



## Externalism and Norms

[Link to publication record in Manchester Research Explorer](#)

### **Citation for published version (APA):**

Macdonald, C. (1998). Externalism and Norms. In A. O'Hear (Ed.), *Current Issues in the Philosophy of Mind* (pp. 273-301). Cambridge University Press.

### **Published in:**

Current Issues in the Philosophy of Mind

### **Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



## Externalism and Norms\*

CYNTHIA MACDONALD

We think that certain of our mental states represent the world around us, and represent it in determinate ways. My perception that there is salt in the pot before me, for example, represents my immediate environment as containing a certain object, a pot, with a certain kind of substance, salt, in it. My belief that salt dissolves in water represents something in the world around me, namely salt, as having a certain observational property, that of dissolving. But what exactly is the relation between such states and the world beyond the surfaces of our skins? Specifically, what exactly is the relation between the *contents* of those states, and the world beyond our bodies?

I believe that the correct view of the relation between certain mental contents, the contents of at least some of our intentional states, and the world beyond our bodies is an *externalist* one. Crudely, externalism is the view that certain of our intentional states, states such as beliefs and desires, have contents that are *world-involving*.<sup>1</sup> Less crudely, it is the view that certain intentional states of persons, states such as beliefs and desires, have contentful natures that are individuation-dependent on factors beyond their bodies. My belief that salt dissolves in water has a content, *that salt dissolves in water*, that is individuation-dependent on a certain substance in the world beyond my body, namely, salt.

The roots of externalism lie in the work of Hilary Putnam, who was concerned to show something, not specifically about the nature of mental states, but about the nature of meaning.<sup>2</sup> He argued that one's meaning what one does by a natural kind word, although intuitively a state of mind, is world-involving. It is world-involving because it is determined in part by the actual, empirically discoverable nature of something in the world external to one's body. So a person's meaning something by a natural

---

\* I would like to thank Graham Macdonald, Graham Bird, Anthony O'Hear, Michael Martin, Scott Sturgeon, Jan Bransen and Marc Slors for comments and discussion of issues in this paper.

<sup>1</sup> The term is Philip Pettit's and John McDowell's. See their 'Introduction' to *Subject, Thought, and Context*, ed. P. Pettit and J. McDowell (Oxford: Oxford University Press, 1986), pp. 1-15.

<sup>2</sup> See 'The Meaning of "Meaning"', in *Mind, Language, and Reality*, vol. 2 (Cambridge University Press, 1975), pp. 215-71.

kind word cannot be determined independently of that person's relation to the physical world around them.

Putnam reinforced this claim by invoking what is by now the familiar strategy of the twin earth thought experiments. These experiments invite us to suppose that the environments of two individuals might differ in certain ways while all the 'within-the-body' physical and phenomenological (or 'felt') facts about those individuals remain invariant. In that case, Putnam argued, the meanings of the words in those individuals' mouths would also vary: these within-the-body twins would then mean different things by their (indistinguishable) utterances.

Tyler Burge took the moral of the twin earth thought experiments one step further.<sup>3</sup> He argued that since, when a person is sincere, what she says is what she believes, the Putnam conclusion about meaning carries over to intentional states such as beliefs and desires. Burge argued that the twin earth thought experiments not only show that meaning is (partly) an external phenomenon, but that mental states like beliefs and desires, whose contents are typically specified by means of words whose meanings are determined by factors external to persons' bodies, are also partly external phenomena. Just as my twin and I might mean different things by our indistinguishable utterances of 'there is salt in the pot' because of differences in the chemical constitutions of superficially and phenomenologically indistinguishable substances to which we are related in our respective environments, so too might my twin and I think different thoughts when we think thoughts with those propositional contents.

The twin earth thought experiments have been used by Burge and others to support the externalist view that certain intentional states have contentful natures that are individuation-dependent on factors external to the bodies of persons who undergo them. Put like this, it may look as though there is one single, clear formulation of externalism and that there is agreement amongst externalists about what it entails with regard to the existence of objects beyond the bodies of persons who undergo intentional states with representational contents. But this is so far from being the case that part of my aim in this paper is to disentangle some of the different formulations and associated commitments of the view from others, in

---

<sup>3</sup> See, for example, 'Individualism and the Mental', in *Midwest Studies in Philosophy* 4 (Minneapolis: University of Minnesota Press, 1979), pp. 73-121; 'Other Bodies', in *Thought and Object: Essays on Intentionality*, ed. A. Woodfield (Oxford: Oxford University Press, 1982), pp. 97-120, and 'Individualism and Psychology', *The Philosophical Review* 95 (1986), pp. 3-45.

order to fix on what I take to be a central commitment common to all of them and to defend that commitment.

Externalist theses can be strong or weak, and they can be strong or weak in different kinds of ways. However, most theses apply in the first instance to contentful intentional types or kinds, such as the kind, *thinks that salt dissolves in water*. Many thinkers can think thoughts with this content, and when they do they think thoughts that fall under a single contentful kind.<sup>4</sup>

---

<sup>4</sup> See for example, Burge, 'Individualism and Psychology', Jerry Fodor, 'Individualism and Supervenience', *Proceedings of the Aristotelian Society*, Supplementary Volume 60 (1986), pp. 235-62, *Psychosemantics* (Cambridge, MA: MIT Press, 1986), and 'A Modal Argument for Narrow Content', *Journal of Philosophy* 88 (1991), and Gregory McCulloch, *The Mind and Its World* (London: Routledge, 1995). Many who are externalists with regard to contentful intentional kinds also endorse externalism with regard to individual states or events that fall under, or are of those kinds. Tyler Burge is one notable example; he is what might be called a *token* externalist as well as a *type* externalist (see 'Individualism and the Mental', and 'Individualism and Psychology', note 7). Token externalism is the view that the natures of individual intentional mental *events* or *states* are individuation-dependent on factors beyond persons' bodies. They are so because they are individuated by the contentful types or kinds under which they fall, which themselves are individuation-dependent on factors beyond persons' bodies. Since to be a mental event is to be an event of a contentful kind, and since contentful kinds are individuation-dependent on factors external to persons' bodies, mental events are themselves individuation-dependent on factors external to persons' bodies.

Despite this natural association of type with token externalism, it is possible to be a type externalist without embracing token externalism, and *vice versa*. Both of these possibilities have been argued for, and in my opinion both positions are defensible. In particular, the combination of type externalism and token internalism is defensible. Whether one is a token as well as a type externalist depends on whether one thinks that it is of the essence of any mental event which is of a contentful type that it be of a contentful type. This is not a question about the truth of the claim that, necessarily, each event that has intentional content has intentional content. That claim is obviously and uncontroversially true. It is a question, rather, about the truth of the claim that necessarily, each event that has intentional content *necessarily* has intentional content. And this claim is not obviously and uncontroversially true. Whether it is true depends on the truth of other views. For instance, it depends on whether non-reductive physicalism is true and contingent. If it is, then token externalism is false, since non-reductive physicalism is committed to the view that the essences of mental events are physical, not mental. It may be true that mental events, *qua* mental, cannot be individuated independently of the contentful types or kinds under which they fall; but it does not follow that these events cannot be individuated independently of the contentful kinds under which they fall. For that depends on whether these events are essentially mental events.

It is this version of externalism that I wish to concentrate on in the remaining sections of this paper. In section I below, I briefly outline a small number of type or content externalist theses, in order to fix on a core commitment that they share. I then formulate type externalism in these terms. Then, in section II, I focus on a debate between two very well-known adversaries, Tyler Burge and Jerry Fodor. This debate concerns the truth of anti-individualism, which differs from externalism in that it concerns how mental kinds are to be taxonomised *for the purposes of a scientific psychology*. However, the debate is instructive, since it helps to identify the *source* of individualism; of why both externalists and anti-individualists disagree with individualists with regard to the core commitment articulated in section I. Then, in section III, I defend Burge by anchoring the source of type externalism in a very general but distinctive argument, one that relies on the rationalistic normativity of the psychological domain. My defence trades on likenesses between psychological explanation and functional explanation in biology. If the defence succeeds, it succeeds equally for externalism and anti-individualism. Finally, in section IV, I conclude with some remarks about the consequences of this particular form of externalism.

## **I Varieties of Type Externalism**

There is a central claim that almost all of the varieties of content externalism share, which concerns the dependency of contentful kinds on conditions or factors in the environment in which subjects are embedded.<sup>5</sup>

---

<sup>5</sup> McGinn (*Mental Content* (Oxford: Oxford University Press, 1989)) is an exception. He distinguishes between what he calls ‘strong’ and ‘weak’ externalism, and argues for the latter and against the former. By ‘strong’ externalism, McGinn means one which takes content-individuation to require the existence, in the environment in which a thinker is situated, of some object or objects external to the thinker’s body. McGinn rejects this view, but endorses the weaker externalist view that content-individuation requires the existence, in the world of the thinker, of some object or objects beyond that thinker’s body.

This departure from most other forms of externalism means that McGinn is not prepared to rest the truth or falsity of externalism on the existence of twin earth examples. Thus, he says:

it understates the case to express the upshot of twin earth reflections as inconsistent with methodological solipsism, since those reflections imply strong externalism, not just weak. Such understatement can be misleading if it encourages the idea that the inapplicability of twin earth arguments to certain

To see this, we need consider only a few of the ways in which type externalism is typically expressed.

Externalism is often expressed in terms of some kind of supervenience claim regarding the contentful natures of certain intentional types.<sup>6</sup> Broadly speaking, the claim is that such types weakly (in the case of the twin earth thought experiments) or strongly (for thought experiments involving other possible worlds) supervene on factors beyond the bodies of persons.<sup>7</sup> However, this claim can itself be interpreted in a number of ways. The reason is that supervenience is a name for a very general covariance relation, one which states that things cannot differ (or vary) in one respect without differing (or varying) in another, and this covers many different types of case.<sup>8</sup> What is related by supervenience, and how it is

---

cases shows that internalism is true in those cases. You can be a weak externalist about a certain kind of content, and so reject methodological solipsism, and yet deny vehemently that a twin earth case can be given for the content at issue: that is in fact my position about certain kinds of content, as will become apparent. (*Mental Content*, p. 9, n. 13).

If McGinn is right, then the truth or falsity of externalism is not decided by whether twin earth examples exist: although the existence of a twin earth example may be decisive for externalism with regard to certain contents, other contents may be externalistically individuated even when a twin earth example is not forthcoming. Although I do not subscribe to McGinn's brand of externalism, I agree with him that the truth or falsity of the thesis is not anchored in the twin earth examples. However, because his version of externalism departs markedly from most others, I set it aside for present purposes.

<sup>6</sup> See, for example, Martin Davies, 'Aims and Claims of Externalists Arguments', *Philosophical Issues* 4 (1993), pp. 227-249, where externalist theses are explicitly formulated in these terms. Also, see Brian McLaughlin and Michael Tye, 'Externalism, Twin-Earth, and Self-Knowledge', in *Knowing Our Own Minds*, ed. C. Wright, B. Smith, and C. Macdonald (Oxford: Oxford University Press, forthcoming), note 39, where externalism is formulated in terms of supervenience, and Burge, 'Individualism and Psychology', where individualism is formulated in terms of supervenience, externalism being the negation of that thesis.

<sup>7</sup> The Putnam twin earth thought experiments concern weak supervenience, since Putnam envisaged twin earth as being a planet in our own universe, and so in the same possible world. Twins are members of different linguistic communities, but communities within the same possible world.

<sup>8</sup> See Frank Jackson, 'Armchair Metaphysics', in *Philosophy in Mind: The Place of Philosophy in the Study of Mind*, ed. M. Michael and J. O'Leary-Hawthorne (Dordrecht: Kluwer Academic Publishers, 1994), pp. 23-42, who characterizes supervenience in similar terms, as lack of independent variation.

related, can differ greatly from case to case; and the strength of that relation may vary also, in accordance with variation in the objects related and the nature of the relation. All of this will make a difference to how the claim of supervenience is to be understood, and whether it is likely to be true in any particular case. In short, supervenience itself is a name for a class of theses that may concern different objects, different kinds of relations between them, and different strengths of relations, each thesis itself requiring independent explanation and defence.<sup>9</sup>

Given this variety, one cannot expect there to be just one externalist thesis associated with any given claim of supervenience. And indeed there is not. Some have held that externalism commits one to the view that contentful intentional properties, properties associated with contentful kinds such as *thinks that salt dissolves in water*, actually *entail* the existence of objects or kinds of objects in the world beyond the skins of

---

<sup>9</sup> For example, there are supervenience relations between logically or conceptually related properties, such as being coloured and being red, supervenience relations between what we might call ‘metaphysically’ related properties, such as moral or aesthetic properties and psychological ones, or psychological properties and physical ones, and supervenience relations between causally related properties, such as those that figure in causal laws. All of these conform to the formula that is thought to characterize supervenience relations generally, namely, no change in supervenient property without a change in subvenient property. So no psychological change without a physical change, no aesthetic change without a physical change, no change in effect property without a change in cause property. But the relations are really very different in these different types of cases. Although they all involve a relation between properties, they differ in the types of properties related, and they differ in the kind of relation that is thought to hold between them. Other theses differ from these in relating, not properties, but regions of worlds or worlds themselves, or events or states.

Global supervenience claims typically concern worlds or regions of worlds. See, for example, Terence Horgan, ‘Supervenience and Microphysics’, *Pacific Philosophical Quarterly* 63 (1983), pp. 29-43.

Matters are more complicated still, since the strength of the dependency relations associated with these different kinds of supervenience relations also varies considerably. For example, the dependency relation associated with being coloured and being red is said to be logically or conceptually necessary. But this is not so for the relation that is thought to hold between moral and aesthetic properties and psychological properties, or between moral and aesthetic properties and physical properties. Here the relation seems to be weaker than one of logical-cum-conceptual necessity. It seems, rather, to be either metaphysically necessary, where this is understood not to require conceptual necessity, or physically necessary, a necessity that is weaker still, requiring only compatibility with the existing laws of nature in this world.

persons.<sup>10</sup> This, it is said, is because externalism is committed to the claim that it is a *conceptual* truth that, for some propositional content *C* (such as *that salt dissolves in water*), and some proposition, *P*, not knowable *a priori* (such as *salt exists*), if a thinker knows that *C*, then *P*. Thus, for example, externalism is committed to the claim that it is a conceptual truth that if a thinker is thinking that water is transparent, then water exists.<sup>11</sup> If this is so, then it must be conceptually necessary that contentful properties supervene on factors beyond the bodies of subjects that undergo states with those properties.

---

<sup>10</sup> See, for example, Martin Davies, 'Externalism, Architecturalism, and Epistemic Warrant', in *Knowing Our Own Minds*, ed. Wright et al., and B. McLaughlin and M. Tye, 'Externalism, Twin-Earth, and Self-Knowledge', same volume. The twin earth thought experiments are standardly construed as supporting conceptually necessary externalist theses. This is what lies behind arguments of the kind that Michael McKinsey has advanced to show that externalism is incompatible with privileged access, or authoritative self-knowledge. His argument depends on externalism being committed to the claim that it is a conceptual truth that, for some thought content, *C*, which has externalistic individuation conditions (such as *that water is transparent*), it is a conceptual truth that if one is thinking that *C*, then *P*, where *P* is a proposition that cannot be known *a priori* (such as *water exists*). See Michael McKinsey, 'Anti-individualism and Privileged Access', *Analysis* 51 (1991), pp. 9-16. For a reply which denies that externalism is committed to such a claim, see Anthony Brueckner, 'What an Anti-individualist Knows *A priori*', *Analysis* 52 (1992) pp. 111-18. But many externalist theses do not purport to be conceptually necessary. See, for example, Fred Dretske, *Knowledge and the Flow of Information*, (Cambridge, MA: MIT Press, 1980) and *Explaining Behaviour: Reasons in a World of Causes* (Cambridge, MA: MIT Press, 1988), Ruth Millikan, *Language, Thought and Other Biological Categories* (Cambridge, MA: MIT Press, 1984), David Papineau, *Reality and Representation* (Oxford: Basil Blackwell, 1987), and Fodor, *Psychosemantics*. There are differences between the view known as anti-individualism, and externalism. Fodor, for example, explicitly distinguishes the two, and claims that externalism is true, but anti-individualism is not (see 'A Modal Argument for Narrow Content'). Externalism is a view about how the contents of intentional states, states with propositional content, are correctly individuated. Anti-individualism, on the other hand, is a view about how the contents of intentional states are, or should be, individuated *for the purposes of a scientific psychology*, i.e. for the purposes of (causal) explanation in psychology. The distinction between externalism and anti-individualism raises important questions about the nature of psychological explanation and the nature of scientific explanation and taxonomy in general. However, these issues are largely irrelevant to the present discussion, and so the distinction will not play a role in the argument to be developed.

<sup>11</sup> McLaughlin and Tye ('Externalism') have pointed out that no type externalist seems actually to have held a view this strong. Brueckner ('What an Anti-individualist Knows'), in his reply to McKinsey ('Anti-individualism') (whose argument is directed at Burge), points out that Burge (in 'Other Bodies') actually argues against this view.



Others deny that externalism is committed to anything as strong as this claim.<sup>12</sup> Although it requires that the contents of certain intentional states be *object-dependent*, this is not a matter that can be known *a priori*, since one cannot know *a priori* that certain concepts, or propositional contents, are object-dependent. This seems especially plausible in the case of natural kind contents.<sup>13</sup>

Still others claim that externalism commits one to something stronger than a mere claim of object-dependency but weaker than a claim of conceptual entitlement, since it requires dependency on objects with which persons causally interact in their environments.<sup>14</sup> Teleological externalist theses, which require that content supervenes on the causal history of subjects and their interactions with objects in their environments, are theses of this kind. These are very different kinds of theses than either of the two just mentioned, and failure to distinguish them can only lead to confusion about the basic commitments of externalism and about whether externalism is itself a plausible or implausible doctrine.

These people disagree about the strength of the relation between the subvenient and the supervenient in externalist theses. Others disagree about the *sorts* of objects related. Externalists may take their commitment

---

<sup>12</sup> Burge ('Other Bodies') is one. See also Fodor (*Psychosemantics*), Millikan (*Language, Thought, and Other Biological Categories*), Papineau (*Reality and Representation*), and Dretske (*Knowledge and the Flow of Information and Explaining Behaviour*).

<sup>13</sup> The argument is this. Whether a concept is a natural kind concept cannot be known *a priori*, since it cannot be known *a priori* that there are natural kinds (and according to at least one version of externalism there can be no natural kind concepts without natural kinds). This can only be known *a posteriori*, if at all, since whether or not there are natural kinds is an empirical matter. But if it cannot be known *a priori* that the concept of salt is a natural kind concept because it is not knowable *a priori* that there are natural kinds, then it cannot be a conceptual truth that if one is thinking that salt dissolves in water, then salt exists. See Brueckner, 'What an Anti-individualist Knows'.

<sup>14</sup> I am thinking of Millikan, *Language, Thought, and other Biological Categories*, Dretske, *Knowledge and the Flow of Information and Explaining Behaviour*, Jerry Fodor, *Psychosemantics*, and Papineau, *Reality and Representation*. It is difficult to know where to place McGinn (*Mental Content*). On the one hand, he rejects the requirement of causal interaction with instances of the natural kind by individuals who possess concepts of that kind (and in this he commits himself to a thesis weaker than Millikan's and others), and on the other he seems to think that a thinker's thinking such contents *conceptually* entails that objects exist beyond the bodies of subjects who think them. For more on this, see McLaughlin and Tye, 'Externalism'.

to externalism to entail the existence primarily of individual things, corresponding to the contents of singular thoughts such as the thought that Cicero was a Roman orator, or demonstrative thoughts such as the thought that this computer has a coloured monitor.<sup>15</sup> Others may take the view to entail the existence of natural kinds of things, such as tigers, salt and water (corresponding to natural kind thoughts), but not necessarily to any individual instances of such kinds.<sup>16</sup> Others still may take the view to entail the existence of both natural kinds and instances of such kinds with which persons who undergo thoughts with contents that are individuation-dependent on such kinds interact causally.<sup>17</sup> Finally (!), still others may take the view to entail the existence of artefactual kinds, such as sofas and chairs (corresponding to thoughts concerning socially determined kinds).<sup>18</sup> Since these views are compatible with one another, externalists may take the view to commit them to some combination of the above commitments.

Despite all of these differences, however, type or content externalists are united in denying that the contentful nature of any intentional kind supervenes only on factors within the bodies of the subjects that undergo states of that kind. So all forms of externalism are committed to some kind of supervenience claim with regard to certain contentful intentional types. The claim is that certain intentional contents supervene on factors beyond person's bodies, in the sense that subjects' intentional states can vary or change with regard to their contents without varying with regard to all of their intrinsic physical properties.<sup>19</sup> Given the variation amongst externalists in what factors these may be, this claim is best formulated in terms of the negation of an *individualist* supervenience thesis. And since variation in supervenient properties requires variation in subvenient ones, so that sameness with regard to subvenient properties prohibits the

---

<sup>15</sup> See John McDowell, 'On the Sense and Reference of a Proper Name', *Mind* 86 (1977), and Gareth Evans, 'Understanding Demonstratives', in *Meaning and Understanding*, ed. H. Parret and J. Bouveresse (Berlin: W. De Gruyter, 1981), pp. 280-303, and *The Varieties of Reference* (Oxford: Clarendon Press, 1982), chapters 4-8.

<sup>16</sup> See McGinn, *Mental Content*.

<sup>17</sup> See Putnam, 'The Meaning of "Meaning"'.  
<sup>18</sup> See Burge, 'Individualism and the Mental'.

<sup>19</sup> I leave open the issue of whether such variation would entail variation in phenomenological, or 'felt' properties. It may be that variation in factors beyond the body of an individual would affect not only contentful states such as beliefs and thoughts, but also sensation states such as perceptual experiences. This is so, for example, for externalists who think that there is no non-conceptual content (see, for example, John McDowell, *Mind and World* (Oxford: Oxford University Press, 1995)).

possibility of difference with regard to the supervenient ones, we can formulate the negation of that thesis as follows:

1. It is not (conceptually, metaphysically, physically) necessary that, for any two individuals  $x$  and  $y$  and any contentful property  $M$ , if  $x$  and  $y$  are indiscernible with regard to all of their intrinsic physical properties  $P$ , then  $x$  and  $y$  are indiscernible with regard to  $M$ .

Or,

2. It is (conceptually, metaphysically, physically) possible that, for any two individuals  $x$  and  $y$  and any contentful property  $M$ ,  $x$  and  $y$  are indiscernible with regard to all of their intrinsic physical properties  $P$ , but discernible with regard to  $M$ .

What this says is that it is possible for two individuals to be the same with regard to their intrinsic physical properties but different with regard to their contentful mental properties. Short of a specific form of dualism, namely an internalist one, this possibility can only be because the natures of contentful kinds depend on factors or conditions external to the bodies of persons who undergo states of those kinds.<sup>20</sup>

Versions of externalism that are articulated in terms of this general supervenience claim are sometimes called *modal externalist theses*.<sup>21</sup> These are concerned with the existence of twin earth examples. Since the twin earth examples make explicit the dependency of contentful kinds on factors or conditions external to subjects' bodies, implicit in supervenience formulations of externalism is a claim which is sometimes called *constitutive externalism*. This is the claim that the correct philosophical account of the natures of certain contentful types takes them to have natures that depend on factors or conditions that exist beyond the bodies of individual subjects that undergo states of those kinds.<sup>22</sup>

---

<sup>20</sup> One might think that dualism alone is sufficient to account for the truth of this claim. However, dualism is silent on the internalism/externalism issue. It is consistent with dualism that mental contents should be individuation-dependent on factors external to the bodies of thinkers (and so external to the mind). See McCulloch, *The Mind and Its World*, p. 227, note 5.

<sup>21</sup> This is Davies's terminology. See 'Aims and Claims of Externalist Arguments', pp. 227-8. See also his 'Externalism, Architecturalism, and Epistemic Warrant'.

<sup>22</sup> As Martin Davies puts it, constitutive externalism says that

Constitutive externalism is the view I want to defend. Although it is a common strategy to employ the twin earth examples to establish it, I want to defend the view in a more direct way. The twin earth examples are best viewed as a kind of counterfactual test of the truth or otherwise of constitutive externalism. This test is meant to flesh out and validate intuitions about the object-dependence of contentful kinds. However, the test is only as persuasive as the intuitions that prompt it. If one is inclined to think that mental contents are object-dependent, then one will be inclined to accept that the twin earth examples are really possible and that they establish such object-dependence. If on the other hand, one is inclined to think that mental contents are not object-dependent, then one will be inclined to think either that the twin earth examples are not possible or that they do not show that mental contents are object-dependent.<sup>23</sup>

---

the most fundamental philosophical account of what it is for a person or animal to be in the mental states in question does avert to the individual's physical or social environment, and not only to what is going on within the spatial and temporal boundaries of the creature. ('Aims and Claims of Externalist Arguments', p. 230).

Davies correctly points out that one can establish a constitutive externalist thesis by establishing that modal *individualism* is false, i.e. that the supervenience claim (1) stated above is true, but that one cannot establish modal externalism just by establishing that constitutive externalism is true. It may be, for example, that although constitutive externalism is true, modal externalism is false because there is a necessary connection between subjects' intrinsic physical properties and factors or conditions beyond those subjects' bodies, so that an environment in which the contents of subjects' intentional states varied would necessarily be an environment in which their intrinsic physical properties also varied.

<sup>23</sup> This emerges in debates such as that between Burge and Fodor concerning the truth or falsity of anti-individualism. Burge argues that attention to actual descriptive and explanatory practices in psychology reveals the taxonomy of both intentional and nonintentionally described behaviour and the taxonomy of intentional states to be non-individualistic. For the interpretation of these practices fails to respect local supervenience, and this is supported by the twin earth thought experiments. However, his arguments for anti-individualism, based on these arguments, have been charged with presuming the truth of anti-individualism. In a similar vein, Burge effectively accuses Fodor's arguments for individualism, which also make use of twin earth thought experiments, of presuming the truth of individualism. Fodor argues that since whether or not twins have type-identical states depends on whether they have the same causal powers, and since sameness and difference of causal powers must be assessed across contexts rather than within them (causal power being a counterfactual notion),

I want, therefore, to ground the intuitions on which that test is based in certain features of actions and their explanation, where the relevant actions are ones based on perception. Like Burge, I see the source of externalism as lying in our actual descriptive explanatory practices. And I believe that attention to these practices can help to explain certain of our intuitions in the twin earth cases. But the argument for externalism can be mounted independently of the twin earth cases. So our intuitions concerning externalism can be vindicated without appeal to them.

The argument that I develop specifically concerns thoughts and other intentional states whose contents, widely construed, concern natural kinds, such as salt and water.<sup>24</sup> I believe that it can be generalized to other sorts of cases, but I shall not attempt that here.

## II The Source of Externalism

It is common in debates between externalists and individualists for both parties to appeal to behavioural considerations in support of their claims about the individuation of contentful kinds. But it is important to see how

---

whether twins have type-identical intentional states depends on whether their states have the same causal powers across contexts. Burge agrees, but argues that twin earth considerations cannot determine and distinguish causal powers of intentional kinds because one cannot decide which contexts are relevant for determining and distinguishing causal powers without making assumptions about the kinds in question. To suppose that the actual environment external to subjects' bodies is not relevant to determining causal powers, and so taxonomy of contentful kinds, is already to assume individualism. The moral for the twin earth thought experiments is that they play a more peripheral role in adjudicating between individualism and anti-individualism. The reason is that their employment is evidently not independent of individualistic/anti-individualistic assumptions. See the debate between Burge and Fodor in *Philosophy of Psychology: Debates on Psychological Explanation*, (ed.) C. Macdonald and G. Macdonald (Oxford: Basil Blackwell, 1995), containing Burge's 'Individualism and Psychology' and Fodor's 'A Modal Argument for Narrow Content', with a commissioned reply by Burge.

<sup>24</sup> In fact, nothing in the argument to follow requires commitment to any doctrine about natural kinds, even though the examples concern what many would consider to be natural kinds. Natural kinds are typically employed in twin earth thought experiments in order to bolster the view that twin earth twins might have thoughts that are distinct despite the phenomenological indistinguishability of the objects or substances to which their thoughts relate in their respective environments. However, the thesis that is being defended here is constitutive externalism, not twin earth externalism. Further, the examples on which the argument is mounted make reference only to the observable effects on normal observers of objects in their environments.

this appeal is put to work in arguments for and against externalism, and how little it establishes in the way of externalist or individualist conclusions.

Consider, for example, the debate between Tyler Burge and Jerry Fodor. Burge maintains that explanatory practices in psychology support externalism/anti-individualism because the explananda in many cases, when they are behaviour, are commonly and clearly understood to be behaviour, relationally understood as involving relations between organisms and their environments.<sup>25</sup> Thus, he appeals to the fact that one distinguishes a heart from a waste pump by its biological function in the organisms in which it performs its function, which cannot be determined to be what it is independently of the causal history of its ancestors in organisms of the same and similar kinds. Its function cannot be specified independently of relations it bears to its surrounding environment, and the way it is embedded in that environment.

However, Fodor does not deny that many of the behaviours in which intentional creatures engage, intentionally described, are to be understood as involving relations between organisms and their environments. What he denies is that such relations are *relevant* to the *taxonomy* of intentional content, at least for the purposes of causal explanation employing such content. They are not relevant because they do not make a difference to the causal powers of contentful kinds. And the individuation of contentful kinds is sensitive only to their causal powers. Thus, he reasons that because twin earth twins are molecular duplicates and so their actual and counterfactual behaviours are identical in relevant aspects, the causal powers of their mental states are identical in relevant respects. They therefore belong to the same natural kind of purposes of psychological explanation, and individualism is true.<sup>26</sup>

Fodor recognizes that this argument can be turned on its head simply by denying that the actual and counterfactual behaviours of me and my twin are identical in relevant respects. After all, when I am thirsty, I reach for water, whereas when my twin is thirsty, she reaches for twater.<sup>27</sup> Since the behaviours are not identical, neither are the causal powers of the mental states which explain them. Inasmuch as externalists and

---

<sup>25</sup> See 'Individualism and Psychology', and 'Intentional Properties and Causation', in *Philosophy of Psychology*, ed. Macdonald and Macdonald, pp. 226-35.

<sup>26</sup> Fodor, 'A Modal Argument for Narrow Content'. Fodor has since given up his commitment to narrow content. See *The Elm and the Expert* (Cambridge, MA: MIT Press, 1994).

<sup>27</sup> See Burge, 'Individualism and Psychology'.

individualists are agreed that differences in behaviour, non-intentionally described, are not what is at issue, but rather differences in behaviour, *intentionally* described, it seems that this argument for individualism does not go through.

Fodor, however, is not perturbed by this. He argues that the question of whether the relevant intentional kinds of twins are the same is a matter of their causal potentialities and that this is to be determined, not *within* contexts, but *across* them. So, for example, the fact that my beliefs on earth cause me to drink water whereas my twin's on twin earth cause her to drink twater does not show that these beliefs have distinct causal potentialities. What is relevant is whether my twin's beliefs *would* cause her to drink water on earth and whether my beliefs *would* cause me to drink twater on twin earth. By this (cross-context counterfactual) criterion, the causal potentialities of our beliefs are the same and the beliefs are type-identical.

This response doesn't quite work, since it is still vulnerable to the charge that when I utter the words 'Gimme water' on earth, I get what I ask for, but when I utter the words 'gimme water' on twin earth, I do not get what I ask for. Similarly for my twin. Our behaviours, intentionally described as water/twater requests, do not have the same causal powers, even across contexts.

Fodor attempts to patch the criterion up by providing a general condition on when differences in properties of causes are differences in causal powers. His claim is that differences in properties of causes are differences in causal powers when those properties are *not conceptually connected* to the effect properties for which they are responsible. By these lights, such differences as there are in intentional behaviour between me and my twin cannot be attributable to differences in the contentful properties of the states which cause that behaviour, widely construed as beliefs about water and beliefs about twater. For those properties are conceptually connected to the properties of the behaviour which makes them intentional, i.e. actions, namely, water requests and twater requests. My water requests and your twater requests may differ, but this difference in behaviour does not mark a difference in content between my water beliefs and your twater beliefs, since the contentful properties of these beliefs, widely construed as beliefs about water and beliefs about twater, are conceptually connected to the behaviour those beliefs cause.

What this debate between Fodor and Burge brings out clearly is that one can agree (1) that intentional content is to be taxonomized by its relation to behaviour, (2) that behaviour is decisive in determining the

truth or falsity of externalism and (3) that behaviour is to be taxonomized for psychological purposes intentionally in ways that involve relations between organisms and their environment, and yet (4) *still* disagree about whether externalism is true or false. Burge and Fodor agree on all of these points, and even on the further two points that (5) mental kinds are to be taxonomized in terms of their causal powers for the purposes of psychological explanation and (6) psychological explanation is causal explanation. But despite all of this agreement, they disagree about whether externalism is true.

What this shows, I think, is that the truth or falsity of externalism, inasmuch as it turns on the broad/narrow content distinction, depends on the issue of the explanatory efficacy of broad or wide content. That is to say, it depends on the issue of whether, in at least some cases of the explanation of action, the contentful kinds implicated in such explanations, to do their explanatory work, must be individuated widely, i.e. by relation to factors that exist beyond the surfaces of the bodies of organisms who undergo states of these contentful kinds. Burge thinks they must because individuation of contentful kinds is individuation by causal powers, but this is not independent of assumptions about the kinds in question. Contentful kinds, like biological-functional kinds, are not only causally but conceptually connected with their effect properties. So the taxonomy of the cause properties is not independent of conceptual connections with their effect properties.<sup>28</sup>

In Burge's view, this makes psychological explanation, like functional explanation in biology, explanation which is causal but which breaches the 'Humean' requirement of connecting effects with causes non-conceptually.<sup>29</sup> Burge acknowledges that it breaches this requirement, but does not see that it presents any problem for the view that

---

<sup>28</sup> Thus he claims,

One could plausibly claim that it is a conceptual truth that hearts differ from twin waste-pumps in that they pump blood. One could plausibly claim that it is conceptually necessary that if something is a heart, then when functioning normally, it pumps blood. ('Intentional Properties and Causation', p. 233).

<sup>29</sup> According to Burge, that is. See 'Intentional Properties and Causation'. In this Burge concurs with Neander ('Functions as Selected Effects: The Conceptual Analyst's Defense', *Philosophy of Science* 58 (1991), pp. 168-84.). But note that Millikan (*Language, Thought, and Other Biological Categories*) denies that there are such conceptual connections between functional properties and the effect properties to which their taxonomy is sensitive. Similarly for intentional properties.



psychological explanation is causal explanation, since he rejects the 'Humean' requirement.

Fodor, on the other hand, thinks that the contentful kinds implicated in the explanation of intentional behaviour or action need to be individuated narrowly, i.e. individualistically. This is because, although he agrees with Burge that individuation of contentful kinds is individuation by causal powers, he accepts the Humean requirement that causes and effects must be individuated in terms of properties that are conceptually independent of one another if the cause properties are to be genuinely causally potent with regard to their effect properties and psychological explanation is to be genuinely causal explanation. Since psychological explanation is genuine causal explanation, it too must meet this requirement. By that standard, widely individuated content gets ruled out from being genuinely causally potent, hence genuinely explanatorily potent.

So the crucial issue that divides Burge and Fodor is whether the explanatory potency of intentional kinds requires that such kinds meet the Humean requirement on causal explanation of being conceptually independent of their effect properties. I think that Burge is correct in his claim that psychological explanation, explanation of actions by means of states with intentional content, works by way of broadly conceptual connections between explanans property and explanandum property. Such contentful properties do their explanatory work because they have causal powers which relate them conceptually to their effect properties.

However, unlike Burge, I believe that attention to the ways in which psychological explanation is like explanation in functional biology shows it to be of a distinctive, normative noncausal type. Moreover, I think that by attending to the ways in which contentful properties are like biological-functional ones, and unlike physical ones, it is possible to mount an argument for externalism that does not lead to the kind of stalemate that seems to be the inevitable result of debates between externalists like Burge and individualists like Fodor.

The dispute is between those who agree that taxonomy of contentful properties is taxonomy by causal powers but disagree about whether this supports externalism because they disagree about whether such taxonomy meets the Humean requirement that for a property to be a distinctive causal power, it must be contingently or non-conceptually related to its effect property. But this dispute seems to me to be unresolvable within the narrow confines of externalism. It simply relocates the disagreement in the issue of whether psychological explanation is like functional

explanation in biology or like causal explanation in such sciences as physics. However, one needs a principled reason for adjudicating between these two alternatives.

I want to try to provide that principled reason by showing that and how the explanation of action by intentional content is like functional explanation in biology in a certain important respect. First, I shall mount the argument. Then I will locate the source of the externalist commitment, and indicate how like it is to the source of externalism in functional biology.

### **III An Argument for Externalism**

I begin with the observation that the truth or otherwise of externalism does not depend on whether the explananda of psychological explanations are actions construed widely or actions construed narrowly (but intentionally). As the debate between Burge and Fodor illustrates, one can agree with an externalist that the explananda of psychological explanations are actions, widely construed, and disagree about whether this shows externalism to be true. However, I think it plausible that contentful states are employed as explanantia of both sorts of actions. Sometimes, for example, we may wish to explain why a subject washes her clothing with water (rather than with sand, or with Coca-Cola), where what seems to need explaining is why she engages in a particular *type* of action with regard to a particular object or type of object. But there are other cases where what we wish to explain is not why a subject engages in a certain type of action with respect to a particular object or type of object, but where we simply wish to explain why that subject engages in actions of a particular type at all. Sometimes, for example, we may want to explain why a subject eats every day, or goes to bed at night, where the actions that serve as explananda are actions, narrowly construed. It may be that a subject cannot eat without eating something, but *what* is eaten is not what one wants to explain. What one wants to explain is the *activity* of eating, or the *activity* of washing, itself. Phenomena like these are actions, narrowly construed.

Narrow actions seem to be just the sort of phenomena whose explanation would only require narrow content, if any phenomena are. So let us concentrate for the moment on actions, narrowly construed. If the explanation of even these cannot be effected without appeal to wide content, then externalism will have been vindicated.

Narrow-act explanation seems to require no mention of any particular object, or of any of a range of objects, on which such actions depend. Because of this, the states which explain and make intelligible such behaviour also seem to be capable of doing so by means of narrow content. That such actions can be construed individualistically evidently supports the view that the contentful kinds that are required to explain them by making them intelligible can also be construed individualistically. For if their taxonomy does not depend on the existence of any particular object or range of objects, then they can evidently be made intelligible, or explained, by means of contents that also do not depend on the existence of any particular object or range of objects.

This idea can be further supported by a twin earth thought experiment. Consider Sue, who washes with water, and her twin, who washes with twater. Although the activity of washing requires that there be *something* that one washes with, the activity itself, *what* Sue and her twin do with the same respective stuff, is the *same kind of thing*. Since the activities are the same, it is plausible to hold that so too are the contentful kinds which explain them.

I do not think, however, that this establishes individualism. The reason is that the individuation of actions, narrowly construed, can only take place against the background of wide-act individuation; and wide acts are only made intelligible by states with wide content.

Actions are not only purposeful; at least sometimes they involve interaction with objects. Further, when these actions are successful, that they are object-involving is not an accidental feature of them. If such actions were not at least sometimes non-accidentally object-involving, they could not be purposeful. But if they could not be purposeful, they could not be actions at all.

The point here is not that there must be successful actions if there are to be actions at all. It is true that actions, in being purposeful movements, *aim* at success. But this is consistent with the possibility that no action is actually successful; that creatures should regularly fail to succeed at what they aim to accomplish by moving their bodies in various ways.

So it is not a necessary, but a contingent matter that there are successful actions in the world. It is a contingent matter that there are objects in the world with which human beings engage, and it is a contingent matter that by engaging with these objects they are both changed by and change the world. But that there are such objects with which human beings engage, and that they at least sometimes do so

successfully, is, I take it, common ground between externalists and individualists.

Given that there are successful actions in the world, intentions that engage with the world are required to make them intelligible. Without such intentions, one cannot make intelligible the non-accidental connection between action and object when an action is successful. This is because without intentions that engage with the world, there is an explanatory gap which leaves it mysterious why that connection is non-accidental. One makes it intelligible by citing intentions concerning objects that match the objects with which subjects non-accidentally engage in their behaviour. The connection between purpose and the world beyond the body is required because successful object-involving action is non-accidentally successful. Wide content is needed to explain successful action, which happens to, but need not, occur.

Think, now, of Sue, who washes every day. We, who want to make intelligible that activity, explain it in terms of her desire to make herself clean and her belief that by washing she will make herself clean. However, that belief and desire will only serve to explain her activity against the background of assumptions she has about the sorts of stuff that make things clean. Water can make things clean. Sand can make things clean. But Coca-Cola cannot make things clean. Nor can tar, mud or hydrochloric acid.

In short, the intelligibility of *what* one is doing, narrowly construed as a successful activity, takes place against the background of assumptions about what it is *appropriate* to do it *with*. One does not make intelligible Sue's activity of washing every day just by mentioning her desire to make herself clean and her belief that by washing she will make herself clean. Her activity simply does not count as an activity of washing if she does it with mud, or with Coca-Cola, or with tar. And here I mean: *successful* activity. Her movements may be the movements of someone who takes herself to be washing. But movements are not actions; and their classification as actions, even narrowly construed, depends on what the appropriate objects are towards which they are directed. One's actions being the successful actions they are *depends* on the appropriateness of such objects to them.

So narrow-act taxonomy is made intelligible against the background of wide-act taxonomy, taxonomy which is object-dependent at least to the extent that it requires appropriate objects in the environment in which agents are embedded towards which they can successfully act. And given that this is so, the explanation of even narrow acts by states with narrow

content is made intelligible against the background of explanation of wide acts by states with wide or broad content. Sue's activity of washing herself is made intelligible by her desire to make herself clean and her belief that by washing she will make herself clean, only if she also has beliefs about what it is appropriate to wash with. Her success in *washing* depends on this. That is to say, her *washing* depends on this, since the taxonomy of her behaviour, narrowly construed, is not independent of what objects are appropriate to wash *with*.

Even her unsuccessful attempts at washing, using inappropriate substances such as tar or mud, require this. For her unsuccessful attempts are ones in which she mistakenly *takes herself* to be washing. But if she takes herself to be washing, then she takes herself to be doing what agents do when they wash. Sue can only be mistaken about whether she is washing if she has correct beliefs about the washing, which require beliefs about what it is appropriate to wash with.

These observations about successful actions and appropriate objects may seem insufficient to establish externalism. For externalism requires that the contents of at least some intentional states be individuation-*dependent* on factors beyond their bodies. Sue's water content must be a *water* content, not just a content that depends on the existence of some substance or other. And it is not clear that appeal to successful actions towards appropriate objects establishes such dependence.

Consider Sue's twin. She successfully does with *twater* what Sue does with water. And *twater* is, on *twin* earth, appropriate to wash with. On twin earth, *twater* gets things clean. Appeal to considerations about successful actions towards appropriate objects fails to discriminate between Sue's behaviour and her twin's behaviour, narrowly construed. Does this not show that the dependency of actions and intentional content on appropriate objects is insufficient to discriminate between water and *twater* contents, so that individualism is still viable?

One cannot respond to this by saying that whether an object is appropriate for a given activity depends on its empirically discoverable nature. The problem that the twin earth examples pose is that although water and *twater* have *different* natures, both are appropriate to wash with. So it looks as though such differences as there are between their natures cannot make a difference to the determination of intentional content.

The problem arises because the twin earth examples are designed to keep firmly in place the ordinary, day-to-day role that certain objects, substances or phenomena beyond persons' bodies play in their activities and, correspondingly, in their thoughts, while varying in their natures in

ways that are hidden to the naked eye. Since our day-to-day activities can be and often are insensitive to such differences in the natures of things that do not manifest themselves to the naked eye, it is not surprising that water and twater should play the same (narrow) role in Sue's and her twin's day-to-day activities.

However, I think that externalism can be defended in the face of this. The twin earth examples are effectively tests of antecedently held intuitions, as I have said. How they are to be interpreted, and what they establish, depends on the intuitions that prompt them. The intuition that the argument for externalism just given sets out to defend is that water contents are *water* contents because our day-to-day descriptive and explanatory practices can only intelligibly explain subjects' successful actions, even narrowly construed, against the background of successful actions which take water as an appropriate object, where these explanations require the employment of contents that are *water* contents. Water actions are made intelligible by states with *water* contents. Unsuccessful attempts at water actions that take objects other than water are made intelligible in part by states with *water* contents.

It is true that on twin earth, it is stipulated that twater plays the role in twin-earthians' day-to-day life that water plays on earth. Why then do Sue and her twin think thoughts with different contents? Because whether Sue's thoughts are *water* thoughts depends on her actual behaviour, in the actual context in which she is situated. Her counterfactual behaviour – what she would do in other circumstances, and in other environments – depends on this. If she were to be transported to twin earth, what would she do? She would wash with twater, drink twater, and so on. But it would not be appropriate *for her* to do so.

Why? Because appropriateness is context-dependent, and the organism is part of the context.<sup>30</sup> Sue's behaviour is appropriate on earth because of

---

<sup>30</sup> Two objections might arise here. One is that what is appropriate behaviour towards an object depends in part on how the type of object involved in that type of behaviour is specified, and that, specified more generally (say, as 'the thirst-quenching, odourless, transparent, colourless liquid'), twater is appropriate for Sue to wash with, because on twin earth, the stuff which satisfies that description gets things clean. The other objection is that appropriate behaviour must be, as the functional behaviour is, capable of allowing for novelty in the range of objects to which such behaviour can become adapted. Creatures move around and may, in new environments, encounter objects of kinds that are distinct from those of the kinds to which their behaviours were initially adapted. It may thus be accidental that these objects serve the needs for which the behaviours were initially selected. Still, engaging with them might prove to

---

be beneficial for these creatures, and so it may be functional for them to behave in the same way towards these new items as they did towards the old ones.

Consider the first objection. Suppose that we are concerned to specify the function of the frog's tongue-flicking behaviour. On one view, the function is to catch small dark moving things. On another, it is to catch frog-food. How we specify the object toward which the frog behaves matters because it makes a difference to whether the frog is functioning biologically normally rather than malfunctioning when it flicks its tongue at black spots in its visual field.

Similarly, it might be argued, for Sue and water/twater. On one view, appropriate behaviour for Sue is behaviour towards water. On another, it is behaviour towards the thirst-quenching, odourless, transparent, colourless liquid. How we specify the object towards which Sue behaves matters here too, because it makes a difference to whether Sue is behaving appropriately when she washes with twater.

Indeterminacy of function-specification is a problem in an environment where both specifications apply precisely because of its consequences for what would count as malfunctioning behaviour. And it is a problem in Sue's environment, since both ways of specifying water are satisfied by water. However, the considerations that may lead one to think that the correct specification of the frog's behaviour is the more general one do not apply with equal force to Sue's behaviour.

In the case of the frog, the inclination to specify its functional behaviour as frog-food catching behaviour seems poorly motivated in the light of the fact that the frog's perceptual system seems to be sensitive only to small dark moving things in its environment. To attribute more specificity in functional behaviour to the frog than this would require us to view the frog as capable of seeing small dark moving things as frog-food. But nothing in its behaviour gives us good reason to suppose that the frog has this capacity.

The situation is different for Sue and water. The frog's environment contains many things which count both as frog-food and also as small, dark, and moving. However, Sue's environment does not contain many stuffs that are phenomenologically indistinguishable from water. Whereas in its actual environment, the frog flicks its tongue at many small dark moving things which may not be frog-food, Sue does not in her actual environment wash with many thirst-quenching, odourless, transparent, colourless liquids which may not be water. Whereas, in the case of the frog, we see no reason to specify its behaviour in the more specific way, in the case of Sue, we have no motivation for specifying her behaviour in the more general way. It all depends on the organism and what is in its actual environment.

So, in the case of Sue, unlike the case of the frog, we do have a reason to specify her behaviour as appropriate behaviour towards water. And so we have a reason for taking her behaviour on twin earth towards twater to be inappropriate, although intelligible. We can make sense of such inappropriate behaviour because, although the environment of twin earth happens to cooperate with Sue, that it does is an accident.

Now consider the second objection. Suppose that the correct way to specify the function of the frog's tongue-flicking behaviour is in terms of its goal in catching frog-food. Still, it might be argued, different things in different environments might count as frog-food. Thus, suppose that the frog were to be placed in a new environment, one

the beneficial effects washing with water has in that environment. But on twin earth those beneficial effects are *accidental* for Sue: on twin earth it is an accident that twater is appropriate for Sue to wash with. It is no accident that on twin earth twater is appropriate to wash with for *twin earthians*. And so it is no accident that Sue's twin acts in ways made intelligible to twin earthians by twater contents. But Sue's actions will not be made intelligible by beliefs and desires of hers with twater contents – not, at least, independently of the fact that Sue's actions are based on misperceptions of twater as water. Given that Sue thinks water thoughts and given that such differences as there are between water and twater do not manifest themselves in the day-to-day role that these substances play

---

where creatures of a different type than those to which the frog's tongue-flicking behaviour was originally adapted nevertheless served to nourish the frog. Would it not then be functional for the frog to flick its tongue at these different creatures?

I want to say here that the behaviour in the new environment, however beneficial to the organism it may be, is not thereby functional for the frog. The reason is that whether a behaviour is functional depends on the types of objects to which the behaviour was initially adapted. It was frog-food, not small dark moving things, to which the frog's behaviour was initially adapted, and for which that behaviour was selected. It was *creatures* of a certain type to which that type of behaviour was initially adapted and for which that behaviour was selected. So it was frog-food of a certain kind to which the frog's behaviour was initially adapted and for which that behaviour was selected. And that is why it is functional for the frog to flick its tongue now in the presence of that kind of frog food. That different organisms in another environment nourish the frog, so that its tongue-flicking behaviour in those circumstances is beneficial to the frog, is fortuitous. It is just good luck for the frog that its new environment obliges its need for nourishment by supplying different, but satisfying creatures for it to eat. (See Ruth Millikan, 'Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics', in *White Queen Psychology and Other Essays for Alice* (Cambridge, MA: MIT Press, 1993), pp. 123-33, especially 125-31.) Happy coincidence between producer and consumer does not thereby make for functional behaviour. Similarly for Sue's behaviour towards twater.

It does not follow from this that the frog's functional behaviour cannot be adapted to new things, and that these things cannot come to figure in the process by which a type of behaviour or trait is selected. They can. And those that do will thereby figure in the specification of objects towards which that type of behaviour is functional. But that they are objects towards which a type of behaviour is functional depends on their role in the selection process, and not vice versa. Similarly for Sue and her appropriate behaviour towards water.

In responding to these objections in this way I am presuming a particular view of biological function, namely a causal-historical view, such as that advocated by Millikan in *Language, Thought, and Other Biological Categories*. It contrasts with ahistorical accounts, such as propensity accounts (see John Bigelow and Robert Pargetter, 'Functions', *Journal of Philosophy* 84 (1987), pp. 181-96).



in the activities of agents on earth and on twin earth, it is not surprising that Sue should wash with twater. It is not surprising; but nor is it true that Sue's act of washing on twin earth is made intelligible independently of her water thoughts.<sup>31</sup>

Is this response question-begging against the individualist, whose criterion for the taxonomy of behaviour and intentional content is counterfactual? I do not think so, and the reason connects the taxonomy of psychological kinds firmly to the taxonomy of functional kinds in biology. In biology, the taxonomy of functional kinds is both teleological and what one might call 'effect-sensitive' in a normative sense of the term. The camouflaging behaviour of this chameleon is camouflaging behaviour, not because it has camouflaging effects in *this* chameleon, nor even because it tends to have camouflaging effects in the majority of chameleons. It is camouflaging behaviour in this chameleon because this type of behaviour had camouflaging effects in a sufficient number of its ancestors to lead to the proliferation of chameleons which displayed that behaviour. That type of behaviour exists *in order* to have camouflaging effects in this and other chameleons, whether it does so or not. And it exists in order to have such effects precisely because its presence in ancestors led to the survival and so to the proliferation of chameleons which displayed this behaviour. Similarly for the deer's flight behaviour, the bee's dancing behaviour, and so on.

---

<sup>31</sup> As Burge puts the point,

Imagine that a heart and an organ that pumps digestive waste (from a completely different evolutionary scheme) were physically indistinguishable up to their boundaries. Clearly they would be of two different biological kinds, with different causal powers, on any conception of causal power that would be relevant to biological taxonomy. Judging the heart's causal powers presupposes that it is connected to a particular type of bodily environment, with a particular sort of function in that environment. One cannot count being connected to such a body to pump blood as just one of many contexts that the heart might be in, if one wants to understand the range of its biologically relevant causal powers. It would show a serious misconception of biological kinds to argue that the causal powers and taxonomically relevant effects of the heart and its physical twin are the same because if one hooked up the waste pump to the heart's body, it would pump blood and cause the blood vessels to dilate; and that if one hooked the heart to the waste pump's body, it would move waste. ('Intentional Causation and Psychology', p. 227).

In biology, teleology arises from the working of natural selection on instances of physico-chemical properties of organisms.<sup>32</sup> Turning green in this environment just in this chameleon's camouflaging itself given that it inhabits a green environment and that instances of this type of behaviour in this chameleon's ancestors helped them to avoid predators and so to aid survival and proliferation of descendants. It is the success of instances of certain types of behaviours in the past that gives rise to teleology in functions, and this teleology persists even when the effects which instances of such behaviour now have regularly fail to occur. Causes are *designed* to bring out certain effects, even when they do not. Still the functional pattern remains, and still the chameleon displays such behaviour, *in order* to have camouflaging effects.

So in biology, certain types of behaviour too *aim* at success: their having teleology just in their aiming at success. To be a functional kind just is to aim at success, and this makes functional behaviour very like action in this respect. In biology, the fit between behaviour and environment is non-accidental when successful because the cause – say, the chameleon's turning green – is designed to have a certain (functional) effect, namely, to match the colour of the environment. So too in the domain of intentional psychology. In successful action, the fit between behaviour and environment is non-accidental when successful because the cause – contentful intentional states – is designed to have a certain (purposeful) effect. The source of the design in the two cases may not be the same, since in biology it is brought about by the process of natural selection. But the design itself – that the cause exists in order to bring out a certain type of effect – is present in both. This is what makes for teleology in biology, and it is what makes for teleology in the domain of the psychological.

If this is right, then the non-accidental fit between activity and object in actual cases of successful action is required for the taxonomy of action itself, just as the non-accidental fit between behaviour and environment in successful behaviours is required for the taxonomy of functional kinds. In the biological case, successful behaviour depends on the actual environment in which it was selected: a chameleon placed in a pink

---

<sup>32</sup> This is the view of biological-functional kinds advocated by Millikan in numerous works. See particularly *Language, Thought, and Other Biological Categories*, and *White Queen Psychology and Other Essays for Alice*. It is a causal-historical (in contrast to a propensity) account of biological function.

environment cannot camouflage itself.<sup>33</sup> And in the case of action, the success depends on there being appropriate objects toward which agents can at least sometimes act. Sue can get it wrong occasionally when she takes inappropriate substances to wash with. She can even get it wrong much of the time. But she cannot get it wrong all of the time. Beyond a certain point, we are no longer prepared to say: she washes. And getting it right or wrong means getting it right or wrong *in her environment*. And so what is appropriate is what is appropriate *in her environment*. Similarly for Sue's twin. And *this* is why a counterfactual criterion is inappropriate for the taxonomy of action and content.

In the case of biology, teleology is there because of functions, and functions arise from the workings of natural selection on instances of physico-chemical properties. Why is there teleology in the case of action? Here there is teleology because there is intentional content. Contentful kinds imbue behaviour with purpose. Behaviour counts as intentional, hence as action, only if it is caused by states with intentional content. But if so, then they aim at success precisely because their contentful causes themselves aim at success. Without teleology in the contentful cause, there is no teleology in the intended effect. This is not, of course, to say that without intentional states there would be no teleology in human behaviour. There would be teleology because human beings are biological creatures. But the teleology would not be the teleology of intentional behaviour. Movements would remain purposeful, but they would not thereby be intended, and so would not be actions.

## IV Conclusion

This completes the argument for externalism. It remains to consider some of its consequences. The externalism argued for here is distinctive in being essentially normative. The objects (etc.) upon which narrow-act taxonomy depends are ones that are *appropriate* to those actions. As I remarked at the end of the last section, this normativity lies firmly in the mind. It lies in the essentially teleological nature of contentful kinds.

However, for reasons which should be apparent from the comparison of psychological kinds with functional kinds in biology, I do not think that

---

<sup>33</sup> Millikan uses the term 'Normal' (with capital 'N') to distinguish the biological-normal from 'normal' in the sense of 'average' or 'usual' or 'typical'. See *Language, Thought, and Other Biological Categories*.

this teleology can exist independently of the actual environments in which agents are embedded. So what counts as an appropriate object towards which an agent can successfully act is not independent of the actual, empirically discoverable nature of that object. What makes Sue's washing with water a successful act is not independent of its being a washing with water; and water's being an appropriate object to wash with is not independent of the fact that it is, unlike hydrochloric acid, H<sub>2</sub>O.

Hydrochloric acid does not have a nature such that it is appropriate to wash with, nor does mud, nor tar. Given what has just been said, it would not be *rational* for Sue to attempt to wash with these substances.

However, it would be intelligible for Sue to wash with twater on twin earth, since we can make intelligible why she might think it appropriate to do so, even though it isn't appropriate *for her*. For twater is phenomenologically indistinguishable from water and also gets things clean. Given this, it is understandable that Sue should do something that is inappropriate for her but appropriate for twin earthians. It is understandable, and perhaps also rational, in much the same way that a person who misperceives salt as sugar and pours salt in her tea can be seen to be behaving rationally because of this misperception. But it is not thereby appropriate, given that appropriateness is context-dependent and the individual is part of the context. There is slack between what is appropriate in the environment of twin earth and what is appropriate for Sue.

This is not to say that whether an object or substance is appropriate to a certain activity reduces to, or can be determined only by, its empirically discoverable nature. It is true that whether a given object or substance is appropriate for a given activity depends in part on its empirically discoverable nature; but different natures can be equally appropriate for the same type of activity. On earth it is appropriate to wash with water; but it can also be appropriate to wash with sand. Apples are appropriate objects to eat, but so are walnuts and mushrooms.

Because objects with different natures can be equally appropriate to a single activity, there is no telling in advance, there is no *a priori* limit on, what object or objects can be appropriate for a given activity. Appropriateness depends on actual effect – for example, in the case of washing, getting things clean – and this in turn depends in part on the nature of the objects with which one engages when one acts. But different objects can have the same effect. And so different objects can be appropriate for a given activity. It is appropriate to wash with water, and

with sand, but not with tar, because water and sand get things clean, but tar does not.

However, the similarity between contentful intentional kinds and functional-biological kinds ends here. Specifically, the attribution of contentful kinds is, whereas the attribution of functional kinds is not, sensitive to both the perspective of others and the perspective of the subject. This makes the norms that govern the attribution of intentional content to subjects' states answerable both to the perspective of those subjects and to the perspective of others. There is no analogue of this dual-perspective constraint on functional taxonomy in biology.

Failure to appreciate this can lead to too close an assimilation of the psychological to the biological, with undesirable consequences. For example, Ruth Millikan has claimed that because externalism is true, a subject who is ignorant of the factors that determine intentional content might fail to recognize that two beliefs with the same content have the same content, and as a result hold contradictory beliefs, one the negation of the other. Many externalists would be prepared to concede this point.<sup>34</sup> But she goes on to infer from this that the norms that govern rationality itself lie beyond the individual subject, in much the same way that the determinants of functional behaviour lie beyond the biological organisms that display it. She holds that externalism has the consequence that whether humans are rational is not determined by, or even partly answerable to, subjects' perspectives.<sup>35</sup>

---

<sup>34</sup> See, for example, Tyler Burge, 'Intellectual Norms and the Foundations of Mind', *Journal of Philosophy* 83 (1986), pp. 697-720.

<sup>35</sup> Thus she says,

it is implicit in contemporary 'externalist' accounts of the contents of thought that what is consistent versus inconsistent, indeed, I will argue, what is rational versus irrational, is not epistemically given to the intact mind. ('White Queen Psychology', in *White Queen Psychology and Other Essays for Alice*, p. 281)

And similarly,

If the White Queen is right, then that Alice has a coherent system of thought, that she possesses, for example, only one thought of each semantic kind, and hence that she thinks in accordance with laws, say, of rational psychology, depends on a felicitous *coordination* between Alice-the-organism and Alice's environment. It depends, in fact, on much the same kind of felicitous coordination that constitutes Alice's thinkings of *true* thoughts; rationality fails to be in the head in the same sort of way as does truth. (Ibid., p. 285.)

But I deny that externalism has any such consequence. It is true that externalism implies that subjects might think thoughts with contents they imperfectly grasp. And since they might, they might mistakenly think that two thoughts have the same content when they do not, or that they have different contents when they do not. Fallibility in knowledge of one's own thoughts of this kind is indeed a consequence of externalism.

But I do not see that the norms that govern rationality do not thereby operate 'from within' the individual subject. For these norms operate in epistemic ways, in ways that make the behaviour of agents intelligible, not only to others, but also to themselves. And it is a constraint on the attribution of content by others that such attribution respect the agent's perspective.

In short, the factors that make a certain content the content it is do not thereby determine the acceptance or rejection of that content by a subject, or the patterns of reasoning in which that subject might engage with regard to that content. Externalism is a metaphysical view about the factors that help to determine intentional content. But rationality, and the norms that govern it, is an epistemological matter. So it does not follow from the fact that externalism is true, hence that the determinants of intentional content lie beyond the individual, that the determinants of reasoning and behaviour lie beyond the individual also. On the contrary, it

---

The illusion that modes of presentation will help save logical possibility also rests on a failure to see that rationality pivots essentially on *referential* content, or *Bedeutung*, and not at all on mode of presentation, that rationality cannot simply be lifted up and attached to mode of presentation. The capacity to reidentify content but only under a mode is a *restriction* on rationality, a lessening of rationality, not a removal of rationality into an inner and safer sphere. (Ibid., pp. 283-4.)

Millikan apparently thinks that *that* rationality, like content itself, is world-involving, shows that it lies beyond the subject altogether ('rationality pivots ... not at all on mode of presentation', 'rationality fails to be in the head in the same sort of way as does truth'). I deny that content externalism has this consequence. That rationality is world-involving does not thereby show that it does not depend in any way on the perspective of the subject, and so does not show that the norms that govern rationality lie beyond the individual. That *what* is thought about when thinking a content lies beyond the individual does not show that *how* it is thought about is not *also* involved in thinking rationally. The rejection of narrow content does not force one to reject any role for the subject to play in rationality; nor is it a 'pernicious Cartesianism' to insist on the importance of that role. Without it, it is difficult to see why subjects should be critically reflective thinkers, and what role critical reflection might serve in an individual's psychology.

is plausible to maintain, in the face of externalism, that these norms are accessible to and operate within the subject, and guide the very formation, rejection, and assimilation, of contentful states.<sup>36</sup>

---

<sup>36</sup> So I am recommending a combination of metaphysical externalism with regard to the determinants of intentional content and epistemological internalism with regard to the norms that govern rationality. Burge seems to pursue a similar strategy. See 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 96 (1996), pp. 91-116.