

The Narrow and the Normative

Karen Neander

Neander@jhunix.hcf.jhu.edu

Paper for NYU series on Mental Causation

Other than for class purposes, please do not quote without the author's permission.

Comments welcome.

(1) Introduction

Intuitively, we act as we do not merely because we have beliefs and desires but because of *what* we believe and *what* we desire. It is because I believed that it would be cold today that I wore a coat. If the content of my belief had been different, had I believed instead that it would be warm, I might not have done so. On one interpretation of the claim that we act as we do because of what we believe and desire the *contents* of our mental states make a causal difference, they have causal efficacy. Intuitive as this may seem, however, it is dubious.

There are a number of problems with the idea that mental properties –intentional properties, in particular— have causal efficacy, even assuming Physicalism. Jaegwon Kim, for instance, lists three. Two of these problems concern non-reductive physicalism.¹ One is to do with the anomalousness of the mental. The problem arises if there are no psychological laws and if causal relations must instantiate causal laws. Another is the problem of causal exclusion, which arises if mental properties are not identical to physical properties given the completeness and closure of the physical world.² If the physical properties of preceding events are causally sufficient for all that happens mental properties appear to be causally redundant. Short of systematic,

pervasive over determination, they can have no causal efficacy at all. The third problem, which is the one that interests me in this paper, is not due to the irreducibility of mental properties but to their extrinsicity. This is the problem of extrinsic causation, which arises in view of the fact that the intentional properties of an individual's mental states are extrinsic to that individual whereas the inner causes of the individual's behavior appear to be intrinsic.

Kim introduces this latter problem via Stephen Stich's Syntacticalism, which is the thesis that computational processes are sensitive to the syntax and not the semantics of the representations involved.³ Representations are transformed and manipulated in computational processes and these transformations and manipulations can be interpreted as semantically cogent. However these processes are causally sensitive to the physical properties of the representations and not to their semantic significance. For instance, in a conventional electronic computer, output depends on such things as whether a pathway is on or off and on how it connects with other pathways. In contrast, it makes no causal difference whether a pattern of ons and offs represents a vote for President, a consumer's preference in the color of tulips, or a record of a move in a game of chess. If our brains are computational systems, the same kind of thing can be said of the neural processes that are computational. Indeed, the problem of the causal irrelevance of intentional properties remains with us even if the computational theory is discarded. Whether or not neural processes are computational processes, if they are physical processes, it would seem that they are sensitive to the ordinary physical properties of the system regardless of their semantic significance.

Actually, there are a number of ways to elaborate this intuition. One way is via Kim's problem of causal exclusion, the second of the problems mentioned above. One reason for thinking that neural processes are not causally sensitive to semantic significance is that having semantic significance is not identical to having a basic physical property or a complex of basic physical properties. Since the basic physical properties of neural events suffice to produce the motor output, their intentional properties seem to be causally superfluous. Another way to elaborate the intuition is via the problem of extrinsic causation, which as I have said is the problem that interests me here. I will say more about the nature of this problem in the next section.

Before I do so, however, let me explain my aims with respect to it. I do not plan to argue that intentional properties have causal efficacy. What interests me here is not so much whether intentional properties have causal efficacy as what follows for their scientific importance on the assumption that they do not. According to Kim,

“A science that invokes mental phenomenon in its explanations is presumptively committed to their causal efficacy; for any phenomena to have an explanatory role, its presence or absence in a given situation must make a difference –a causal difference.”
(1998, p.31)

Thus, according to Kim, what follows from their lack of causal efficacy is their lack of scientific importance. I will argue that this is wrong.

There is, of course, a weak presumption along these lines, in the sense that there is at least a *prima facie* puzzle as to why a science might appeal to properties that lack causal efficacy. In so far as science is in the business of providing causal explanations, properties that have causal efficacy have an obvious role in science and properties that lack causal efficacy lack that obvious role. However, the presumption does not survive a careful analysis of scientific practice. Scientists themselves feel no presumptive

commitment to the causal efficacy of all of the properties they invoke. Of course, as amateur philosophers of science, they are as likely to be impressed by Kim's presumptive principle as is anyone else. But they do not abide by this principle in practice and they are right not to do so. In a while I will turn to the case of normal (or proper) function in biology in order to show that this is so. Having a normal function is an extrinsic property, and there is no presumption that it has causal efficacy, yet appeals to normal function play an important explanatory and classificatory role in biology, as I will explain.

As some readers will already have realized, I have not chosen the notion of a normal function at random from among those I might have chosen to illustrate my general point. On my view, there is an important connection between function and content and so, on my view, the fact that they are both extrinsic is not accidental.⁴ Philosophers are still apt to think that neurobiology is a non-intentional science. They are still apt to think that it does not traffic (as we say, suggesting that the practice is somewhat disreputable) in talk of representations. While this may once have been true, when neurobiologists were more exclusively focussed on understanding the molecular details of neural processing, it is true no longer. Neurobiologists are now trying to determine the functions of large-scale neural components, and such functions are often –perhaps invariably— described in representational terms. This talk of the representational functions of neural components is where content –representation and the possibility of misrepresentation— makes its first appearance in the scientific picture of the brain as we move in the direction from “lower-level” analyses to “higher-level” analyses of brain functioning. Representational functions thus seem likely to form the conceptual bridge between what might be

considered straightforwardly biological descriptions and straightforwardly psychological descriptions of the brain. So at least one promising approach to understanding the scientific significance of content is to try to understand the scientific significance of the representational functions of neural components. And if that is our aim, it is useful, indeed imperative, to try to understand the scientific significance of functions in general. Once we have done so, we will be better placed to tackle the question of what role specifically representational functions play.

Thus, what follows has a dual motivation. One goal is to illustrate the way in which one extrinsic concept plays a significant scientific role. I hope that this motivation suffices for those who reject the line of reasoning offered in the preceding paragraph. If my comments on function are correct, I will have offered support for the claim that a property can be extrinsic, and can clearly lack causal efficacy, and yet can have scientific significance. However, there is also another goal, which I will not reach in this paper, but to which this paper is also directed. This more ultimate goal is to cast some light on the scientific significance of content itself.

(2) Twin-cases and semantic norms

Roughly speaking, the intrinsic properties of a system at time t , are those of its properties that are wholly realized within it at t . Consider a particular individual at a time: for example, yourself at this very moment. Now consider some of your properties (loosely construed) at this time. Perhaps you are a sibling, perhaps you are sitting on a chair, and perhaps you slept well last night. In addition, it may be that you are less than six feet tall and you certainly have many carbon-based molecules between your ears.

Some of these features are wholly realized within you (“from your skin in”) at this moment and some are not. The contrast is often explained in terms of what it would take to duplicate these features. Imagine that we have God-like powers and that we want to make a second individual who is as exactly like you as is possible. We could duplicate some of your features just by duplicating the arrangement of elementary physical particles that presently constitutes you. This would be enough to capture the height and the carbon molecules in the head. But it would not suffice to make your duplicate a sibling, because for that we would need to give him or her a history of having been born of parents, one of whom at least had had another child. We’d also need to duplicate some of your environment to ensure that he or she was seated on a chair and some of your recent history to give him or her a good sleep last night. Sometimes the concept of an intrinsic concept is defined with respect to supervenience. One set of properties is said to supervene on another set of properties if the second “fixes” the first. Or, more precisely, if there can be no difference between two individuals with respect to the properties from the first set without a difference in them with respect to the properties of the second set. An intrinsic property of an individual, A, at time, t, is one that supervenes on the spatially and temporally local properties of A at t.

Some mental properties may be intrinsic. Qualia are plausible candidates. However, intentional properties are plausibly regarded as extrinsic. This claim is usually justified by appeal to certain modal and semantic intuitions concerning the familiar “twin cases” of the kind offered by Hilary Putnam and Tyler Burge.⁵ Of course, these “twin cases” are not really cases of *twins*. They are instead duplication scenarios of the kind I was just discussing. In this case, however, we imagine ourselves varying certain

contextual conditions in order to see what, if anything, changes with respect to the content of the duplicate's mental states. It seems we can alter the content of a duplicate's mental states by altering aspects of the original's history and/or social and/or physical environment.

Take Putnam's famous twin-Earth case, for instance. In this case we imagine an individual, Oscar, who has always lived on Earth and who, like the rest of us, has a term that refers to water—i.e., “water”. Oscar knows nothing of the molecular constitution of water. In addition, we imagine another individual, Toscar, who has always lived on another planet, Twin-Earth, somewhere in the universe. Toscar and Twin-Earth are as exactly like Oscar and Earth, respectively, as is possible, with just one exception: unlike Earth, Twin-Earth has no H₂O. In its place is a foreign molecule, designated “XYZ”, which somehow plays the same causal role as H₂O (at least as far as one could tell, if one were restricted to the epistemological resources available to Oscar and Toscar). The liquid composed of XYZ looks and tastes the same as the liquid composed of H₂O. It too falls from the skies, flows in the streams, quenches the thirst, etc. Following Kripke, Putnam's claim is that the vocable “water” as spoken by Oscar refers to H₂O, whereas the vocable “water” as spoken by Toscar refers to XYZ. Indeed, since water (i.e., what we refer to as “water”) *is* H₂O, the English word “water” refers to water whereas the Twin-English word “water” does not. When Toscar says in Twin-English, “I want a drink of water”, he does not refer to water (i.e., to what we refer to as “water”, H₂O) but to twin-water (i.e., to what he refers to as “water”, XYZ). For parallel reasons, their corresponding thoughts are also about different things. When Oscar and Toscar head for the village well to fill a jug, it is water that Oscar believes he will find and it is twin-water

that Toscar believes he will find. Thus, the contents of their thoughts are different, despite the fact that they are, at least in all relevant respects, intrinsic duplicates.

The content of their inner representations thus seems to be extrinsic not intrinsic, since it seems to be determined in part by the environment in which they have lived. Yet, the extrinsic character of content casts its causal efficacy into doubt. The worry is partly about the contiguity of causation. Their past and their environment cannot affect their behavior except by means of intrinsic mediation. Our earlier history can affect our later behavior, but only by virtue of the fact that it can alter us which can cause us to behave in certain ways. (My learning about the antique show on Tuesday was a cause of my going to the show on Wednesday, but only because my learning of the show on Tuesday resulted in memory traces that persisted on Wednesday.) So too our environment can affect our behavior, but only by virtue of the fact that it can impact upon our inner state and this inner state can cause us to behave in certain ways. (The ringing of the doorbell caused me to open the door, but only because I heard the doorbell because my sensory receptors transformed the sound into neural firings.) Two individuals who are intrinsic doppelgangers will produce the same motor output *qua* motor output given the same input, regardless of their intentional differences.⁶ So it appears that intentional difference *per se* is inconsequential. The difference between Oscar's thoughts of water and Toscar's thoughts of twin-water makes no difference, or no causal difference, to their motor output.⁷ It's a puzzle, therefore why scientific psychology, which is presumably concerned with explaining and predicting behavior, makes reference to intentional properties.

One response that has been made to this argument is to maintain that behavior, as opposed to mere movement (or what I am calling “motor output *qua* motor output”) is intentional.⁸ There is an important sense in which one does not count as going to fetch water unless it is water that one intends to fetch. So, it is not true that intrinsic doppelgangers will always produce the same behavioral output in the same situation. Oscar reaches for water and Toscar reaches for twin-water, so they do not produce the same behavior when, desirous of what they each refer to as “water” they think they see some within reach. Were they switched, so that (unknown to them) Oscar were teletransported to Twin-Earth and Toscar to Earth, Oscar would still be reaching for water, although it would now be XYZ he grasped, and Toscar would still be reaching for twin-water, although it would now be H₂O that he grasped. For what each of them would be aiming at would be what each of them desired and believed to be before them, and that depends on the content of their beliefs and desires, and not on what is there before them. Thus, it is argued, the behavioral effects of intentional states are different in so far as there are intentional differences in their causes.

It is true that differences in intention “make a difference” to behavioral outcome in the way described. However, this does not show that content has causal efficacy, because there are non-causal as well as causal ways in which something can “make a difference”.⁹ Someone might claim that water drunk on Tuesdays has special effects that water drunk on Wednesdays does not have. The water drunk on Tuesdays has the effect of producing water-drunk-on-a-Tuesday-effected-thirst-quenching whereas the water drunk on Wednesday has the effect of producing water-drunk-on-a-Wednesday-effected-thirst-quenching. It is apparent that, in this case at least, that there is no causal difference, just a

conceptual difference. That is, the difference is just a difference in how the outcome can be classified due to a difference in how the cause can be classified, which is a purely conceptual difference.

Some philosophers have tried to argue that scientific psychology can forego the distinctions among contents that the various “twin-cases” generate. In doing so, they tend to assume that, if so, it can do without an extrinsic (or “broad”) notion of content and make do with an intrinsic (or “narrow”) notion. Thus they argue that psychology can make do with “disjunctive” contents such as [*water or twin-water or ... etc.*] or with allegedly “neutral contents” such as [*transparent, tasteless liquid that falls from the sky, quenches our thirst, ... etc.*].¹⁰ However, this is like arguing that we can do without dessert when the real question is whether we can forego eating altogether. Even if scientific psychology could manage without a distinction between Oscar’s thoughts about water and Twin-Oscar’s thoughts about twin-water, it would not follow that psychology could make do without broad content.

Suppose that, for the purposes of scientific psychology, Oscar and Toscar can count as thinking of the same thing when they walk to the well. Then, in the substitution scenario, Oscar will not be mistaken when he reaches for XYZ thinking that it’s water, and Toscar will not be mistaken when he reaches for H₂O thinking it’s twin-water. We are supposing that scientific psychology can sacrifice this much error. It does not follow that scientific psychology can forego all possibility of representational error. It does not follow that it has no need to conceive of some inner states as misperceptions, misrememberings, misrepresentations, and so on. Even if we may think of Oscar and Toscar as correct as long as it is water or twin-water that they grasp, it does not follow

that we need not consider them in error when what they grasp is gin. Yet scientific psychology cannot treat such an error as an error unless it has a normative notion of content and a normative notion of content is by all extant accounts broad.

A normative notion of content is one that permits the possibility of error. Indeed, there is no content unless there is the possibility of error. For a representation to have content is for there to be certain things to which, were it applied to them, it would be correctly applied, and certain other things to which, were it applied to them, it would be incorrectly applied.¹¹ Meaningful distinctions among representations require contents that exclude as well as include. Oscar's and Toscar's mental representations of water or twin-water, however "neutral" or disjunctive their contents might be, must be such that *they would be wrong if they were applied to some things* (gin, cats, potatoes, buckets, chimneys, etc.) or else they are meaningless.

It is generally accepted that so-called "narrow content" is not normative.¹² I will leave it to readers to cast their minds over whichever theories of broad content they find most plausible (or, perhaps I should say, least implausible). Some theories tell us that we need to know about Oscar's linguistic community, or about his learning history, or about his evolutionary history, or about the asymmetric dependencies that hold when his perceptual and cognitive systems are function normally, or about some combination of these things. But each even halfway plausible theory of normative content makes some appeal to extrinsic features.¹³ Moreover, this is probably what we should expect. In attributing content to a representation we are attributing a standard of correctness to it. In doing so, we are distinguishing between the way in which the system that tokens the representation actually responds and the way in which it is, in some sense, "supposed" to

respond. The system's intrinsic properties determine the former, but it is hard to see how they could supply a sensible standard for the latter.¹⁴ Those who argue that psychology can do without broad content are obliged to show either that psychology can do without content altogether. Or they are obliged to do what has not yet been done, which is give us reason to think that normative content could be narrow.

Notice that the point is not just that, in order to assess whether an individual is in error, we need to examine his or her context. Of course, that's also true. Even when we know that Oscar is thinking, "that's water in the glass", we need to check to see if it is really water in the glass to see if he is right or wrong. Thus to classify mental states as correct or incorrect is to classify them on extrinsic grounds. However, in order to know that he is thinking, "that's water in the glass", and hence in order to classify his thought as a thought of that kind, we already need to know something of his context. For he cannot be thinking, "that's water in the glass", unless he *would be wrong* were it gin, or any other number of things. We need to know something of his context because we need to know whatever it is that we need to know in order to attribute normative content to his inner state.

Now let's turn to the related or at least somewhat parallel case of function in biology. This too is, in a sense, a normative notion.

(3) Narrow and Normal Functions

The concept of function that I have in mind is the concept that is in play in biological contexts in statements of the form, "the/a function of X is Z", e.g., "a function of the pineal gland is to secrete melatonin". This is the concept of function that permits the

possibility of malfunction and that underlies a distinction between normal and abnormal functioning and hence involves some notion of functional norms. In this sense, my pineal gland can both have the function of secreting melatonin and lack the disposition to perform that function. Indeed, this is what is required for the possibility that it might malfunction. Thus the normal function of a trait is distinct from its synchronic dispositions. It is more a matter of what something is “supposed” to do than what it is disposed to do.

Some find my use of the terms “normative” and “norms” objectionable in this context, since they like to preserve these terms for prescriptive contexts. But as there is a sense of “norm” in which even statistical norms are norms I think it is they who are, and not I who am, departing from standard use. In any case, we need to be very clear that I am not claiming that the relevant notion of normal function is prescriptive. On the contrary, I believe that functional norms are fully naturalistic and hence are not essentially prescriptive. I should add that nothing in particular hangs on the words themselves. So readers are welcome to make appropriate substitutions—e.g., “quasi-norm” and “quasi-normative”. (Note that the same terminological issue arises with respect to talk of semantic norms, assuming that they are also naturalistic.)

It is uncontroversial that some concept of function is at the heart of physiology, for physiology just is the study of the functional organization of living systems. Those engaged in the physiological side of biology will often express their main aim in terms of this concept. Consider, for instance, these avowals:

The goal of modern molecular cell biology is nothing short of understanding the biochemical, cellular and organismal functions of all the proteins encoded in the genome.
(Lodish, et. al. 1999, p. 281.)

[W]hen we try to understand how any organism works, the same two themes apply. One is the correlation of structure and function.

(Campbell, 1987/90, p.781.)

[T]he problem of the neurobiologist is to learn how the order and complexity relate to the function.

(Hubel, 1995, p.4)

This concept –the one that is central to the very *raison d'être* of physiology— is the key concept here. It is, as I have said, a normative concept, in the weak sense already described, as is shown by its association with locutions such as *normal function*, *functioning properly*, *abnormal functioning*, *functional deficit*, *dysfunction*, *functional impairment*, *loss of function*, *malfunctioning*, and so on. Those who are unfamiliar with its use in physiology and neurophysiology can gain a quick impression of its use –or at least of the fact that it is used—if they browse the following sample of sentences taken from physiological and neurophysiological research.

Once the foreign DNA is inside the host cell, enzymes that probably *function normally* in DNA repair and recombination join the fragments of foreign DNA with the host cell's chromosomes.

(Lodish, 282.)

These head groups may interact with the membrane proteins in ways that are essential for the *proper functioning* of the proteins.

(Shepherd, 45)

Glutamate is generally conceded to be the most important transmitter for *normal brain function*.

(Purves et. al., 9.)

In fact, substantial evidence indicates that neurotrophic factors can protect and even restore *impaired functions* resulting from trauma, aging, and ischemia.

(Kipiranova et. al. 1999, 21)

Certain chromosomal aberrations *lead to deficits or malfunctioning* in nerve cells, as in Downe's syndrome (mongolism).

(Shepherd, 48)

Most of our knowledge about these functions has come from the study of individuals who suffer from *malfunctioning* caused by cortical lesions.

(Gleitman, 1999, 33.)

Published studies have largely replicated earlier work, while at the same time underscoring two important trends: delineating the separable neural substrates of depression and anxiety and examining relations between patterns of regional brain function and specific types of *cognitive dysfunction*. ... New evidence supports

previous work documenting diminished right parietal activity and *compromised right brain function* in depression. (Davidson, 1999, 229.)

It is indisputable that physiology makes use of some notion of normal function. The open questions are (1) how we are to understand the notion? And (2) what role or roles does it play? I will give a brief introduction to responses to the first question in the remainder of this section and then address the second.

There are two main options on offer by way of an account of functional norms. One is that functional norms are statistical and the other is that they are teleological. In either case they are extrinsic.

I have long defended the view that functional norms are teleological.¹⁵ In my view, the relevant concept of function is the concept of what something is *for*. In the context of contemporary somatic biology, it is more particularly the concept of what something was *selected for by natural selection*. Such an account of normal functions is known as an *etioloical account*. On this kind of account, it is a function of the pineal gland to secrete melatonin because pineal glands were selected for secreting melatonin by natural selection. Your pineal gland has the function of secreting melatonin, not because it does so, but because past pineal glands did so and in doing so they contributed to the fitness of our ancestors and so to the selection of the genotype responsible for the phenotype. Whether a token trait (your pineal gland) has a certain function depends on its causal connections with past traits (ancestral pineal glands) and on their selection history.

The other main option is that functional norms are statistical. The idea in this case is that what is normal is a matter of what is numerically typical within a species (or

within a sex or age group or some such part of a species). On this kind of account, it is the function of the human pineal gland to secrete melatonin because secreting melatonin is a contribution that pineal glands typically make to the fitness of humans. Although this kind of account was once popular it now has few supporters. The main problem is that atypical functioning can be normal and typical functioning can be abnormal.¹⁶ Certainly, there is nothing conceptually incoherent in the idea that a particular form of malfunction can become typical within a species: for example, AIDS could become typical among humans and, if it did, the malfunctioning of the immune system that is involved would still count as malfunctioning. Those who favor statistical accounts need to tell us how to abstract the statistics from a population so as to avoid these kinds of consequences. In any case, a statistical account also renders normal function extrinsic. Whether a token trait (your pineal gland) has a certain function depends on what other items (other pineal glands in other humans) are doing or are doing or are capable of doing.

Some accounts of biological function that are on offer make no mention of either selection history or statistical norms. But a close inspection invariably reveals that they are under-specified with respect to functional norms. Robert Cummins' popular account of functions [1975] is a case in point. When he offered his account, Cummins was not focussed on the problem of accounting for functional norms and his account is under-specified with respect to that problem. Since I will want to refer to Cummins' account again later, let's take a look.

Cummins' most important contribution to the philosophical literature on functions was his recognition of the role of function attributions in operational explanations of complex systems. Others also saw that function attributions had such a role, but

Cummins account of it has been especially influential. An operational explanation is –as the name suggests- an explanation of how a system operates. Although the contrast is perhaps too simple, it helps to think of operational explanations as more synchronic and less diachronic than certain other styles of explanation, e.g., historical explanations of evolutionary change. Operational explanations explain certain overall capacities of a system at a time in terms of the capacities of the parts of that system at that time. Operational explanations are, for that reason, often thought of as ahistorical as opposed to historical explanations.

According to Cummins, to give a statement of the form, “X has the function to Z in a system, S”, is to contribute to an operational explanation of S, or more precisely, of some particular capacity of S that is under analysis. Cummins thinks of these operational explanations in terms of the conventional divide-and-conquer-strategy, which he refers to as the ‘analytical strategy’. It is also often referred to as the reductive strategy.¹⁷ This is a recursive conceptual decomposition of a system into its component parts and their respective contributions to the overall working of the system. This is how Cummins puts it,

The biologically significant capacities of an entire organism are explained by analyzing the organism into a number of "systems" --the circulatory system, the digestive system, the nervous system, etc., --each of which has its characteristic capacities. These capacities are in turn analyzed into capacities of component organs and structures. Ideally, this strategy is pressed until the analyzing capacities are amenable to the subsumption strategy. (Cummins, 1975, p.760-1)

He suggests the following definition.¹⁸

Cummins’ DEF: The function of X in S is Z relative to an analytical account, A, of S’s capacity to Y just in case,
(a) X can Z in S, and
(b) A appropriately and adequately accounts for S’s capacity to Y by, in part, appealing to the capacity of X to Z in S.

According to Cummins, the relevant overall capacity of a system (Y) is simply that which researchers are interested in explaining via the analysis (A). Thus the overall capacity is a capacity to survive and reproduce if that is what scientists are interested in explaining. He also tells us that function attributions are more or less appropriate to the extent to which the analyzed capacity is complex—i.e., to the extent that it is made of many different parts doing many different kinds of things. So, according to this definition, a trait's function is its capacity to contribute to a more complex capacity of the containing system, such that the second capacity is under analysis and the trait's said capacity features in that analysis. Your pineal gland has the function of secreting melatonin because it has the capacity to do so, and its capacity to do so features in an adequate and appropriate analysis of your diurnal cycles and, in turn, your capacity to survive and reproduce. Assuming, that is, that someone is interested in explaining that.

Are functions of the kind that Cummins defines intrinsic? It would seem that extrinsic factors enter into the definition in several ways. First, through the interests of the researcher and, second, through the capacities of traits (for example, to contribute to survival and reproduction) which in part depend on their surroundings. However, we might maintain that functions are not so much extrinsic as relative in these respects. That is, we might say that traits have intrinsic functions relative to research interests and environments. One research interest is in explaining survival and reproduction. So we might, for example, refer to all Cummins-functions that are relative to that research interest as SR-Cummins-functions. My heart right now has the SR-Cummins-function of pumping blood whether or not anyone is interested in explaining my survival. Were they to be so interested, this capacity of my heart would feature in an analytic explanation of

my ability to survive. In a similar way, we might make function attributions environment-relative. Relative to its arctic environment, the polar bear's fur has the SR-Cummins-functions of providing warmth and camouflage, since in that environment it has these capacities and they contribute to the bear's comfort and hunting prowess. Even if the bear were not in that environment its fur would have these capacities if it were, and so such environment relative capacities are possessed independently of which environment the bear is in. Thus, we could perhaps rescue an intrinsic concept of function by this move.¹⁹ I don't think such a suggestion is promising as an account of how functions are employed in biology, but I will not pursue this here since my present aim is not to evaluate these accounts, as opposed to spell out their implications for extrinsic concepts in science.

Whether or not we can render Cummins-style functions suitably intrinsic with respect to those aspects mentioned in the preceding paragraph, there is another way in which extrinsic factors are relevant. If we are to interpret Cummins as offering an account of normal function, we must read his definition as applying to a type and not a token trait. To see this, let's see what follows when we do the opposite. Assume that the "X" of his definition stands for a token trait (e.g., your pineal gland). On that assumption, malfunction is impossible on his account. As you will recall, malfunction is only possible if a token trait can at one and the same time both have a function and lack the capacity to perform it. But this is impossible, given Cummins' first requirement, because that says that a trait must have the capacity to Z if it is to have the function of Z-ing. So, either your pineal gland has the capacity to secrete melatonin, in which case that might be its function, but in which case it is not malfunctioning. Or your pineal gland

lacks the capacity to secrete melatonin, in which case that cannot be its function, so that again it is not malfunctioning.

We introduce functional norms by treating token traits as tokens of a type and by deriving a standard of normal functioning for the tokens from facts about the type. I suggest that this is the central source of externalism in the case of normative properties of this kind. At any rate, it appears to be the central source of externalism in the case of this particular normative property. Extrinsic features now enter in two ways. One is via whatever connections token traits must have to each other in order for them to count as tokens of the same type. I have not said anything about this so far. But, for instance, if they must be homologous or if they must be traits of organisms in the same species or of organisms in some larger taxonomic group, extrinsic considerations are involved. The other is via the kinds of facts about other tokens of the type that determine the norms. If, as I said above, the two options are that this is a matter of the selection history of the phenotype or that it is a matter of what is typical of instances of the phenotype in the species, then extrinsic considerations are involved again. Given Cummins' staunch holdout against etiological accounts, he is likely to opt for the second alternative.²⁰ But, in either case, the concept becomes extrinsic at this point. Moreover, I see no way to employ the same technical maneuver as was employed before without patently collapsing the distinction between intrinsic and extrinsic concepts. We could technically import possible evolutionary histories and possible species memberships inside the concept of a normal function. So that we might speak of SR-Cummins functions that are relative to environments and species and/or evolutionary histories. But at this point it is hard to see why anything should count as extrinsic, since this move will always be available.²¹

In what follows I will sometimes refer to a non-normative Cummins-style function. Consider this to be an actual causal capacity of a token trait. Or, a little more fancy, consider it to be an actual causal capacity of a token trait to contribute to the containing organism's ability to survive and reproduce relative to a certain environment. We have not settled whether this should be considered an intrinsic notion, but we can assume that it is for the sake of the argument. In contrast to this, I will also refer to a notion of normal function, which I will assume is extrinsic. In particular, I will assume that the normal biological function of an ordinary somatic trait depends of what ancestral traits were selected for. Although I have not argued for this beyond casting a few quick aspersions in the direction of the statistical alternative, I have argued for it elsewhere.²² The reader should keep in mind that my present aim is not to show that such an account is correct. My aim is to show that an extrinsic concept can be of scientific importance. The main thing for readers to appreciate is that some concept of normal function is in play in physiology and it seems it must be an extrinsic concept.

Of course, there is a *prima facie* puzzle about the scientific significance of such an extrinsic concept. It makes no difference to a trait's causal powers that it has a certain normal function. Whether I live or die does not depend on whether I have a heart that has the function of pumping blood, in this sense. Whether I live or die depends on whether I have a heart that is actually capable of pumping blood. So, what useful theoretical role might a concept of the extrinsic and normative kind play? In what follows I claim that it has an important role both in classifying traits and in providing explanations of the functional organization of kinds of living systems.

(4) Normal Function and Classification

Philosophers have long believed that biology is richly endowed with functional categories. The idea has been that a heart counts as a heart because it is for pumping blood and eyes count as eyes because they are for seeing, and so on. More generally, on the traditional view, the traits of organisms are often classified into types wholly or in part on the basis of their biological function. This view, which extends at least as far back as Aristotle, has long been entrenched in contemporary philosophy of biology (see e.g., Morton Beckner, 1959, 112-118). Indeed, it has become accepted philosophical wisdom more generally. For instance, we find Kim (1996, 77) explaining that "... even many biological concepts (e.g., the gene, heart) appear to have an essentially functional component". And Tyler Burge, in defending a scientific role for extrinsic concepts, asserts that "[t]o be a heart, an entity has to have the normal, evolved function of pumping blood in a body's circulatory system" (1989, ICP, 312).

Ask almost anyone in philosophy of mind or cognitive science why we want functional categories when we do and you will almost certainly be told that it has something to do with wanting categories that are multiple physical realizable. The claim is that we want functional categories when we want to make salient the shared dispositions of certain items that have diverse structural designs and physical realizations. Consider the components in computers, for instance. It can be useful to classify adders together, on the basis of the function they perform, whatever the physical constitution and architecture of these adders. For certain purposes, it can be sufficient to know that an adder will add; i.e., that it will take as inputs physical symbols representing numbers and generate as output a physical symbol representing the sum of those numbers.

This standard motivational story is too simple in the biological context. For one thing, it tells us that items that are co-classified in a functional category have the same disposition. True, it does allow that there are differences in disposition among functional items that are due to the differences in the underlying mechanism. It is not meant as a denial, for example, of the fact that an adder made of paper tape will burn more readily than an adder made of silicon. However, this standard motivational story implies that all of the items in the functional category possess the same broad, defining, disposition. It implies that all adders can add. It implies that all pineal glands can secrete melatonin, that all hearts can pump blood, that all eyes can see. If only.

There is some dispute over whether functional categories are really as pervasive and important in biology as was traditionally supposed. Elsewhere, I argue that claims to the contrary are misguided.²³ In any case, this dispute is cold comfort to those who would argue for intrinsic classification since the alternative to function that is proposed is usually homology and homology is also an evolutionary notion.²⁴ The standard definition of the term tells us that the same trait (or similar traits) in two or more separate species are homologous when that trait (or those traits) have been derived without interruption but with or without modification from a common ancestor. (Thus the wings of pigeons and parrots are homologous but the wings of birds and bats are not.) I do not argue against classification of traits by homology, rather I argue that homology requires assistance from other complementary principles, especially but not exclusively to do with function. We can easily see that homology requires assistance when we consider that even in order to specify whether two traits are homologous we need to characterize the trait in question. For instance, the forelimbs of bats and birds are homologous *as*

forelimbs but they are not homologous *as wings*. Thus, traits are not homologous *simpliciter* but only as this or that type of trait.

Another way to express this point is to say that homologous categories have a hierarchical structure. We can identify smaller homologous groupings within larger homologous groupings. The forelimbs of bats and birds are both members of the larger category, vertebrate forelimbs, but they belong in separate smaller categories also (avian wings, etc.). Thus we differentiate homologous groupings from related homologous groupings. If we think of a homologue as a branching lineage of a phenotype, the issue is how to stop and start the lineage for the purpose of creating useful categories. The same lineage can be segmented up in different ways. My suggestion is that one of the main ways by means of which we segment phenotypic lineages is by drawing conceptual lines at those places where there have been significant shifts in selection. In other words, we often start and stop trait types at those places where there has been a change in what there was selection for. Token traits are classified according to the segment of the lineage they derive from. Thus a token trait is an appendix (possibly a malfunctioning one) as opposed to a vestigial appendix if it belongs to a segment of the phenotypic lineage in which there is/was selection for the digestion of cellulose. What there has been selection for has a structural as well as a functional component; traits are selected for their dispositions but they are also selected for the structures that bestow those dispositions on them: these are really just two aspects of the same thing.

Categories that are intuitively functional categories or that have traditionally been regarded as functional categories are *ab/normality inclusive categories*. In the biological case, these are categories of traits of organisms that include both normal and abnormal

items. Biology has very many ab/normality inclusive categories. For example, the categories of hearts, livers and neural pathways are ab/normality inclusive because they include abnormal hearts, livers and neural pathways, respectively, along with normal ones. In contrast, the category of normal hearts is not ab/normality inclusive since it excludes abnormal hearts. And the categories of livers with cirrhosis and cAMP deficient neural pathways are not ab/normality inclusive since they exclude normal livers and normal neural pathways.

Notice further that if these ab/normality inclusive categories are indeed functional categories the concept of a *normal function* is clearly implicated. An intrinsic Cummins-style concept of function cannot play the definitive role since it is not sufficiently ab/normality inclusive. Suppose it is true that a heart counts as a heart in virtue of its function. Since it still counts as a heart even if it is unable to perform its function, it must count as a heart in virtue of its normal function and not in virtue of its present causal capacities. It must count as a heart in virtue of what it is “supposed” to do as opposed to what it is disposed to do. Thus, if such categories are functional categories, they are extrinsic categories. If such categories are functional categories, it is a requirement on membership that an item possess a certain extrinsic property, a certain normal function.

In view of this, there are really two *prima facie* puzzles about extrinsic classification. When we focus on intrinsic twin cases, the worry about extrinsic classification appears to be this. If we classify traits on the basis of extrinsic features – e.g., normative function or content— our distinctions seem to be *more fine-grained than they need to be*. Predictions and causal explanations appeal to generalizations and a key question is how best to frame these generalizations. We want to know what kinds of

kinds (or types or categories or classifications) such generalizations should refer to. One reasonable principle is that, other things being equal, a generalization of greater scope is to be preferred to a generalization of lesser scope. It's true that dry wood burns well and it's also true that dry wood burns well in the evening, but why use the second generalization, which has a much more limited application, when the first is equally accurate? Similarly, generalizations expressed in terms of normative function or content seem to limit the scope of the generalization without increasing the accuracy. Since two items that share the same intrinsic features share the features that suffice to ensure the same output given the same input, it would also seem that they should be co-classified for explanatory and predictive purposes.

Donald Davidson's Swampman illustrates the problem for functions as well as for content. Swampman is an imaginary creature who is the product of an entirely random collision of elementary particles. Purely coincidentally, when Swampman coalesces into existence, he is a molecule for molecule replica of Davidson at a certain point in time, t^1 . Assume that it is a requirement on membership in the category of hearts that something *has the normal function of pumping blood*. Assume also that it is a requirement on membership in the category of hearts* that something *has the disposition to pump blood*. Then Davidson has a heart at t^1 but Swampman does not since Swampman has no ancestors and hence no selection history. In contrast, if Davidson has a heart* at t^1 then on arrival so does Swampman, for Davidson's heart and Swampman's heart* are molecule for molecule replicas, so if Davidson's heart can pump blood, Swampman's heart* can too. In view of intrinsic twin cases of this sort, the issue seems to be the same as with extrinsic mental classification. Why use classifications that exclude items with the same

causal dispositions as those that are included? Why exclude Swampman's heart*? Why frame generalizations in terms of hearts rather than hearts*.

There is, however, a second issue. Recall that we reasoned that, if ab/normality inclusive categories are functional categories, the concept of normal function is implicated. Such categories require an extrinsic concept of normal function because they co-classify items with radically different dispositions. Membership in an ab/normality inclusive category may not guarantee any particular intrinsic properties. It certainly does not guarantee possession of the relevant defining disposition/s –i.e., the disposition/s for which the phenotype was selected. So the *prima facie* problem with the category of hearts is not merely that it excludes items that have the same disposition as its members (hearts* that are not hearts). The problem is also that it includes items with radically different dispositions (hearts that are not hearts*). For the purpose of making precise predictions, therefore, functional categories seem, in this respect, to be *not as fine-grained as they need to be*. Extrinsic categories like this seem to sacrifice predictive precision. For predictive purposes, the category of hearts is nowhere near as accurate as the category of hearts*. Again, there is a parallel issue with the classification of psychological states according to normative content. Knowing that an inner representation has a certain content may tell us something about what causes and effects it is supposed to have but this may or may not correspond to the causes and effects it actually has. Once again we are co-classifying items that may have radically different dispositions.

These considerations have led some commentators to conclude that such categories are not good categories from a scientific perspective. They have led others to conclude

that in so far as biology or psychology employs such categories it cannot be aiming at prediction.²⁵ Neither of these opinions is correct in my view, but it will take a little time to explain why. As I have just been saying, the category of hearts excludes some things that the category of hearts* includes. It excludes Swampman's heart*, for example. But the category of hearts also includes some things that the category of hearts* excludes. It includes severely malfunctioning hearts, for example. So there is a trade-off in membership here. Swap the category of hearts for the category of hearts* and we rule in Swampman's heart* but at the same time we rule out the hearts of all of the real men and women and other creatures who are suffering from heart failure. The key to understanding the role of these ab/normality inclusive categories lies here. We need to ask why we want to co-classify abnormal items with their normal counterparts. In relation to the biological case, I have three quick points to make in relation to this and another point that will emerge toward the end of the next section.

First, there is no real loss of predictive power. Although ab/normality inclusive categories embrace traits with very different causal dispositions, we can and do use them to generate more fine-grained categories that are of more predictive value. Knowing that something is a heart does not tell us much about what its actual causal capacities are. But knowing that it is a normal heart or that there is ventricular fibrillation (a rapid and fluttering beat) or coronary artery occlusion (blockage) or stenosis (partial blockage) or myocardial infarction (muscle cell death) tells us much more. We can therefore co-classify malfunctioning items with normal items of the type at one level of abstraction, while still enjoying the benefits of categories that distinguish them at another level of abstraction. Indeed, we can do more than this. We can also distinguish very many

different kinds of abnormality. It is by means of such quite fine-grained classifications of syndromes or pathologies that medical research has what degree of predictive power it has.

Second, classifying according to normal function gives stability of classification over the lifetime of a trait. The traits of organisms can malfunction but they can also regain their capacity to perform their function, either in the natural course of events as a result of the body's own mechanisms of repair, or as a result of medical intervention. A token trait could gain and lose the capacity to perform its function many times over. Classification by normal function but not classification by actual causal capacities will allow a token X to continue to count as a token X over time.

Third, if we compare the category of normal hearts with the category of hearts* it is true that the latter has wider application than the former. For something to be a normal heart it must (a) have the normal function of pumping blood and (b) be capable of performing that function. For something to be a heart* it merely needs (b), to be capable of performing that function. So the category of normal hearts excludes some fantastic imaginary cases such as Swampman's heart*, and in that sense it has a narrower application than the category of hearts*. But the category of hearts, as opposed to normal hearts, is not narrower in application than the category of hearts*, not at any rate with respect to items in the real world. It embraces all those malfunctioning hearts that are excluded from the category of hearts*.

Our ab/normality inclusive categories therefore give us greater generality in terms of real world applications and greater stability of classification without any loss in predictive power. It is true that it is not immediately clear what advantage such greater

generality or stability brings, given that these more stable and more general ab/normality inclusive categories do not seem to be ideal for predictive purposes. Even though they do not cost us anything in predictive power, since we can always use other more fine-grained categories for predictive purposes, we can still ask what the point of them is. But perhaps we can uncover some advantages if we examine the role of normal function in explanations of living systems.

(5) Normal Function and Explanation

Functional explanations –by which I here mean explanations that appeal to biological functions— have often been regarded with suspicion. Carl Hempel argued that they are almost always shabby efforts at explanation since they do not fit his formula. And in the years that followed, throughout the fifties, sixties and seventies of the last century, the majority of philosophers of biology argued that explanations that appealed to a teleological concept of function were not scientifically respectable. They were accused of being Aristotelian, of invoking reverse causation, of harking back to Creationist thinking, of being value-laden, or in some other way of not being naturalistic. It is a case of real progress in philosophy, I believe, that we have largely put this worry about naturalism to rest. Of course, there is now concern over the fact that appeals to function are extrinsic.

When we inquire into the explanatory role of functions, the first question to ask is what is being explained. What do we at least purport to explain by referring to normal functions? Is it perhaps the operation of the system in which a trait is embedded? For example, when we are told the function of the primate's optic tectum, is it the operation

of the primate's perceptual system that is being explained? Or is it perhaps the trait itself—the trait to which the function is attributed—that is being explained by the function attribution? For example, when we are told the function of the primate's optic tectum, is it the existence of the optic tectum that is explained? Views on this have differed, with those in favor of the latter opting for a teleological (or etiological) concept of function and those in favor of the latter opting for something more like a Cummins-style or statistical concept of function.²⁶

Roughly speaking, these two alternative explanatory aims correspond to what Ernst Mayr refers to as proximate and ultimate explanations.²⁷ As Mayr says, a question of the form “why does X do Y?” is ambiguous. On the one hand, it might be requesting a description of the mechanism in each individual by which X does Y, and, on the other hand, it might be asking why certain organisms evolved Xs with the capacity to Y. In general, these are seen as quite distinct explanatory enterprises. Explanations of the first—proximal—kind are seen as the task of the physiologist, and explanations of the second—ultimate—kind are seen as the task of the evolutionist. Moreover, the explanatory role of the teleological (or etiological) concept of function—is generally thought to be restricted to explanations of the second—ultimate—kind. For instance, after explaining Mayr's distinction, Elliott Sober continues:

Nonetheless there is a division of labor between the physiologist and the evolutionist. Each answers one but not the other of the following two questions: (1) What mechanism inside ivy plants cause them to grow toward the light? (2) Why do ivy plants contain mechanisms that cause them to grow toward the light? Question (1) calls for details about *structure*; question (2) naturally leads one to consider issues pertaining to *function*. (1993, 8)²⁸

Perhaps most philosophers would now agree that normal functions, understood according to an etiological account, provide ultimate explanations. According to an

etiological account, when we are told the function of a thing, we are told what the relevant phenotype was selected for. Normal functions are not, of course, the immediate causes of the traits whose functions they are. They are not the causes, immediate or otherwise, of anything. But function attributions provide ultimate explanations in the sense that we can unpack them to learn the remote causes of the traits concerned. When we learn a/the function of a trait, we acquire an elliptical sketch of some of the remote causes of the persistence of the phenotype. Suppose, for instance, that we learn that it is the function of the primate's optic tectum to locate the visual signal source. Then we learn that ancestral optic tectums responded in certain systematic ways to the location of the visual signal source, and that they thereby contributed to the viability and fecundity of the primates who possessed them and hence to the selection of the responsible genotypes.

What role, if any, does the concept of normal function play in a proximal (or what I have elsewhere referred to as an operational) explanation of a system? Most philosophers deny that normal function, understood as a teleological and etiological notion, has any role in such explanations. They would agree with Sober that such questions of mechanism are not evolutionary questions at all. Perhaps part of their thinking is that, since the property of having a normal function is not a property that has causal efficacy, it cannot be a cause of anything within an organic system and it can therefore have no role in a causal explanation of its operation. According to Sober, in the passage cited above, descriptions of structure, not function, are required for proximal explanations. However, one popular view in philosophy of biology is that *a* concept of function *is* employed in such explanations, but it is a Cummins-style concept of function not the teleological one. This view is often accompanied by the view that the teleological

concept has a role in evolutionary explanations, and together they go by the name of *conceptual pluralism*.²⁹ According to this view, there are two concepts of function that have distinct explanatory roles. There is the teleological concept, which plays a role in evolutionary explanations of traits but not in operational explanations. And there is a Cummins-style concept of function, which plays the lead role in operational explanations of living systems.

At first glance, this seems like a sensible position, but it seems to me that it has to be an oversimplification. Of course, I agree that both concepts of function have a role in biology. If you think that the teleological concept plays a role, that much is trivial. To deny that there is also a concept of the sort defined by Cummins' analysis would be tantamount to denying that biologists give reductive explanations of living systems and employ a concept of a causal capacity in doing so. Moreover, if one is a supporter of an etiological account of normal function, to admit that biology requires a concept of a normal function is to admit that it also requires a concept of a disposition to contribute to fitness. For, on an etiological account, normal functions are dispositions that were selected because they contributed to fitness.

What I reject is the claim that teleological functions play no role in physiology. This has to be wrong if teleological functions are normal functions, as physiology is replete with references –explicit and implicit—to normal functions.³⁰ We have to ask what such references are doing there. It is not enough to suggest that normal function must play a classificatory role. For one thing, the concept is used explicitly, in a manner that appears to be in addition to any classificatory role it may have. Besides, it would be puzzling if it played an important classificatory role if the resulting classifications did not in turn play

an explanatory role. In what follows I will argue that normal function plays an important role in operational explanations of living systems. It plays an important *idealizing* role, as I will explain.³¹ To see this, we need to consider the challenges that the physiologists face in constructing operational explanations. I will begin with some reasons why physiology requires functional categories of some kind, and then give some reasons why it requires externalist, ab/normality inclusive categories.

Consider a single living thing, such as you or me. We are staggeringly complex organic systems. Each cell is vastly complex in itself, and each of us is composed of about 200 different kinds of cells (muscle cells, skin cells, nerve cells, etc.) in a complex arrangement. The human brain, for instance, has about 12 billion neurons with synaptic connections perhaps numbering about 12,000 billion. Of course, we believe that this complexity is the result of natural selection, but it does not follow from the fact that such systems result from natural selection that we will need to refer to natural selection to explain their operation. However, when a system is so complex, a reductive explanation is called for, and a reductive explanation will require functional categories of some kind. Or, more weakly perhaps, it will at least require that functional considerations play a leading role in determining which components of the system we choose to identify. We must conceptually decompose the system into parts, and these parts into their parts, and so on. There are an infinite number of ways to arbitrarily decompose an organism into its parts but the physiologist needs to do so in a way that is useful for the task at hand. Since she wants to understand how the system works, she needs to decompose the system into parts along functionally significant lines³². Since the heart as a whole works as a whole to pump blood around the body, it is a useful to identify it as a component. A

component composed of a small part of the heart, a bit of the nearest rib and some of the breast tissue beyond will not be so useful for explaining how the system works, and nor will an infinite number of other alternatives.

Notice that this need for decomposition along functionally significant lines remains, whether or not the instances so identified are multiply realizable. There would be such a need even if we were dealing with just one organism as it presented itself at just one time, and hence with only one realization. Even if the entire resources of physiology were devoted to the sole task of explaining your physiology as it is this moment it would still be necessary to conceptually decompose you into parts that are illuminating with respect to your functional organization. The sheer complexity of the system demands a hierarchical analysis and a decomposition into parts that is sensitive to what's done, what does it, and how it does it.

There are two ways in which structural heterogeneity can be an issue. One is the standard and much discussed way in which the same function might be performed by mechanisms that vary in their structural design. This is what is ordinarily referred to as 'multiple physical realizability'. But there is also another kind of 'heterogeneous realization'. In this case, a component of a given type is structurally heterogeneous, not in the sense that items of the type come in different designs, but in the sense that the design of each token of the type is complex, for each token consists of structurally heterogeneous sub-components. In large part, such complex components are "meaningful" components, not because they are tied together with something like structural string (e.g., a cell membrane) but instead because they are functionally coherent units. The various heterogeneous parts may collaborate closely in the

performance of some overall function or small set of functions. We have already looked at the example of the heart. At the lower end of the hierarchical analysis, consider genes. Genes are individuated in part in structural terms on the basis of being contiguous base pairs on a chromosome, but they are also individuated in part on the basis of the proteins that they specify, which is a functional consideration. At a far higher level of analysis, consider the immune system. It too is in part individuated with respect to the overall function to which its various parts contribute, i.e., immunity.³³

For all that has been said so far, a simple notion of a causal capacity or a causal disposition would surely suffice. But the physiologists' task is of course much more demanding than this. Their task is not to describe just one individual as it is at a given time, hard enough as that might be. Their task is to describe a colossal diversity of living forms. There are 2 or 3 million species of plants or animals at present, descended from only a small minority of those that have existed in the past. Moreover, there is a multitude of instances of each kind. There are about six billion human beings, just counting those alive today, and we are hardly the most populous species. I have no idea how many sugar ants or houseflies or mice or sparrows or daisies there might be, let alone how many bacteria of any given strain, but precise numbers are irrelevant. They are enormous and the question is how are the physiologists to describe even a minute fraction of so many living systems?

That there are so many instances of each kind of living system would not add to the physiologist's difficulties if each instance of a kind were the same as the others. But this is far from being the case. No two complex organisms are precisely the same. Given our combinatorial DNA-RNA system of inheritance, and the occurrence of mutation and

re-combination, the potential variation is practically unlimited. Further, the development and maintenance of the organism is the product of a tremendously complicated interaction between the genome and an environment that is itself capable of infinite variation. Consider also the unlimited possibilities of pathology—the many different forms of cancer, the range of possible injuries, the countless ways in which we could be ravaged by viruses and bacteria, which are also evolving forms of life. Add to this the fact that each individual is not a fixed product of all these sources of variation, but is an ongoing developing and eventually deteriorating concern.

The problem is a little like that of the car manufacturer faced with the task of providing an owner's manual for all the cars of a certain model on the road, only the difficulties are vastly magnified. Cars are extremely simple as compared to organisms: they have parts but their parts are not composed of micro-machines. Each car of the same model at least begins from the same blue print, under fairly uniform factory conditions, and is built from basically the same material supplies. However, even here, individual cars will differ as they roll off the assembly line and differences will tend to accumulate as they are driven about, are neglected or cherished, are worked on by different mechanics, suffer minor bumps and major crashes, etc. Even in the case of the car manufacturer, it is not practical to attempt to provide a manual that accurately captures all of the idiosyncrasies of each car as it leaves the assembly line, let alone as it changes over its "life-time". Some differences can be described. For instance, one can buy the same basic model with different accessories and these different accessories can be described. Or the engineers might describe common faults that develop over time. But there are practical limits to what can be done along these lines and the solution, of

course, is to provide a description of the car's design, a description that abstracts away from harmless idiosyncrasies, manufacturing faults, developing mechanical flaws and wear and tear and dirt and rust. The car manufacturer leaves each individual owner and his mechanic to figure out in what ways his car approximates or fails to approximate to that design.

The physiologist can and does investigate the actual structures and the actual causal capacities of the components of individual living systems. But no such description of any given individual will accurately describe all of the corresponding actual structures and actual causal capacities of all of the components every other individual of that kind. Since describing the actual structures and the actual causal capacities of the components of every individual is out of the question, the task is to provide a representative description or a few representative descriptions. How is this best done? Physiologists could arbitrarily select an individual, on the principle that any given individual is as representative as any other individual. They might select Jane from off the street and start analyzing her physiology. They might describe the ulcer she's had since teaching in Zimbabwe, her asthma in reaction to pollen, the results of her hysterectomy, etceteras. Beyond a certain point, there would be insurmountable practical problems (not to mention moral ones) because the individual chosen would not survive the kind of investigation required. And even if she could, she would have changed before the investigation had barely begun, and be dead of old age (if not illness or injury) long before it finished. But even if this plan were practical it would not be a good plan. For an arbitrarily chosen individual is not the most useful individual to describe. Jane may be about average in terms of her overall health, but a more useful description is a description

of the normal system (or systems) for the kind –i.e., the system (or systems) that is (or are) normal in all respects.

On the understanding that an etiological theory of normal function and a corresponding etiological theory of normal structure is correct, this is a description of a system whose component parts are all such that they have the structures and the capacities for which there has been selection. This system is more representative of the relevant kind because variation tends to converge on normality. Few if any individual organisms, at least few if any multi-cellular organisms, are normal in every respect. Some abnormality is usual. However, different parts are affected and when the same parts are affected they are affected in different ways. In contrast, most individuals are mostly normal most of the time. A description of a normal system therefore has a good chance of accurately describing more of most individuals most of the time than a description of an arbitrary individual of that species does.

This would be trivial were the description of the normal system a description of the typical system. Indeed, it is a good question why, in that case, the normal system should not be the typical system. But note that a statistical concept of normal function is also an extrinsic concept, so the main point would still hold if it were. Note also that, for the most part, what is typical and what is normal correspond. Systems that are not mostly normal in the teleological sense very quickly die and cease adding to the numbers. However, there are reasons for preferring a teleological as opposed to statistical notion of normality for this task of providing a representative description and explanation of the functional organization of a kind of living being. I'll list four.

(1) There is a problem with providing a non-circular definition of the relevant reference class from which the statistics are to be abstracted. It won't do to uniformly use the entire species as the reference class as what is normal functioning for one sex, or age group or population within a species, and so on, can differ. We do not want to count male (or female) reproductive organs as abnormal just because that sex is in the minority in a species. But on what basis do we determine that sex differences are normal and yet that the distinct features of, say, Down's syndrome, are not?

(2) A description of the system that is typical at one time may not remain a description of the system that is typical at a later time. When the typical deviates from what there was selection for, it is liable to comparatively fast change. It can change with demographic fluctuations, climactic variations, environmental changes, technological innovations, the spread and control of diseases, and so on. Of course, the design of a species is far from immutable, but in comparison it generally changes gradually over a grand evolutionary time frame.

(3) Whether a particular individual counts as typical in a certain respect will also depend on which time slice of the population we use. An individual may count as typical in a certain respect relative to the last fifty years, but not relative to the fifty years that follow. It might count as typical relative to the last fifty years but atypical relative to the last one hundred and again as typical relative the last thousand.

(4) Finally, and I think this is the most fundamental consideration, there is an important way in which a description of the normal system is more explanatory than a description of the typical system. The physiologist seeks not merely to describe what happens in a system, she seeks to explain how the system *works*. Descriptions of

breakdowns in the organization of the system are of interest, whether they are typical or not, but such descriptions need to be distinguished from descriptions of how the system operates so as to make survival and reproduction possible.

Mere complexity, in the sense of mere heterogeneity, is not the same as organized complexity. Organized complexity is required for the maintenance of a high degree of heterogeneity, so the two tend to go together. But we must add something to the idea of heterogeneity to capture the idea of organized complexity. In an organized system the diverse parts collaborate, communicate and coordinate their activity. This can, of course, break down, in an infinite number of possible ways. But when the system *works*, in the success sense of that term, its huge cast of diverse characters is intimately and intricately orchestrated in a fairly harmonious collective enterprise. This is orchestrated by natural selection. It is the result of the co-adaptation of each part to all of the other parts in the internal environment of the organism. This is the historical – the evolutionary explanation— of the organization. But to describe what each part was adapted for does more than give us an elliptical sketch of each part's history. When we learn what each part was adapted for, we also learn its causal role within the organized system, within the system that works.

When non-biologists think of the way in which a trait of an organism is adapted, they generally have in mind the way in which it is adapted to the rigors of the organism's external environment. This is an important aspect of its adaptation. If an organism were not well enough adapted to its external environment, no amount of fine internal organization would keep it alive. However, the adaptation of each of the components of a living system to all of the other components of the system is as important as their

collective adaptation to the outside world. To specify what a part was selected for, is to specify its contribution to the system's ability to "make a living" in two senses. First, in the sense of making life possible by sustaining and not disrupting the system's internal organization and, second, in the sense of making life possible by enabling the organism to survive and reproduce within some particular environmental niche.³⁴

Thus, *to describe the normal system, is to explain how the system works*, in the sense in which "works" is a success term. Individuals approximate such a description to the extent that they are normal. To the extent that they are normal, such a description explains how they work and predicts their working. To the extent that an individual deviates from such a description, we have a clue as to why that individual's physiology might not be working. A clue as to how his or her internal organization might be disrupted and as to what kinds of alterations might restore health. To classify the components of individuals according to their normal functions is to make such idealized descriptions and explanations applicable to them. Membership in such classifications locates token traits within such idealized descriptions and explanations. If our hearts failed to count as hearts when they malfunctioned, if in general our Xs failed to count as Xs when they malfunctioned, nothing in the description and explanation of the normal system would appear to apply to them. We would lose our directory to their place in the normal system. In the absence of ab/normality inclusive categories, their classification would not point to what they do when they function normally. It would not point to whatever differences there might be between what they would do were the system working and what they are actually doing that might prevent the system from working. Thus we have further reasons for thinking that,

The study of adaptation is not an optional preoccupation with fascinating fragments of natural history, it is the core of biological study.
(Pittendrigh, 1958, p.395).

And that,

Nothing in biology makes sense except in the light of evolution.
(Dobzhansky, 1973)³⁵

.....

Whether or not there are parallel reasons to be found for classifying mental states according to their contents remains to be seen. However, I do believe we can conclude that content's lack of causal efficacy is no bar to its scientific significance. Moreover, in the light of reasons given in the introduction to this paper, I believe we should look to see whether there are parallels between the theoretical roles of function and content. When we focus on Putnam and Burge inspired twin cases we are apt to ask whether psychology can get by with disjunctive or neutral contents, and whether we should we co-classify mental states that differ in content if they have the same dispositions. But these questions do not go to the heart of the matter. Instead we should be asking why we want to co-classify token representations that have different dispositions, why we want to co-classify correct and incorrect applications as tokens of the same type.

¹ Jaegwon Kim, Mind in a Physical World (MIT Press, Cambridge, Mass, 1998). For elaboration of the three problems, see Kim, (1998) pp. 32-38. I have listed the three problems in a different order to the order in which Kim lists them. Kim sees all three problems as problems for non-reductive physicalism but as I remark below the problem of extrinsic causation has a different source. I am grateful to Crystal L'Hote for pointing this out to me. Note that the notion of reduction used in this paragraph of the main text of this paper is that of classical reduction, or what is often referred to as type-type reduction, e.g., an identity between mental kinds or properties and neurophysiological kinds or properties.

² This problem is Kim's focus in Mind in a Physical World *op. cit.*

³ Stephen Stich, From Folk Psychology to Cognitive Science (Cambridge, MIT Press, 1983).

⁴ Views of this kind have been advanced, for example, by Fodor, J. in "Psychosemantics—Or Where do Truth Conditions Come From?" finally published in Lycan, B. Mind and Cognition: A Reader (Oxford; Blackwell, 1990); Millikan, R., in Language, Thought and Other Biological Categories (Cambridge, MA; MIT Press, 1984) and White Queen Psychology and Other Essays for Alice (Cambridge, MA; MIT Press, 1993); and Dretske, F., in "Misrepresentation", in Radu Bogdan (ed.) Belief (Oxford; Oxford University Press, 1986) and Explaining Behavior: Reasons in a World of Causes (Cambridge, Mass; MIT Press, 1988). Fodor rejects the idea in his (1987) Psychosemantics, The Problem of Meaning in the Philosophy of Mind, (Cambridge Mass; MIT Press). For a response to Fodor's objection on the grounds of functional indeterminacy, see Neander, K., (1995) "Misrepresenting and Malfunctioning", Philosophical Studies 79, pp.109-141.

⁵ See Hilary Putnam, "The Meaning of 'Meaning'", in his (1975) Philosophical Papers, vol. 2 (Cambridge, Cambridge University Press) and Tyler Burge, "Individualism and the Mental", Midwest Studies in Philosophy 4 (1979).

⁶ Rather, this would be true were determinism true. We can ignore this complication since indeterminism does not help the cause of content causation in any way.

⁷ See Stich, *op.cit.* and Fodor, J. (1991) "A Modal Argument for Narrow Content", The Journal of Philosophy, LXXXVIII, no 1, 5-26.

⁸ I think it is Burge who first makes this point.

⁹ I take the following to be in the spirit of Fodor's reply to Burge.

¹⁰ For example, Segal advocates the disjunctive approach in his 1991, "Defence of a Reasonable Individualism" and neutral contents in his (2000) A Slim Book on Narrow Content.

¹¹ We can construct exceptions to this –e.g., "it is now" and "this is anything or nothing", but these depend for their meaning on being complex. It is true that we could create a simple lexical item to stand for them, but the meaning of the simple lexical item would remain molecular, i.e., derived from the meanings of defining terms.

¹² Not everyone agrees. See Michael McDermott's account of narrow content for an exception. It is an interesting attempt to capture the content of narrow belief attributions in terms of expectations about what kind of neural inputs will follow certain neural outputs.

¹³ Cummins's concept of content in his Content, Targets and Attitudes is an exception, but only because he introduces the –I think potentially very useful– notion of

a target of a tokening of a representation. His account of a target is under-specified, but it is clear that the main work of providing for the possibility of misrepresentation is shifted on to this notion.

¹⁴ As Barry Loewer pointed out to me, it is not that *no* concept of content can be intrinsic. For example, here is an intrinsic account: every neural event means, “today is Tuesday”. This guarantees an abundance of error six days a week.

¹⁵ First in a widely circulated but unpublished paper titled “Teleology in Biology”, initially delivered to the New Zealand Division of the Australasian Philosophical Association, 1980. Subsequently in my (1983) Ph.D. dissertation, Abnormal Psychobiology, La Trobe, and my (1991) “Functions as Selected Effects”, Philosophy of Science, 58, pp.168-184.

¹⁶ Boorse raised this problem against his own account. He did not provide a satisfactory solution to it when he did so.

¹⁷ Note that the possibility of ‘reduction’ in this sense does not entail the possibility of reduction in the sense mentioned in footnote 1, above. Reductive analyses of systems in this sense do not require identities between kinds at one level of analysis and kinds at lower levels of analyses.

¹⁸ I have changed the symbols.

¹⁹ This relativization move is made by David Walsh and Andre Ariew, A. in their paper titled, “A Taxonomy of Functions”, reprinted in

²⁰ In fact, he did so in a colloquia discussion at U.C. Davis when I presented a version of my paper, “Types of Traits”, see fn. 11.

²¹ There is a parallel moral to be drawn concerning certain attempts to make content narrow (see e.g., Frank Jackson and David Braddon-Mitchell attempt to provide an account of narrow content) but I won’t try to elaborate on this here.

²² See fn. 15 above. For discussion of these theories, see three recent anthologies: Nature’s Purpose: analyses of function and design in biology ed., by Colin Allen, Marc Becoff, and George Lauder, MIT Press; edited by David Buller (ed) Function, Selection, and Design (The State University of New York Press); and Functions in Philosophy of Biology and Philosophy of Psychology, edited by Andre Ariew, Robert Cummins, Mark Perlman, (Oxford University Press, forthcoming).

²³ See Neander (forthcoming) “Types of Traits: Function, Structure and Homology in the Classification of Traits” in (eds.) Ariew, Cummins & Perlman (forthcoming 2002).

²⁴ See Amundson and Lauder “Functions without Purpose”.

²⁵ See Millikan (1993) “Explanation in Biopsychology”, Mental Causation, edited by J. Heil and A. Mele, p. 222.

²⁶ For example, Larry Wright has argued for the latter, and Cummins for the former.

²⁷ Ernst Mayr, “Cause and Effect in Biology”, Science 134, 1501-1506.

²⁸ Sober, E. (1993) Philosophy of Biology, (Westview Press; Boulder, San Francisco) 6-8.

²⁹ Both Ruth Millikan (19XX) and Peter Godfrey-Smith (19XX) do so, for instance.

³⁰ In one place (in his book, _____) Peter-Godfrey-Smith claims that a Cummins-style concept of function does permit the possibility of malfunction. However, he offers no elaboration. As we saw earlier in this paper, Cummins’ account can easily be developed so that it permits the possibility of malfunction, but then the obvious ways to

do so are by making the norms statistical or teleological. If Godfrey-Smith agrees that the norms are not statistical, which he seems to do elsewhere, it is unclear how he still has two distinct concepts.

³¹ I make this claim in my (1991) “The Teleological Notion of ‘Function’”, Australasian Journal of Philosophy, 69, No. 4, p.467. Millikan says the same in her (1993) “Explanation in Biopsychology”, Mental Causation, edited by J. Heil and A. Mele, 230-231 although this differs from her earlier position.

³² For useful discussion see Cummins, The Nature of Psychological Explanation.

³³ Several of the preceding points are taken verbatim from my “Types of Traits”.

³⁴ Naturally, the focus will shift from trait to trait. To say what the echidnas’ spikes are adapted for requires more of a focus on the relation between the echidna and its environment. To say what estrogen or mRNA or coenzyme Q10 is adapted for requires more of a focus on internal relations within the organism.

³⁵ This is the title of his paper in American Biology Teacher, 35, 125-129.

Dobhansky’s point was not my point, but perhaps he would have been sympathetic.