

# *Donald Davidson's Theory of Mind Is Non-Normative*

*Timothy Schroeder*

*Philosophers' Imprint*  
<[www.philosophersimprint.org/003001/](http://www.philosophersimprint.org/003001/)>  
*Volume3, No.1*  
*May 2003*  
© 2003 *Timothy Schroeder*

What makes a theory of mind normative? A clear-enough answer would seem to be this: for a theory of mind to be normative is for it to appeal to something normative in explaining the nature of the mind. This is not, however, quite adequate. Getting a better understanding of what makes a theory of mind normative requires a better understanding of norms and, in particular, a distinction between what will be called the *categorisation scheme* of a norm and the norm's *force-maker*. The distinction is a very important one. Once it is drawn, a surprising conclusion follows: the theory of mind developed by Donald Davidson is not a genuinely normative theory.

## *1. Davidson's Theory of Mind*

Donald Davidson's theory of mind is well known, and my summary will be correspondingly brief. Davidson takes the view that individual mental events are identical to individual physical events. Davidson is thus a materialist and a monist. He holds that mental event *types* are not identical to physical event *types*, however, and so does not look to neuroscience to individuate mental events. Rather, the mental events an organism enjoys are those it is best interpreted as having, where the best interpretation invokes an open-ended set of principles for interpreting individual behaviour, the most important for present purposes being an assumption that the organism is by and large *rational*. Since mental types are individuated via principles of interpretation, they cannot be expected to map in any regular way onto physical types, and since causal relations depend entirely upon the physical type of an event, mental types do not enter into causal laws. Mental events, on such a view, are said to be anomalous. Davidson is thus an *anomalous monist*.

Normativity is said to enter Davidson's theory of mind through his interpretationism. Which propositional attitudes

*Timothy Schroeder is Assistant Professor of Philosophy at the University of Manitoba.*

an organism has depends upon what the best interpretation of it is, the best interpretation being one that is constrained to make the organism “like ourselves” (Davidson 1980, 239) in various respects. One central feature of ourselves, and any creature we can interpret as like ourselves in having a mind, is that we are largely rational. Thus, which propositional attitudes an organism has depends, in part, on which it would have were it largely rational. Since rationality is an evaluative standard, Davidson’s theory of mind invokes a norm. Since the invocation of the norm is not theoretically superfluous, but rather central, Davidson’s theory is a normative theory of mind—or so it would seem.

The appearance of normativity in Davidson’s theory of mind is not an accident. Davidson is not particularly shy about holding his theory of mind to be normative (see his 1985 especially), or chiding theories of the mind for lacking normativity (1980, 241). He calls rationality a “constitutive ideal” (1980, 223), the “standard” beliefs must fit (1980, 239), and failures of it “blunder,” “error,” and “bad thinking” (1980, 221). He remarks on “good reasons” (1980, 233) as being an essential ingredient of action explanation, along with the notion of a “good argument” (1980, 241). And while not every philosopher interested in Davidson has picked up on this strand in his thinking, a number have. Some recent examples include Hornsby (1997), who takes the normativity of a Davidsonian theory of mind to be some evidence that the mind cannot be “naturalised” (1997, 87); Hurly (1998), who frequently alludes to the normativity of the mental within a Davidsonian approach; and Gampel (1997), who worries that Davidson’s notion of normativity is insufficiently clear.

This is all something of a confusion. In the most straightforward sense of what it is to be a normative theory, Davidson’s theory of mind is not genuinely normative at all. To see this, it will help first to say something about norms in general.

## 2. Norms

Some philosophers of mind are put off by the word ‘norm’ but are comfortable with ‘function’ (in the sense that admits of malfunction), ‘rule’, or the expression ‘supposed to’. I will here use ‘norm’ in what I take to be the most generic and therefore innocuous sense possible, meaning something like ‘evaluative standard’. Biological functions, artifact functions, and rules are all types of evaluative standards, and so (on the terminology I will use) types of norms. Other types of norms include the systems of etiquette found in various cultures, and professional codes of conduct. Norms in this sense are diverse, evaluating everything from toasters to kidneys to beliefs. Some norms appear to be created by mere convention, others seem independent of human efforts. Some norms are action-guiding, others apply to objects incapable of actions.

As a rough beginning, the simplest norms may be thought of as dividing up domains into mutually exclusive and jointly exhaustive categories. Etiquette, for instance, divides actions into those which are polite, those which are impolite, and those which are neither. The function of the heart divides possible heart-contractions into those in accord with the function and those not in accord with the function. Norms for student papers divide them into a discrete range of grades. And so on. In dividing a particular domain into categories, norms involve sets of objects, actions, or events. Politeness involves roughly three sets of actions: the set of polite actions, the set of impolite actions, and the set of actions which are neither. Collectively, these three sets exhaust all possible actions, actions being the domain of evaluation for etiquette. The three sets may thus be thought of as collected together by the norm to form a categorisation scheme, and to a first approximation there is nothing more to etiquette than the existence of this scheme.

This rough beginning is obviously rough, however. It is a familiar fact about etiquette, for example, that there are

certain practices that are polite in some societies but rude in others. In some societies, the polite person eats loudly with an open mouth; in others, this sort of behaviour is rude. We can readily imagine two competing systems of etiquette in which every specific way of eating, greeting friends, passing the time with strangers, and so on that is rude in one is polite in the other, and vice versa, while both are silent about the same things.<sup>1</sup> These two systems would divide exactly the same domain (human actions) into exactly the same groupings. In each one, belching in public would be found in the same category as failing to thank a host and ignoring a friend in an attempt to impress a new acquaintance. Likewise, both systems would co-categorise writing thank-you notes and saying 'please'. Both systems would take an interest in the same actions, and both would distinguish the same three sets of actions, yet there would be an important difference between the two, a difference that would not lie in how the systems divide up their domain. Norms involve categorisation schemes, but they also involve something more.

The same conclusion is suggested by considering what it takes for certain norms to exist. If I use my shoe as a door-stop, then while I am so using it, the shoe is supposed to hold open the door: a norm has been created. This norm did not exist until I began using the shoe in just this way. Likewise, when I finish using my shoe as a doorstop, it no longer retains that function: the norm that governed it no longer

<sup>1</sup> According to Foot (1978, Ch.7), there are some practices which are rude in any society, such as "when one man pushes another out of the way" (1978, 102). But there are societies in which this sort of behaviour is conventionally acceptable: Icelandic patrons of movie theatres, in my experience, struggle to gain the best seats with substantial pushing and shoving. It seems better to say that what is rude is a matter of convention. Foot suggests that rude behaviour is simply the kind of behaviour that "causes offence by indicating lack of respect" (1978, 102) and this may well be right, but what offends and what indicates respect is a matter conditioned by culture, and this is all I require to make my point.

exists. The two sets of event-types of interest here, door-held-open-by-shoe events and door-not-held-open-by-shoe events, do not come into being and then go out of existence, however. The categorisation scheme is stable over time (in whatever sense such abstracta can be said to persist), but the norm is not. Again, we are led to the conclusion that norms are distinct from the mere existence of the categorisation scheme having normative significance. Something more seems needed.

The "something more" is what makes a mere categorisation scheme into a norm. It is what takes one category and makes it true that it is the *good, to be preferred, correct*, or otherwise normatively positive category, and makes it true that the other is the *bad, to be avoided, incorrect*, or otherwise normatively negative category.<sup>2</sup> This something more is what explains why eating this way, greeting that way, etc. in one society could be polite while the same actions<sup>3</sup> could, in another society, be impolite. Likewise, this something more is what explains how holding open a door or not could go from being a normatively insignificant event to a normatively significant one. Call this "something more" the *force-maker*. The force-maker for a normative system is whatever creates the distinction between a mere categorisation scheme and a normative system: it puts the *normative force* into the categories.<sup>4</sup>

<sup>2</sup> In normative systems which are not simple binary systems, this "something more" is what gives each of the categories its own particular normative status, making it true (e.g.) that a grade of A is better than a grade of B, which in turn is better than a C, and so on; or, if the evaluation comes on a continuum, it makes it true that it is better to be further along one side of the continuum, and worse to be further along the other side.

<sup>3</sup> That is, actions of the same physical type.

<sup>4</sup> Allan Gibbard (1990, 112-17) makes a related distinction between the (non-normative) states of affairs represented by thick normative judgements and the normative governance of behaviour loaded into such judgements. My own distinction is not completely equivalent, however. In addition to the more superficial distinction between my emphasis on norms versus Gibbard's on normative judgements, there is a more important difference, namely, that I do not presup-

To name something is not yet to explain it. What, exactly, is a force-maker? What can take the several sets of objects making up a categorisation scheme and give them a particular valence, turn the scheme into something with normative significance? I have no general answer to offer, but a certain amount of clarification is available for simple cases. Consider intentional design: if one thing is designed and built with the intention that it do something, then it becomes true that the object is *supposed* to do that thing. Possible doings and non-doings of the thing by the object existed (in whatever sense such possibilities exist) before the object was built, and continue to exist whether or not it performs its function. However, because the object has been designed and built for a purpose, the event types now take on a special significance. Holding open a door (if the object is a doorstop) now becomes a function-fulfilling event.<sup>5</sup> Something about intentional design thus takes a pre-existing categorisation scheme and gives it normative force in a particular case, at a particular time, for a particular domain (in the case just imagined, a single object). Perhaps it is the fact that, in designing an object, one *governs* or *regulates* the object through designing it in a certain way, and this governance or regulation is *aimed* at some particular *end*. In this way, one set of possibilities becomes privileged (that at which the governance aims), and this puts the force behind the categorisation scheme. This speculation need not be right to make the basic point: intentional design is a potential

---

pose Gibbard's account of normativity. My account of force-makers permits Gibbard's account of normativity, but it also permits many other conflicting accounts.

<sup>5</sup> I skip over certain complications here. For instance, a doorstop does not have the function of holding open doors all the time, but only when it has been initially wedged under a door under conditions within those intended by the designer.

force-maker, able to take categorisation schemes with no intrinsic normative significance and give them such significance in a certain situation.

Intentional design is not the only familiar sort of process which creates norms. Widespread institutions of social control, for instance, would appear to be the force-maker backing the normativity of etiquette. It is the fact that parents encourage their children to perform actions of one type (including saying 'please') and not of another type (including belching in public), and that adults likewise apply more and less subtle forms of pressure to the same ends on one another, that appears to make it true that actions of the one type are required, and of the other type are forbidden, by etiquette.<sup>6</sup>

Beyond the examples provided by design and etiquette, at least one other clear example of a force-maker may also be discerned: the process of natural selection has famously been held to be the force-maker for biological functions (see, e.g., Ayala 1970; Godfrey-Smith 1994). It is worth noting that, insofar as natural selection is a design-like process, it is one which governs which individuals pass on their genes in a fashion that directs a population toward better adaptation to the immediate environment; natural selection may thus be analogous to processes of intentional design, and so it is perhaps not so mysterious that natural selection should be a force-maker as well.<sup>7</sup>

These three illustrations of force-makers do no more

<sup>6</sup> We might speculate that norms are created through social control because these sorts of social interactions are processes of governance or regulation, aimed at particular ends. That is, we might speculate that there is a deep commonality between what makes intentional design a force-maker and what makes social control a force-maker. But at present, I can do no more than offer this as speculation.

<sup>7</sup> It is also worth suggesting that a central project of meta-ethics may be thought of as determining the force-maker for morality (if it has one at all).

than scratch the surface of explaining what it is to be a force-maker. But they make the notion of a force-maker clear enough for present purposes. A force-maker is something quite distinct from a categorisation scheme, putting the normative *oomph* into a scheme. A norm is what results when a force-maker puts its weight behind the set of objects making up one category (set of categories, end of a continuum) and denies it to the set of objects making up some other category in the categorisation scheme.

Can the distinction between categorisation scheme and force-maker be defended in general? Perhaps not. There are some (e.g., McDowell 1998, chapter 10) who would hold that there can be no conceiving of certain normative categorisation schemes, such as that separating courageous people from cowards, without also appreciating their normative significance. If certain categorisation schemes are conceivable only with their canonical normative significance, perhaps this shows that there is no ontological distinction between categorisation scheme and force-maker in the particular case. Even if ontological distinctness could be shown, perhaps it is still true that there is no room to make the distinction between categorisation scheme and force-maker in conception. For the purposes of argument, I will not insist otherwise. It is clear that the distinction applies to etiquette and artifact functions; it is far less clear that the distinction applies to the virtues. It will suffice for my purposes to show that there is a robust distinction between categorisation scheme and force-maker in the case of rationality, the norm of most relevance to Davidsonian theories of mind.

Consider, then, the norm (or norms) of rationality.<sup>8</sup> Rational beings, according to Davidson, are consistent in their

<sup>8</sup> By writing 'norm (or norms)' I mean to avoid choosing between the view that there is one norm which governs belief, action, intention, desire, emotions, and the like, and the view that there is one norm for belief (epistemic rationality), another for action (practical rationality), and so on. Such a choice is not relevant for present purposes.

beliefs, and their actions are at least minimally coherent with their beliefs and desires in that they have rationalisations. Is there room to draw a robust distinction here between the categorisation scheme of rationality and its force-maker? There is. Logical consistency is not, of itself, normative: things which can be categorised as consistent or not are not always subject to *any* force-maker, much less the specific force-maker for rationality. Consistency is sometimes normatively required, but not always. The first full statement on p.33 of *Word and Object* may be consistent with the last full statement on p.88 of *On the Plurality of Worlds*, but this is a fact (if it is one) about consistency with no normative significance. Nothing requires the two statements to be consistent, or requires that they not be consistent, for that matter. Likewise, coherence between beliefs, desires, and actions is sometimes normatively required, but not always. My action need not cohere with your beliefs and desires, for instance: there is no normative failure if I act in a manner you could not rationalise. Neither mere consistency nor mere coherence is normatively significant on its own. Sometimes consistency and coherence are backed by a force-maker, and so required, but not always. Hence there is room, both conceptually and ontologically, to distinguish between the categorisation scheme for rationality and its force-maker.

Similar reasoning applies to other, related normative concepts used by Davidson, such as that of correctness. We must interpret others' beliefs as by and large correct, according to Davidson, but correctness of belief can be decomposed into two components: the beliefs being true, and the truth of the belief being in some way good or appropriate. That is, we can again distinguish categorisation scheme and force-maker. Beliefs can be seen as true or false, and then again they can be seen as correct or erroneous, and these are quite distinct matters. To see that truth and falsity make no normative contribution on their own, consider sentences in works of fiction. Works of fiction are full of both

truths and falsehoods, but these true and false sentences are rarely either correct or in error. The statements found in works of fiction are not, by and large, in the same domain of evaluation as earnest utterances.<sup>9</sup> Just how to understand all the domains of evaluation to which works of fiction *are* subject is a difficult problem in the philosophy of art, but no deep aesthetic theorising is necessary to agree that when Henry James begins *The American* with "On a brilliant day in May, a gentleman was reclining at his ease on the great circular divan which at that period occupied the centre of the Salon Carré, in the Museum of the Louvre," what he writes may be true or false, but it can neither be correct nor an error. To hold that its truth is required or preferable, in any normative sense, would be absurd. Once again, we find that the normative concepts relevant to Davidsonian interpretation permit one to distinguish between the categorisation scheme required by the concept and the force-maker.

Finally, consider Davidson's suggestion that we must interpret others as lovers "of the good" (1980, 222), at least by our own lights. Once again, a distinction can be made between the categorisation scheme implicit in "the good" and its normative force. For Davidson, interpreting others as lovers of the good *by our own lights* is interpreting them as desiring the sorts of things that we too desire. This matching of the content of desire between two individuals, interpreter and interpreted, is not in itself a normative matter. The mere having of matching contents in matching attitudes is no more normative than the mere having of consistent contents in belief. That both contents be *good*, at least by our lights, is a distinct matter, requiring some force-maker not required by the mere fact of the contents being the *same* (they could be the same even if, in a particular case, they

were desires for what is bad).

Thus, while the distinction between categorisation scheme and force-maker may be controversial in particular cases, it is perfectly straightforward in others, including in the cases of the norms Davidson appeals to in his theory of mind.

### 3. *What Makes A Theory Of Mind Normative?*

With this improved understanding of normativity, return to the question of what makes a theory of mind normative, in the broad sense I have described. I suggest that, for a theory of mind to be normative in this broad sense, it is necessary that it make use of (1) a categorisation scheme for which there is some force-maker, and (2) the fact that the scheme has a force-maker. Admittedly, one could invent many requirements for use of the phrase 'normative theory of mind' that would differ from the one I propose. But the requirement being offered here is, I think, the most natural and the most interesting. If a theory of mind does not fit the first criterion, then it makes no mention of any entity for which there is a force-maker, and so makes no mention of any entity for which notions like 'better' or 'worse' or 'permitted' or 'malformed' are applicable. Such a theory hardly answers to the name of a normative theory of mind: the identity theory would count as a normative theory of mind under these conditions. If a theory of mind does not fit the second criterion, then there is a more subtle but still important respect in which it fails to be normative. There is nothing substantively normative about a theory of mind that appeals to a categorisation scheme that happens to have a force-maker for it, if the existence of the force-maker is theoretically insignificant. To call such a theory normative would be akin to calling functionalism a materialistic theory of mind, just because the objects fitting its categorisation scheme (realising or not realising a certain complex functional organisation) happen to be material objects. But func-

<sup>9</sup> If the reader wishes to hold that no statement in a work of fiction can be an instance of an intellectual error, that is compatible with the present account. All that I wish to point out is that there is a separation between falsities and errors.

tionalism is commonly and properly understood to be neutral (in principle) between materialism and dualism, precisely because there is nothing about realising a functional organisation that requires the realiser to be made of physical matter. Functionalism's truth is indifferent to the truth of materialism, and so it would be a gross distortion of the theory to claim it to be a materialistic theory of mind. Likewise, if a theory of mind makes mention of a categorisation scheme which has a force-maker behind it, but is theoretically indifferent to the existence of that force-maker, then it would be a gross distortion to call that theory normative. Turning again to identity theory, it could be pointed out that the events falling under the categorisation scheme of type-identity theory (the firing of V4 neural columns, perhaps) are subject to biological norms, but since this is no concern of identity theory, it would be a gross distortion to conclude that the identity theory of mind is a normative theory.

The philosophy of mind contains a number of theories that meet these criteria for being normative. Teleosemantic theories of mind talk about neural structures fitting or not fitting certain categorisation schemes (indication, correspondence, etc.) and about neural structures being *required* to fit these schemes, natural selection being the force-maker.<sup>10</sup> Sellarsian theories of mind invoke categorisation schemes for utterances, whose force-makers are understood to be certain sorts of social practice.<sup>11</sup> Both sorts of theories are, as a result, straightforwardly normative. Davidsonian theories of mind invoke interpretability in the language of rationality, correctness and goodness of desire, but do not seek to explicate the force-maker for behaviour conforming to these norms. This, as we shall see, is no accident, for Davidsonian theories of mind are not genuinely normative theories. While they make use of categorisation schemes that have

force-makers—rationality etc.—they fail the second criterion, for they are insensitive to the fact that rationality etc. *have* force-makers.

#### 4. Davidson's Theory of Mind is Non-Normative

To see that Davidson is interested in rationality etc. only insofar as it provides a categorisation scheme, and therefore has no theoretical use for the fact that rationality is a *norm*, consider a sampling of quotations from his writings on the subject.

[T]he whole set of axioms, laws, or postulates for the measurement of length is partly constitutive of the idea of a system of macroscopic, rigid, physical objects. ... [Likewise,] we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions. ... [W]e make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest. ... [T]he content of a propositional attitude derives from its place in the pattern. (1980, 221)

Just as the satisfaction of the conditions for measuring length or mass may be viewed as constitutive of the range of application of the sciences that employ these measures, so the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of applications of such concepts as those of belief, desire, intention and action. ... [I]f we are intelligibly to attribute attitudes and beliefs, or usefully to describe motions as behaviour, then we are committed to finding, in the pattern of behaviour, belief, and desire, a large degree of rationality and consistency. (1980, 236-7)

Only by studying the *pattern* of assents to sentences can we decide what is meant and what believed. (1984, xvii; italics in original)

<sup>10</sup> See Dretske (1988, 1995), Lycan (1987, 1988), Millikan (1984, 1993), Papineau (1987), Sterelny (1990), and others.

<sup>11</sup> See Sellars (1963), Brandom (1994).

[Radical interpretation] is intended to solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning. This is accomplished by assigning truth conditions to alien sentences that make native speakers right when plausibly possible. ... What justifies this procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement. ... If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything. (1984, 137)

In these passages, Davidson makes clear his theoretical interest in norms such as rationality.<sup>12</sup> Davidson holds that just as one can understand an object as having a certain length only if that makes it shorter than some objects and longer than others according to a certain pattern (a pattern created by the transitivity of the relation *longer-than*), so one can understand an organism as having a certain belief only if it also has certain other beliefs, desires, etc., according to a certain pattern (that instantiated by predominantly rational creatures). His interest in rationality is thus an interest in it only insofar as it picks out a certain set of propositional-attitude clusters (those which it would be fairly rational to hold) and distinguishes them from a different set of propositional-attitude clusters (those which it would be wildly irrational to hold). The fact that the patterns exhibited by the propositional attitudes of a rational organism are normatively commanded—that there exists a force-maker for the patterns—is of no significance in Davidson's theory. That is,

<sup>12</sup> Davidson (1985) is another essay in which it is especially clear that Davidson is interested in rationality only insofar as rationality is a matter of internal consistency (itself a non-normative property).

Davidson holds that it is in the very nature of the propositional attitudes that any collection of them belonging to a particular creature will be a collection with a characteristic pattern of a certain sort—if there is a belief that *P*, there will almost never be a belief that not-*P*, and so on—just as any collection of comparative lengths must have an abstract structure of a sort—it must exhibit transitivity. Whether or not length measurements are normatively required to exhibit transitivity is not a serious question; neither, in Davidson's theory of mind, is the question whether propositional attitudes are normatively required to be found in their own characteristic pattern. The fact that there *is* a normative requirement that propositional attitudes exhibit the sort of characteristic structure of interest to Davidson is not itself of theoretical interest.

Likewise, when it is asked whether Davidson requires that beliefs be largely *true* (as opposed to false) or whether he requires that they must largely be *correct* (as opposed to mistaken), the answer best supported by the text is that Davidson is interested in truth, not correctness. He makes much of his claim that “If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything” (1984, 137). Mere falsity of belief (by our own lights) makes interpretation difficult or impossible: the fact that such falsities would involve errors is not a significant extra burden. In the same spirit, it is the fact that others have desires with the same content as our desires, and not desires for what is good, that makes others interpretable to us. Davidson's theory is thus not concerned with the existence of the force-makers for the norms it invokes, and so is not a genuinely normative theory of mind.

It may be easier to see the point with the help of an analogy, a comparison, and an explanation. First, the analogy:

imagine a person who is deeply concerned with social appearances. "What would the neighbours think?" is always the first question to cross his mind when he considers a novel course of action. Imagine further that this person just happens to live in a community with substantial concern for morality; what the neighbours would think depends on how they would morally evaluate the man. If the neighbours hold a true theory of morality, then the man will be concerned with whether his actions are moral, but only in a very limited sense. He will want his actions to be of one type (acts that are condoned by the neighbours, which just happen to be those acts which are moral) and not of another type (acts not condoned by the neighbours, these also being the acts that are immoral). He will not, however, care that these acts are required by the force-maker for morality; it is enough for him that his neighbours approve of these actions.<sup>13</sup> Hence, while he will monitor whether or not his actions fall within a certain category, that happening to be the category which morality commends, he will not do so because morality commends such actions. This man's concern for morality is akin to Davidson's theoretical interest in rationality. Davidson's theory of mind takes an interest in the categorisation scheme of rationality, but not in the fact that such a scheme is within the scope of some force-maker.

Second, the comparison. Consider Fred Dretske's (1986, 1988, 1995) teleosemantic theory of mind, according to which a mental representation is a structure, some state of which has the function of carrying the information that  $P$  for some proposition  $P$ . As a concrete example, suppose that the firing of a certain cluster of neurons in area  $SI$  of the somatosensory cortex has the biological function of carrying the information that there is a touch to the tip of the left index finger. The categorisation scheme made significant by this function divides its elements, firings of that neural clus-

ter, into two sets: the set of those carrying the information that there is a touch to the tip of the left index finger, and the set of those not carrying this information. If Dretske's theory were, like Davidson's, not genuinely normative, then this categorisation scheme would be all that mattered theoretically. Whether one had a mental representation of touch to the left index finger's tip would depend solely upon whether that cluster of neurons in  $SI$  were carrying that information. However, this is not Dretske's theory. The actual theory holds that the relevant fact is not whether or not the neural firings carry the information, but whether the neural firings are *supposed* to carry that information—whether they have that function, whether there is some force-maker behind the categories. Since they do have that function, since there is a force-maker (i.e., natural selection), the firings constitute a mental representation, whether or not they are actually carrying information about the fingertip. The existence of a force-maker is theoretically crucial: if there were no force-maker there would be no function, and if there were no function, no mental representation would be held to exist. Other genuinely normative theories of mind, such as Sellarsian theories, take their force-makers equally seriously. In a truly normative theory of mind, the presence of a force-maker is crucial for establishing the presence of intentionality in a system. Davidson's theory is not of this type.

Third, the explanation.<sup>14</sup> Why does Davidson write of the normativity of the mind when it turns out that he is not interested in normativity *per se* after all? In Davidson's view, a primary project for the philosopher of mind is to produce a theory of mind that allows beings like ourselves, already endowed with minds, to interpret other beings as mental or non-mental beings in a principled manner. Any such interpretative project is bound to need interpretative

<sup>13</sup> I here assume that the force-maker for morality, if one exists, is not local peer pressure, but this seems a safe enough assumption.

<sup>14</sup> Thanks to the editors of *Philosophers' Imprint* for suggestions at this point.

constraints, for without constraints all sorts of interpretations of all sorts of beings could be advanced: massive indeterminacy of interpretation looms. Davidson finds these interpretative constraints in rationality and the like. By requiring that a good interpretation make a being approximately rational, correct, and a lover of the good, Davidson guarantees that interpretations will be restricted, that it will not be true that anything goes, that every being his theory claims to have a mind will be one that has something recognisable *by us* as a being with a mind. As a theoretical strategy, this is entirely reasonable. However, for the theoretical purpose of constraining interpretation, it is not essential that rationality, correctness of belief, and goodness of desired objects have force-makers: all that is essential is that these features determine non-trivial categorisation schemes. Although Davidson's theory uses normative terms, it is ultimately only concerned with solving the problem of underdetermined interpretation through using ourselves as a default interpretative template, and this is not, ultimately, a use that is concerned with normativity. This is why Davidson's theory is ultimately unconcerned with the truly normative, despite appearances to the contrary.

### 5. *Objections and Responses*

Time for objections. In this section, I briefly consider three possible objections to this attack on the normativity of Davidson's theory of mind.

First objection: Davidson's theory of mind is genuinely normative, because there is a *right* way to interpret organisms and a *wrong* way to do so according to the theory. More subtly, some interpretations are better and some are worse; since the theory of mind is (in part) a theory of interpretation, normativity enters the theory here, at least.

Response: There is a right way and a wrong way to ascribe *any* property. There are better and worse interpretations of radio signals as being caused by quasars, for exam-

ple; it does not follow that there is something especially normative about the domain of study of radio astronomy. Likewise, the fact that there are better and worse interpretations of the propositional attitudes of organisms in no way implies that propositional attitudes themselves are normative entities. All it implies is that theorising (about anything) is a normative enterprise, which is a point in the philosophy of science, perhaps, but not necessarily one of significance to the philosophy of mind.

Second objection: It behoves a being to speak and act rationally, if it is to be interpretable as having a mind. Thus, there is a normative requirement on every interpretable being to be consistent in thought and deed. Perhaps this will suffice to show that Davidson's theory of mind is normative.<sup>15</sup>

Response: The most obvious problem facing this second objection is that the norm appealed to is a means/end norm, and so exists only insofar as a being wishes to be interpretable. A being wishing to be inscrutable can still have a mind, on Davidson's view (so long as it is, in fact, interpretable), but it will not be subject to the norm in question. Hence the norm proposed has little to do with rationality itself, the norm of interest for present purposes.

Third objection: Requiring the Davidsonian to make reference to the force-maker for rationality is, in effect, requiring the theory to explain what reason we have to be rational. This, however, is an unreasonable requirement. Being able say why one should be rational is impossible, for being able to give justifications of any kind presupposes being rational, being in the "space of reasons," as it is sometimes known.<sup>16</sup>

Response: At least two responses are appropriate here. First, I have not gone so far as to ask the Davidsonian for a complete explication of the force-maker for rationality. All I

<sup>15</sup> I owe this second objection and response to an anonymous referee.

<sup>16</sup> I owe this third objection to an anonymous referee.

have held is that a normative theory of mind must not be theoretically indifferent to the existence of force-makers. Davidson's theory of mind does not merely manifest an inarticulacy about the nature of rationality's force-maker (which would certainly be understandable, given how difficult it appears to be to identify the force-maker for rationality), but manifests complete indifference to the existence or non-existence of the force-maker. This is why it is a non-normative theory. Second, I am not convinced that it is any harder to use reason to explain why one is required to be reasonable than it is to use one's mind in order to explain what a mind is, or to engage in other, similarly self-referential tasks. In fact, it is clearly trivial to give one true account of a force-maker for rationality, for plausibly rationality is required by evolutionary design (see, e.g., Millikan 1984). Why should giving a deeper account of the distinguished or special force-maker for rationality (assuming it is not evolutionary design) be so much harder?

### 6. *The Significance of Normativity*

Suppose Davidson's theory of mind is not, in fact, a normative theory. Is this really a blow to it? The glib answer is that it need not be a blow at all. After all, a number of theories of mind are non-normative, and these have many adherents. It may prove, in the end, that no force-maker need be invoked by a complete explication of the mind. On the other hand, many of Davidson's followers take the purported normativity of the theory very seriously, as does Davidson himself. For these philosophers, the discovery that Davidson's theory of mind is not genuinely normative at all may be a rude shock, and a good reason to re-evaluate their commitments to the theory as it stands.

The loss of normativity harms the Davidsonian position in at least two ways. First, the general arguments put forward by other normative philosophers of mind for the conclusion that the mind must essentially involve something

normative, arguments which have served as attacks on functionalism, asymmetric dependency theory, and the like, are now also attacks on the Davidsonian. Consider the Sellarsian thesis that thoughts are "positions in the 'game' of reasoning."<sup>17</sup> Sellars' point might be put as follows: any thought is subject to a norm, is rational or irrational, simply by virtue of being the thought of a reasoning creature. "I ought to buy socks," "my spouse would never abandon me," "I am dreaming"—these are all things that necessarily fall within the scope of the norm of theoretical rationality. Any belief one has is, by the very nature of belief, one for which one *should* have *good* reasons to believe it; there can be no such thing as a thought which is not subject to the laws of reason.<sup>18</sup> The point has struck many as forceful. There is not merely something odd, but something incoherent sounding, in the idea of a belief that is not rationally evaluated. The Davidsonian seems to be able to agree, and say that the reason there cannot be a belief which is not rationally evaluated is that to have a belief is, in part, to be interpretable as largely rational. However, this is misleading. With the distinction between categorisation scheme and force-maker made clear, all the Davidsonian can say is that to have a belief is, in part, to be interpretable as consistent in certain of one's belief contents—to conform to the categorisation scheme of rationality. And consistency, as we have seen, does not entail the existence of norms on its own, much less pick out rationality uniquely. If, as the Sellarsian would have it, to be a belief is, in part, to be in the domain of a certain evaluative standard, then the Davidsonian has no way to accommodate this fact.

In the same vein, consider the conclusion reached by Fred Dretske, that to be a belief is to be subject to the possibility of misrepresenting.<sup>19</sup> What makes something a belief

<sup>17</sup> Sellars (1963, 324).

<sup>18</sup> Idle fantasies, of course, are another matter.

<sup>19</sup> Dretske (1986; 1988, Ch. 3).

is, in part, its susceptibility to misrepresenting the world. Misrepresenting is something only things with functions can do, Dretske purports to show, and to have a function is to be subject to a norm. Thus, to be a belief is, in part, to be subject to a norm, and Dretske reaches the same conclusion as Sellars, though by a substantially different route. Again, it is a conclusion the Davidsonian has imagined her theory could accommodate; again, the Davidsonian has been mistaken. The possibility that a belief be false is something essential to the Davidsonian picture of the mind, but falsity is not, on its own, the same as error or misrepresentation; some force-maker must be introduced for something false to be a mistake, and the Davidsonian introduces no force-maker into her theory of mind.

These quick arguments for the normativity of the mind are not intended as conclusive. The point has been simply that the Davidsonian, having thought herself a normative theorist of mind, is likely to have had a certain sympathy for these sorts of arguments. The discovery that these arguments are now hostile to her preferred theory of mind, rather than supportive of it, must give pause.

The second way in which the loss of normativity harms the Davidsonian position is by raising an awkward question which was not previously posed. A Davidsonian theory of mind holds that to have a mind requires, in part, that one's propositional attitudes fit a certain sort of pattern. It just so happens, though it is no part of the theory that this should be so, that the patterns one's propositional attitudes must have in order that one have a mind are patterns backed by a force-maker: the force-maker for rationality. This seems a remarkable coincidence. Of course, Davidson himself does not see it as a coincidence, but the argument of this paper has been that, so long as Davidson's theory does not concern itself with the existence of a force-maker for rationality, the fact that the categorisation scheme of interest to Davidsonians is one which is also subject to a force-maker, and

hence is an evaluative scheme, is (theoretically speaking) purely coincidental. The awkward question, then, is this: why does it happen to be the case that the patterns relevant to the theory of mind are backed by a force-maker? The patterns relevant to other non-normative theories of mind, such as those advanced by functionalists or asymmetric-dependency theorists, do not happen to be backed by force-makers. Why does the Davidsonian theory of mind differ?

My point, in raising this question, is not that no answer is possible for the Davidsonian, but rather that the question is hard to answer yet difficult to brush off. A perhaps cynical thought which comes to mind is that the coincidence exists because a Davidsonian theory of mind needs to incorporate the normative aspect of rationality, and Davidson was at least partly sensitive to this need but (subtly) failed to meet the theoretical need when composing the theory, hence the coincidence. If the Davidsonian wishes to retain his theory without modification, something must be done to ward off this cynical assessment.

It would be easy to dismiss the coincidence if the normativity of rationality were purely a matter of social pressure. If that were the case, then one might postulate that people (inchoately) grasped the essence of the mind, and then through social interaction created norms matching the essence, commanding that the patterns central to mentality actually be instantiated. It would then be as if health required certain sorts of hygienic practices, and then etiquette came to command that these practices be carried out—the coincidence of the natural pattern in the world and the normatively commanded pattern would be explained by the norm being developed by people to fit the pre-existing natural pattern. This is not a likely account of rationality, however, for there is nothing conventional about whether one is intelligent or foolish, prudent or rash. Convention might praise a person who rejects *modus ponens*, but the person remains negatively evaluated by rationality for all that.

A more sophisticated attempt to dismiss the coincidence might point out that the coincidence is natural given the twin facts that Davidsonian interpretation requires that we interpret beings with minds as being like ourselves, and that beings like ourselves are by and large rational. We are also *required* to be rational, and this is of no ultimate interest to the Davidsonian, true. But since we are *in fact* largely rational, and since good interpretation requires thinking of others as largely the way we in fact are, the coincidence is only to be expected. This more sophisticated response is still not wholly satisfactory, however. For one might point out that various forms of functionalism also require that we interpret others as largely like ourselves. In, say, Lewis' (1972) familiar formulation, we must interpret other creatures as having inner states satisfying at least half of our commonsense platitudes about psychology (derived, naturally enough, from ourselves). This is another way of working out what it takes for others to be enough like ourselves to justify interpreting them as having minds, but it accords no privilege to the patterns of rationality, and in fact implicitly privileges a commonsensical level of irrationality as only to be expected from other creatures with minds. What, then, is specifically superior about Davidson's stress on the patterns of rationality, when the normativity of rationality is not appealed to?

As with the previous problem advanced for the Davidsonian, this one is not a knock-down argument that something is wrong with the Davidsonian's theory. The Davidsonian can, without inconsistency, hold that it just happens that the pattern of rationality is the best pattern to look for in applying mental concepts, and so the patterns of interest to the philosopher of mind are also subject to a non-conventional force-maker. The reply might stretch credulity somewhat, but formally there is nothing improper. My hope in raising the objection has thus not been to refute the Davidsonian but to inspire the Davidsonian to seek a substantive answer to it, or perhaps to modify his theory.

The direction such a modification would have to take is the positive lesson which may be drawn from the considerations put forward in this paper, a lesson which will be of interest to philosophers wishing to be fully normative Davidsonians and with which I will conclude. Suppose one wishes to hold as closely to a Davidsonian theory of mind as possible but also to hold a genuinely normative theory of mind. At a minimum, such a theory would hold that any set of propositional attitudes must be a set which is *required* to be (roughly?) rational, and that this is an essential feature of the propositional attitudes—nothing not subject to this requirement could be a set of beliefs or desires. If one wishes to hold such a theory of mind, what needs to be done? With the understanding of force-makers acquired earlier, this question can be answered. To hold that the propositional attitudes are essentially required to be rational, one must be ready to give an account of what makes something a propositional attitude that makes mention of some force-maker as a necessary ingredient. The force-maker must apply normative force to the categorisation scheme of rationality, making logically consistent or otherwise rationally related propositional attitudes (belonging to a single organism) good, and logically inconsistent or otherwise irrationally related propositional attitudes (belonging to a single organism) bad. The account of this force-maker need not be reductive, and might even be interpretationistic, but it must be part of any genuinely normative Davidsonian theory of mind.

The upshot of the present work is thus a question for would-be normative Davidsonians: What is the force-maker for rationality? This is a question the answer to which should interest everyone, Davidsonian or not.<sup>20</sup>

<sup>20</sup> Thanks to Nomy Arpaly, Jon Kaplan, the Editors of *Philosophers' Imprint*, and two anonymous referees for helpful comments on earlier drafts of this paper.

**References**

- Ayala, F. 1970. "Teleological explanations in evolutionary biology." *Philosophy of Science* 37, 1-15.
- Brandom, R. 1994. *Making It Explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Davidson, D. 1985. "Incoherence and Irrationality." *Dialectica* 39, 345-354.
- Dretske, F. 1986. "Misrepresentation," in W. Lycan (ed.). 1990. *Mind and Cognition: A reader*. Cambridge, MA: Blackwell Publishers.
- Dretske, F. 1988. *Explaining Behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Gampel, E. 1997. "The Normativity of Meaning." *Philosophical Studies* 86, 221-242.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.
- Hornsby, J. 1997. *Simple Mindedness: In defense of naïve naturalism in the philosophy of mind*. Cambridge, MA: Harvard University Press.
- Hurley, S. 1998. *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- Lewis, D. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50, 249-58.
- Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. 1988. *Judgement and Justification*. Cambridge: Cambridge University Press.
- Millikan, R. 1984. *Language, Thought, and Other Biological Categories: New foundations for realism*. Cambridge, MA:

- MIT Press.
- Papineau, D. 1987. *Reality and Representation*. Oxford: Basil Blackwell.
- Sellars, W. 1963. *Science, Perception, and Reality*. London: Routledge and Kegan Paul.
- Sterelny, K. 1990. *The Representational Mind: An introduction*. Cambridge, MA: Basil Blackwell.