognitive being.

The second, and I think more systematically important, way in which a goal can involve the environment is by way of goal-acquisition. How could the object acquire that goal? Just as how we acquire the concepts we employ is a major constraint on the interpretation of our mental states, how we acquire the goals we have is a constraint on the functional diagnosis of our activity. The goal structure that we attribute to artifacts are built into the artifact by us. The goal structures of natural organisms, on the other hand, cannot be acquired by explicit design. The great advance of Darwinian biology is that it provides a strategy for giving a natural explanation of the goal structures of organisms. Natural selection gives us a general understanding of how goal structures could emerge from the primal ooze, maintain themselves, and develop towards increasing complexity over time and generations. We have a

set of strategies for explaining how specific goals might be acquired, just as we have a set of strategies for explaining how specific concepts might be acquired. Any intelligible analysis of functional structure must respect those strategies just as intelligible interpretation of mental states must respect the strategies by which concept-acquisition is explainable.

The upshot of this argument is that any proper and complete functional decomposition of an object must be able to explain for each sub-process how it acquired the goal it subserves. Now obviously, most of our functional analyses of organic functions do not go this far, and most probably could not easily develop an explicit explanation along these lines. But we do know the general strategy for such explanations, and we therefore have reason to reject any functional decomposition of an organism that seems incompatible with such an explanatory strategy. And that's where we find the swamp creature. We can't explain why it has such an elaborate system capable of ingesting decomposable organic material and passing the energy obtained on to its other parts to sustain the on-going chemical reactions; we can't explain why it has a system capable of combining with similar beings to produce new beings that derive their structures from those of their progenitors. The mechanisms by which things acquire goals, either via natural selection or via some intentional design and construction process, have been ruled out from the start.⁷ The creature is *per hypothesis* a chance collocation of atoms. Thus I am unwilling to grant that the swamp creature's behavior is produced by the same physiological processes that produce my behavior. For physiological processes are individuated in part by reference to their goals, and it is a reasonable constraint on goal-attribution that goal-acquisition be explainable. Or, more precisely, it is a reasonable constraint on goal-attribution that the acquisition of that goal not conflict with recognized mechanisms of such acquisition. There is an implicit promissory note in the scientific analysis of a functionally individuated process that it will fit into a bigger and broader causal story.

What makes cases such as these so confusing is that we often tend to focus on the forward-looking aspects of the functional, and the swamp creature seems to duplicate all our forward-looking properties. The swamp creature will apparently behave in just the way A1 would if he were in the creature's oozy situation, and it would surely be the best

strategy for dealing with the swamp creature to treat it *as if* it were just like A1 (or at least as if it were a normal human being). In the light of these pragmatic considerations, the peculiar past history of the swamp creature seems inconsequential. But to ignore it altogether, to say it makes no difference is to ignore part of the rich structure of our concept of functional teleology. Perhaps most important, it is to abandon an important part of naturalism, for it lifts functionality out of the causal network as itself immune to further natural explanation.

There is an obvious objection to the position taken here: "Slice open the swamp creature and what do you see? A heart and lungs, no? But if you're right, the swamp creature does *not* have a heart and lungs, for these are functionally characterized physiological items, and the swamp creature doesn't have the same physiology as us. Haven't you reduced your position to absurdity?" Granted, we would be sorely tempted to described its internal structures as heart, lungs, etc., but this temptation can be easily overcome with soberer reflection.

First of all, we certainly can say that the structures inside the swamp creature function *like* a heart and lungs; we can even go a bit further and admit they function *as* its heart and lungs. But does that beating thing inside the swamp creature's chest really *have the function* of circulating its blood (admitting, only for the moment, that it has blood)? How could we explain how it got and why it has such a function? We can't. The biologist would say that these structures are at best *analogous*, but not *homologous*,⁸ and she could on that basis deny that the swamp creature has a heart and lungs *like* ours in the biologically interesting sense. But the biologist should not stop there: the swamp creature's organs, structures, and processes are homologs of *no* organic structures or processes; its entire being is merely analogous to biological organisms. Classifying it and its structures with their biological analogs and marking no distinction abrogates the fundamental principles we have discovered to govern natural biological kinds. We really should not say that the swamp creature has a heart and lungs, unless we footnote it to explain that we mean here only that it has heart- and lung-analogs?⁹

There is a lemma to my argument that I alluded to earlier but is now worth expanding. There are at least two different notions of functionalism available in the literature. One, I shall call it *abstract* functionalism,

holds that a functional role is defined solely in terms of a pattern of relations, where not *all* the relations a role-filler may have will count. Which relations are important will be given some more or less abstract specification in a theory of the relevant domain. The functional role can then be identified via a Ramsey sentence of that theory specifying a set of relations among inputs, outputs, and internal states,¹⁰ I call this *abstract functionalism* because the functional roles are defined in an abstract, context-free manner. Turing machine states are simple and standard examples of abstract functionally defined roles. The functional roles of objects, states, or events are determined by setting them in one-to-one correspondence with the states specified in the Ramsey sentence. Abstract functionalism, of course, can be as individualistic as you please in its modes of individuating functional roles.

Though I believe our physiological processes (and our mental states) are functionally defined, I do not believe they are abstractly functionally defined. For one thing, abstract functionalism is simply too liberal.¹¹ Arbitrarily constructible isomorphic structures are more numerous than the blades of grass, the drops of rain, or probably the natural numbers. There would be organs and minds everywhere if abstract functionalism told the whole story.

More importantly, though, abstract functionalism does not fit in well with the methods and goals of an explanatory natural science. It may be a fascinating revelation that a certain set of states of the world matches up one-to-one with the states of some abstractly specified system, but that need not be anything more than a colossal accident. If the functional characterizations of the relevant states are themselves to have explanatory value (and if they don't, what role are they serving?), the functional relations among the states must be nomological; they must support counterfactuals. In individual cases it might be plausible that the nomological force of the functional characterization is simply passed upwards from particular law-like relations among the realizing entities. But then there is no particular explanatory job performed by the functional characterization. The functional vocabulary has explanatory force only when it is independent of the realization in some significant manner. Abstract functionalism gives us this medium independence, but at too great a price, for it cuts the functional characterization off from its context and

makes it impossible to explain why the functional characterization holds in the case at hand. Explanatory functionalism requires functions that are interestingly medium- or realization-independent while sufficiently tied to the rest of their causal situation to fit into the other stories we tell about the world. The *unity* of science is probably too much to demand, but the *continuity* of the sciences seems essential. Abstract functionalism cannot provide that, but the teleological functionalism espoused in this paper does.

There are probably a number of ways in which goal-attribution and content-attribution are similar, though I will not try to spell out details here. The satisfaction-conditions of the goal attributed probably play a role similar to that played by the truth-conditions of content-attributions. Thus, we should also expect that the ways in which goals in functional explanations involve the environment are very similar to the ways in which contents in intentional attributions involve the environment. This is hardly surprising, since many intentional attributions are themselves parts of functional-style explanations of behavior. The parallels are impressive, but one shouldn't get carried away with them. Just as there seems a need to posit some *de re* mental contents with direct relations to the environment and which supply the elements out of which *de dicto* contents can be constructed, so there may be a need for certain "*de re*" goals that relate directly to the environment and on the foundation of which "*de dicto*" goals without explicit ties to the environment can be built. Possible example: dogs can have "*de dicto*" goals in their dreams of catching that rabbit, only because they have real *de re* goals of eating, etc. I don't think it is yet clear just what role in our belief-system *de re* beliefs play, but their presence seems unavoidable. The same goes for *de re* goals. Similarly, as I have argued, just as one of the major arguments for non-individualism in mental contents comes from considerations of concept acquisition, so the problem of goal-acquisition motivates non-individualism for goals.

My points in this section have been that abstract functionalism is not suited for use in any naturalistic explanatory enterprise, and thus that functional analyses and explanations contained in the empirical disciplines will never be context-free. This gives us reason to deny that the apparent similarities between A1 and the swamp creature force us

to give the same functional analysis to each of them. Just as functional similarity doesn't entail physical similarity, so physical similarity, it turns out, doesn't entail functional Similarity.

3

We have seen reasons now for denying the swamp creature intentionality, intelligence, and even functional similarity to us. But many people still resist the notion that there is really nothing mental going on in the swamp creature. It seems to them that the fact that the only explicit differences between the swamp creature and A1 are either external or historical makes them irrelevant to comparisons of their current mental states as such. At least, such people contend, there is some aspect of the mental independent of external and/or historical relations with respect to which we can compare A1 and the swamp creature. I shall call such people Autonomists, since their central thesis is the autonomy of the mental - its fixity and determinateness independently of its physical and historical contexts.

Autonomism comes in different strengths. A hard-core Cartesian dualist would claim the in-principle complete autonomy of the mental. Individualism is a far less stringent autonomism in that it merely limits the range of past, present, and future physical conditions that are essentially involved in the mental to those now internal to the body. Individualism seeks, as it were, to limit the dependence of the mental on the physical to the smallest physically respectable portion of the world. The arguments against individualism that have been launched all depend on the facts that the intentional content of many mental states essentially involves conditions outside the body and that without taking such involvement seriously we cannot individuate contents or the respective mental states. To hold on to the relative autonomy of the mental, the individualist needs to find some set of mental states that are proof against anti-individualist arguments and might plausibly be construed as fundamental to all the others. A set of mental states proof against the anti-individualist arguments would have to be systematically shorn of all content with reference beyond the body. No states that make

overt reference to external things are allowed; contents are restricted to the strictly internal: sensations, other intentional states, the concept of self and such things. These states, it is claimed, surely survive the anti-individualist critique. Because such states make such minimal claims, it is difficult to show that they are, at root, also world-involving in ways that are inconsistent with hard-core Autonomism.¹²

What I propose to do is aim at this smaller target by broadening my target field in general. And this is where the considerations advanced in the previous sections become relevant. In brief, I will argue that the content-poor states the Autonomist hopes to save from environmental contamination are nevertheless contaminated because of the *kinds* of mental states they are, not their contents. Roughly, our distinctions among the various kinds of mental states (sensations, beliefs, intentions, etc.) are functional distinctions, and I hope to show that the arguments that functional individuation is in general anti-individualistic also can be turned to show that in individuating a state as a sensation we have already made reference to environmental ties. Thus, paring down the content will not suffice to save either the internality of the mental or its autonomy, for the distinctions we make among the attitudes already involve the world.

I have to begin by discussing the typology of mental states. A fairly coarse, rough-and-ready set of distinctions enables us to trace the reasons for which Autonomists often focus on sensations as the most problematic states for my approach. I then argue that they are nonetheless infected with the general anti-individualism of the functional. This part of the paper can promise no more than a programmatic outline, I'm afraid, and leaves open a number of questions: How strong is this infection? How does it affect the ontology of sensations? What does it tell us about the role of sensation in our cognitive lives?

What kinds of mental states can we have? There are a number of classifications of mental state kinds available. (I'm not sure what to say about, e.g., moods, because I'm not sure just what they are or how they relate to our intentional states; does a mood have a content?) I will use a very simple classificatory scheme of contentful states to make my points. We may classify contentful states roughly into input-oriented perceptual states, action-oriented volitional states, and "purer," inter-

mediary intentional states like beliefs. This three-fold classification is quite rough, and the boundaries among these groupings vague and overlapping. A fear may be highly action-oriented as a direct and immediate cause of flight, or it may be fairly removed from action, perhaps an existential angst.

It seems fairly commonsensical to individuate perceptual and volitional states by reference to the external realities they are oriented towards (and this is, in fact, how we do it). But it has also seemed reasonable to believe that the pure intermediary states could be individuated on their own. Holding on to an individualism of the purely intentional, however, means that somewhere along the line from the world's effects upon us to our effects upon the world we can drive a significant enough wedge to isolate pure intentional states from their contextual surround. As I see it, much of the Cartesian, Autonomist tradition in both metaphysics and epistemology is engaged in precisely this task (and then trying to show that despite the wedge, the pure intentional states are not so isolated from the world as to justify skepticism).

The accomplishment of the recent anti-individualist arguments is to give us good reason to think that that project can't succeed. Pure intentional states are not autonomous or individualistically individuated. We are better off going with the intuition that perceptual and volitional states are properly individuated by reference to the external realities they are oriented towards, and to let that involvement of the external infect the apparently purely intentional as well. This approach is consonant with the thoroughgoing naturalism that Autonomism rejects, and locates the "purely" intentional states of an organism right where they belong: inextricably tied into the causal nexus of the world.¹³

There has been a kind of subliminal understanding among the Autonomists of the difficulty of isolating the purely intentional, evident in the widespread agreement that, if one is going to drive in a wedge that separates the mental realm from the physical environment, it needs to be done early, right at the "interface" between world and mind. The classical story is that the mind "interfaces" with the world through sensations, which are purely mental but not purely intentional. For the latter-day Autonomist who doesn't want the mental to be entirely autonomous from the physical, the "interface" between mind and world is to be

found at the interface between the body and the world, in the sensory organs. In fact, this whole notion that there is *an* "interface" between mind and world - even an apparently physicalistically respectable one - is yet one more episode in the ongoing Myth of the Given that Sellars (1956) attacked in his classic "Empiricism and the Philosophy of Mind." The Autonomist standardly tries to establish some basic or foundational level of the mental at the mind-world "interface," the content of which does *not* refer to anything external. Whether conceived of as sense-data, *ProtokoUsiitze*, non-comparative lookings or seemings, or "narrow contents," such states are supposed to be capable of remaining constant across contextual differences precisely because they are on the other side of the "interface."¹⁴

The naturalist, in contrast, strives to understand humans and other organisms as particularly interesting and complex natural objects. Perception, dispositional belief, and volition are all pieces of the functional articulation of the cognitive continuum connecting the world's effects on us with our effects on the world. Isolating a particular, apriori specifiable point of "interface" with the world holds no interest for the naturalist, however. As instrumentalities of an organism's survival in a complex world, cognitive structures are as functionally defined as physiological structures, but as intentionally or informationally characterized structures, there are peculiarities to them that do demand particular attention. How the informational structures of the organism "interface" with the world is a major problem for the naturalist, but it is an empirical problem to be solved by empirical methods, and there can be no apriori assumption that the "interface" remains constant in the face of cognitive change.

Between the traditional prejudices of western philosophy and the complexities of the mind's involvement with the world, sensory consciousness seems inevitably the last hope for the Autonomist and the most difficult test case for the anti-individualist. Is consciousness itself for something, merely a functional structure of our organism, tied like all functional structures to its context?

It is not news that we typically individuate sensations by reference to the properties of the distal objects that typically or commonly cause the sensations. Sensations of red, of heat, of texture are parasit-

ically individuated by reference to their typical or common causes, for instance.¹⁵ In this regard sensations should be very susceptible to twin-world type arguments. It is implausible that we typically individuate unicorn-thoughts by reference to relations with actual unicorns, but it is easy to believe that sensations of red are typically individuated by reference to relations with actual red things. Yet, when we construct such scenarios, something funny seems to happen, for they seem less convincing than the arguments aimed at intentional states. The phenomenal quality of a sensation seems intrinsic to it, and even if, in our language, we individuate the sensation by referring to its typical or common cause, we seem perfectly well able to imagine the phenomenal quality remaining constant in tremendously different circumstances. Thus, many people are willing to countenance twin-world stories in which the twin worlders' spectrum is inverted relative to ours, although their linguistic and behavioral reactions to environmental stimuli are just what ours would be. Beliefs and other relatively "pure" intentional states without any phenomenal quality have nothing to them that could hold constant through a change in relational properties. To the extent that perceptual and volitional states seem to afford something that could remain constant across significant contextual changes, it is because of their affective, sensational component.

The presence of a phenomenal quality in the case of a sensation state allows us to think that the relational properties that the state may bear to the context and that we normally rely on to individuate it serve merely as contingent *reference fixers*. This differs from the case of beliefs: with beliefs, there is nothing about them other than relation-based content ascription that we could hope would prove us access to them.

While beliefs seem necessarily caught up in the cognitive continuum between input and output - we wouldn't know how to think about them apart from their place in the continuum - sensations seem capable of an independent existence, because there is, apparently, a way to think about them apart from their place in that continuum- direct acquaintance with their phenomenal quality.¹⁶ There is a price to pay here, however. In so far as sensations have an intrinsic nature independent of their role in the cognitive continuum, it is unclear what that nature, their phenomenal quality, has to do with their role in cognitive activity. After all, sensations

are usually thought to make an essential contribution to cognition and to do that in virtue of their essential phenomenal quality. But spectrum inversions, stories about the possibility of seeing sounds and hearing colors, and science fictions about sensationless intelligences seem to indicate that the phenomenal quality of a sensation is not directly tied to its particular contribution to cognition. Sensations of red in different people may have different intrinsic characters; they are sensations of red because of their typical proximal causes, and they make their contribution to knowledge in virtue of this functional role. They provide, *ceteris paribus*, a ground for belief in the presence of a red object because of their typical, functional connection to red objects. What exactly is the cognitive contribution of the phenomenal quality itself ?

If phenomenal qualities are (in principle) independent of their role in the cognitive continuum, why is it that there are no sensations outside of cognitive processes? Of course, one could reply that there are, but sensations are detectable only when they occur within a cognitive system. I, however, and most other people, I believe, find the notion of wild, undetectable sensations pretty close to absurd.¹⁷ Sensations are states of cognitive processors, not independent entities. Let us not slip on what Lycan calls the banana peel. But if sensations are essentially states within cognitive systems, shouldn't we expect them to share in the functionality of the system? Of course, not every state nor every aspect of every state of a system needs to be functional; we could opt for a kind of epiphenomenalism, relegating phenomenal qualities to accidental excrescences of the organism. In that case, cognition without sensation should be no problem. But most people's intuitions say at least empirical cognition needs sensation. We are stuck with strong pulls in two different directions: the phenomenal qualities definitive of sensations seem without a clear role in cognition and in any case independent of whatever role they may play, but cognition without sensation seems impossible, and any intersubjective access we have to sensations is definitely through their role in our cognitive and affective behavior. Sellars once said something to the effect that the mind-body problem is not the really tough nut to crack, it is the mind-sensorium problem that is the real poser, and we now see why.

In the midst of these problems I would like to maintain that the notion of a sensation is the notion of a functional kind, the notion of something that plays a certain role in our cognitive processes. It may be that particular sensations or phenomenal qualities also have nonfunctional individuating conditions; I do not want to take that problem on here. My point is that the context "sensation of" picks out a type of inner state of a cognitive system functionally, that is, in terms of typical ties to the external world and the internal cognitive economy. We attribute sensations to beings that have sensations of kinds we can't share (bats have sonar sensations; aliens might have sensations of parts of the electromagnetic spectrum quite unavailable to us) because we can see or imagine how such states must enter into the cognitive-behavioral processes of the organism.

Applying this functional characterization assumes a complex set of goals and subgoals for the system, some of which cannot be characterized individualistically. In particular, attributing sensations in general to a being assumes that there is an appropriate functional decomposition of the internal states and structure of that being as a semantic processor. Even if we refuse to treat the content of a sensation as a kind of intentional content (and I am not in favor of conflating these two forms of "content"), and thus subject to external contamination via semantic holism, we still cannot extract sensation as a state-kind from contamination with the external. Attributing sensations to a being attributes a general functional, cognitive structure directed at certain kinds of goals and determined by a complex historical process.¹⁸ The being's possession of this structure ought to be naturalistically explainable, unless we believe in magic. Sensation as a state-kind does not have the kind of autonomy that would allow us to posit sensations without ties and relations to the physical world. These considerations show one way in which the notion of a disembodied but sensory mind is, if not incoherent, at least contrary to all the sensible, established procedures we have going for us.

In brief, the argument is that sensations are items of a kind individuated by their role in complex cognitive systems. Nonabstract functional description and explanation is methodologically committed to nonindividualistic individuation: However well one might be able to individu-

ate parts and states of a system solely on internal grounds, there will be some parts or states of the system, the functional individuation of which requires reference to things outside the current state of the system. Thus all nonabstract functional individuation is at least nonindividualistic by inheritance. Our actual practice in individuating sensations is straightforwardly nonindividualistic, for we individuate sensations by typical causes. Given the general anti-individualism of nonabstract functional description and explanation, we should take it that sensations are nonin-dividualistically individuated unless someone can demonstrate a viable, clearly individualistic mode of individuating sensations. Despite the strong intuitions of many that such a mode of individuation should be possible, I have never seen it. The burden of proof lies with those who insist that there is a set of nonfunctional individuating conditions for sensations. And if they are right, the price to be paid is that sensations lie forever beyond the reach of any naturalistic form of inquiry.

If the first thing the swamp creature does after solidifying is step on a muskrat trap, won't it have plenty of painful sensations? My answer is simply, No. It may have states that are sensation-analogs in important ways - we could certainly expect the same physical movements from the creature we would expect from Al, were he to step on a trap. But the swamp creature's states do not serve the purposes of sensations, they do not play any role in a cognitive system. In denying that the swamp creature had a real heart and real lungs, I called on principles of biological theory, but it is not clear that "sensation" is a biological category. Many people would not care that calling the swamp creature's inner states sensations violates principles of biological explanation. Sensations are *psychological* states, and their ties to biology are precisely what is in question here. Nonetheless, at a higher level of abstraction the point still holds: if we grant that the swamp creature's inner states are full-fledged sensations, we have thereby cut psychology off from the rest of the natural sciences.

This affords us adequate justification for denying to the creature any particular sensations, at least in the first few moments of its being. Though the creature possesses some of the forward-looking qualities of cognitive creatures, it lacks important etiological qualities that tie the cognitive to the natural. Lacking these etiological aspects, there are no

WILLEM A. DEVRIES

states of its cognitive structures that can properly be called sensations, that is, that can be called sensations without differing in methodologically significant ways from the kinds of states we normally would call sensations. One could opt to extend our term "sensation" to cover these states, but I don't see why we would want to, for we would then be putting sensations out of reach of any naturalistic methodology. To assume to begin with that sensations are beyond the reach of naturalistic methods is to beg important questions. The swamp creature is cognitive only if cognition is not ultimately a natural affair.

The obvious reply from the Autonomist is that it is equally question-begging to assume that everything, sensations included, gets a naturalistic treatment. The long tradition and strong intuitions that back the autonomy of the mental have to count for something and cannot be simply discarded. At this point the debate seems to have become a game of burden-tennis. Picking this debate up and progressing with it involves a careful analysis of the roots and motivations of naturalism and Autonomism, an analysis that would go far beyond the bounds of this paper.

Stuck between Autonomistic (or Cartesian) and naturalistic intuitions, the temptation may be to say that, since the creature possesses some but not all of the usual set of properties of familiar cognitive creatures, our concept of cognition and its contributory states are simply indeterminate in this case, so we don't know what to say. Some are even tempted to go a step further and say that such confusion about what to say shows us that there really isn't anything to say: there is no fact of the matter here. As I indicated above, this is a last-resort position. A thoroughgoing naturalism that is committed to the community of nature, however, has every reason to distinguish the swamp creature from its twin. Larger commitments to naturalism or anti-naturalism (supernaturalism?) hang in the balance, and I plunk down here on the naturalistic side. If only I knew exactly what my version of naturalism really amounted to!¹⁹

REFERENCES

- Baker, L. R. (1987) *Saving belief: A Critique of Physicalism*. Princeton: Princeton University Press.
- Block, Ned (1978) Troubles with Functionalism. In: W. Savage (ed.) *Perception and Cognition: Minnesota Studies in the Philosophy of Science, Vol. IX*.

Minneapolis: University of Minnesota Press.

- Block, Ned (1990) The Computer Model of the Mind. In: D. N. Osherson and E. E. Smith (eds.) *An Invitation to Cognitive Science, Volume 3: Thinking*. Cambridge, MA: MIT.
- Block, Ned (1992) Reply to deVries. Comment on an earlier, abbreviated version of this paper presented at the Eastern Division of the APA, Washington, DC.
- Burge, T. (1979) Individualism and the Mental, *Midwest Studies in Philosophy* IV: 73-121.
- Burge, T. (1982) Other Bodies. In: A. Woodfield (ed.) *Thought and Object*, pp. 97-120. Oxford: Oxford University Press.
- Burge, T. (1986) Cartesian Error and the Objectivity of Perception. In: P. Petit and J. McDowell (ers.) *Subject, Thought, and Context*, pp. 117-136. Oxford: Oxford University Press.
- Lewis, D. (1972) Psychophysical and Theoretical Identifications, *Australasian Journal of Philosophy* 50: 249-58.
- Lycan, W. (1987) *Consciousness*. Cambridge, MA: MIT/Bradford Books.
- Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT/Bradford.
- Millikan (1993) *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT/Bradford.
- Sellars, W. (1956) Empiricism and the Philosophy of Mind. In: H. Feigl and M. Scriven (eds.) *Minnesota Studies in the Philosophy of Science* I: 253-329. Minnesota: University of Minnesota Press; reprinted in W. Sellars (1963) *Science, Perception, and Reality*, pp. 127-196. London: Routledge and Kegan Paul.

Department of Philosophy
University of New Hampshire
Hamilton-Smith Hall
Durham, NH 03824
USA

NOTES

1. Burge seems willing to allow some form of intentionality to such creatures, albeit at a high price. The burden of my argument will be that that price is too high to take seriously, and Burge need not be so cautious.

2. See, e.g., Dretske (1981).

3. How to specify what counts as the right *kind* of causal contact to establish a semantic connection remains a - perhaps - *the* - major problem for this approach to intentionality.

4. Ryle, of course, tried to analyze away all mention of inner, mental states or processes in the notion of intelligence. He suggested instead an *adverbial* analysis of that notion: intelligent behavior, he claimed, is not a matter of the process generating the behavior, but of how the behavior itself takes place in its context. But his arguments do not exclude the possibility that people (or other organisms) can behave intelligently *because of the* kinds of processes that generate their behavior, and that our concept of intelligent behavior is sensitive to the aetiology of the behavior.

5. It is ironic that Block is himself responsible for many of the classic counterexamples to machine or abstract functionalism.

6. If my claim about functional analysis in the natural sciences is correct, then if intelligence is capable of treatment in the natural sciences, the notion of a trapped intelligence - an intelligence unable to give evidence of itself to any other external being - must be understood as an *abnormal condition*. For any natural intelligence, there must be some set of normal conditions in which that intelligence would exhibit itself via relations to its normal conditions. It might be that there are intelligences for whom normal conditions are such that *we* could not spot their intelligence in those conditions - they may be conditions inimical to human life or inaccessible to human instruments. But the notion of a natural, 'trapped' intelligence forever beyond all possible detection is as empty as the notion of the undetectable demon in my watch. I conclude that intelligence is context-flexible, but not *context-free*.

7. There may be a 'mechanism' that led to the swamp creature - namely a monumentally improbable series of quantum events. But that does not therefore justify a redescription of the thing in terms of the vocabulary of goals, organs, etc. One might reply that "different mechanisms can lead to the acquisition of the same goals." And that's true - but even if there are different ways to acquire goals, that surely doesn't mean that *anything goes*. Quantum mechanisms don't normally lead to the acquisition of any goals unless there is also an appropriate redescription of

the events in the vocabularies of biology and perhaps psychology. Why should they in this case? We have a surd here. At the very least the burden of proof should fall on the shoulders of those who claim that in this one-of-a-kind event we do have a goal-acquisition mechanism operating.

A student in Jennifer Armstrong's philosophy of mind seminar raised the interesting question of what one should say if quantum accidents were a relatively common mode of the origin of species. I believe a proper naturalism would have to take that into account and therefore would not have grounds for distinguishing the swamp creature from other organisms. The naturalist's rejection of the swamp creature's intentionality and biologicality is not in that sense an *a priori* dictate at all.

8. Analogous organs play the same general function and may even be quite similar in their structure and operation, but do not have a common evolutionary descent; homologous organs share a common evolutionary ancestor.

9. Ruth Millikan has been developing these kinds of points in many of her pieces. See Millikan (1984) and (1993).

10. For more details, see David Lewis (1972) or Ned Block (1978).

11. Lots of people have noticed this. See, for example, Block (1978) or Lycan (1987), especially chapter 3.

12. Phenomenalism can be seen as an Autonomistic individualism, for it attempts to reconstruct all our concepts and beliefs about extra-mental reality in terms of narrowly mental sensation concepts. Phenomenalism seems at least moribund, if not altogether dead, having succumbed to the criticisms of Chisholm, Sellars, and others. But Autonomism holds on through the efforts of those who think that, e.g., how one is appeared to is (in principle) capable of individuation independently of one's context.

13. It is no accident that this picture of intentional state individuation parallels the holistic picture of theoretical-concept or -statement individuation that we get from Quine, Sellars and their descendants. Both are cases of individuation by semantic properties, and semantic properties inevitably depend on 1) internal relations among the items in the representational system and 2) connections to the world.

14. Fodor offers a peculiar halfway house to more traditional Cartesian Autonomism. There is little Cartesian about his epistemology, and his metaphysics of the mind also seems deeply committed to naturalism. But in fact he balks at a thoroughgoing commitment to a naturalism of the mental, for he believes mental states are *formally defined*. As long as mental states are essentially formal, they are only contingently natural, and their ties to the environment are accidents of them.

15. I am assuming here the common sense standpoint from which such proper sensibles are simply out there to be discovered by us. Once we start wondering just what the ontological status of *red* is, things get very messy.

16. The full story here needs a little more detail: we could think of a belief in a nonrelational way by, e.g., thinking of it as a neurophysiological state, but the neuro-physiological characterization only contingently picks out the belief. The only way to guarantee you're picking out a belief is via the intentional (and therefore ultimately relational) characterization. But while we may think about the sensation via a relational characterization that mentions its typical causes, we also think that there is some non-relational mode of access to it that is not just contingently tied to the sensation. If there is direct acquaintance with a phenomenal quality, it is a way of thinking about the sensation that is noncontingently related to the sensation.

17. The notion of an actually undetected sensation that is a state of a cognitive system does not seem at all absurd to me. But I take it that such a sensation would still be a functional state of the cognitive system that is counterfactually, though not actually, involved in the cognitive processes of the organism.

18. My own intuitions may be irredeemably corrupted, but the kinds and complexity of sensations I am willing to attribute to an organism correlate directly with the kinds and complexity of cognitive structure I am willing to attribute to it.

19. I am grateful to all those who have contributed to my thinking on these issues. My colleagues Jennifer Armstrong, Drew Christie, Val Dusek, Paul McNamara, Tom Sullivan, Timm Triplett, and Ken Westphal initiated these reflections in our study group and several of them read and commented on multiple drafts of the evolving paper. Armstrong's philosophy of mind seminar read and discussed the paper with me in a very probing fashion, so thanks to Michael Laughy, David Meulenbroek, Eileen Tierney, and Kevin Wilcox. An earlier version was read to the University of Cincinnati Philosophy Colloquium, where I received helpful comments from Lynne Rudder Baker, Tyler Burge, Chris Gauker, Bernie Kobes, Bob Richardson and others. Special thanks to Ned Block for his elegant commentary and his patience with my criticisms.