PHILOSOPHICAL ISSUES, 4
Naturalism and Normativity, 1993

Externalism, Naturalism and Method*

Kirk A. Ludwig

> Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics and leads the philosopher into complete darkness.
>
> Wittgenstein, *The Blue Book.*

1 Introduction

This paper is concerned with certain arguments and motivations for externalism in the philosophy of mind, and with the proper method for answering questions about the conditions for having mental contents. I am interested in particular in the interplay between arguments for externalism and the demand that the mental be naturalized. Broadly speaking, we naturalize the mental by showing how it can be integrated successfully with the rest of our picture of the natural world. Arguments for externalism often seem to presuppose that the naturalistic project, cast in the particularly strong form of providing a conceptual reduction of the mental to the non–mental, can be successfully carried out. I find the arguments for externalism unconvincing, and the motivations for pursuing the naturalistic

*I would like to thank John Biro and Martin Davies for helpful comments on this paper.

project in this form in which it is often cast, which would buttress these arguments for externalism, unpersuasive.

In the following, I first provide an account of the externalist thesis, distinguishing it from two other positions, which I call 'strong individualism' and 'internalism' —the distinction between which is easily overlooked— and reject a further distinction sometimes advanced between 'modal externalism' and 'constitutive externalism'. As we will see, getting clear about the relations among these views is crucial to any adequate evaluation of arguments for externalism. Next, I turn to certain thought experiments which purport to establish externalism specifically about perceptual content. I argue that they fail, for two reasons, one of which can be traced partly to the failure to observe the distinctions between the different views mentioned above. These results generalize to externalist arguments of the same form about other sorts of content. My primary target in this is a series of recent papers by Martin Davies, culminating in the one delivered at the conference which occasions the publication of this volume.[1]

2 What Externalism Is and What It Is Not

The externalist holds that an individual's thought contents are at least partially logically determined by his relations to events, objects, kinds, and so on, in his environment. The externalist thesis is, in short, that content properties are in part relational properties.[2] A property P is a relational property just in case, necessarily, for any object O, if O has P, then there is an X such that X is (i) not an abstract object and (ii) X is not identical to O or to any part of O.[3]

A remark about condition (i) is in order. I exclude abstract objects because they are necessary existents, and there would be some abstract object that would satisfy condition (ii) for any property of

[1]Martin Davies, 'Aims and Claims of Externalist Arguments', this volume; 'Externality, Psychological Explanation, and Narrow Content', *Proceedings of the Aristotelian Society*, Supplementary vol. 60, pp. 263-83; 'Individualism and Perceptual Content', *Mind*, vol. 100, pp. 461-84; 'Perceptual Content and Local Supervenience', *Proceedings of the Aristotelian Society*, vol. 92, pp. 21-45. I do not attribute to Davies all the motivations for externalism which I discuss.

[2]A property P is a content property iff there is some representational state R with content C such that, for any x, x has P iff x has R.

[3]Throughout I will be using 'necessarily' and 'possibly' in the sense of conceptual or 'broadly logical' necessity and possibility. It is only for claims about conceptual necessity and possibility that our intuitions about thought experiments can be appropriately used as evidence.

any contingent or necessary existent. I assume that only abstract objects are necessary existents, so that if any other object satisfies (ii) for a given property P, it will be in virtue of a constitutive connection between the instantiation of P and the existence of some object. This characterization of relational properties might be thought to be too exclusive because of the possibility of there being relations between contingent existents and abstract objects, or between abstract objects. Beliefs, for example, are often characterized as relations between individuals and propositions. But even if this is correct, it is clear that the issue between externalists and their opponents has never been about whether attitudes should be understood as relations between individuals and abstract objects. The issue is whether in specifying the content of an individual's thought or perceptual experience one must refer to his relations to his environment. Thus, no change in our definition is required for present purposes. If there are abstract objects and we bear relations to them, the definition above can be taken simply to characterize that class of relational properties which are at issue in the debate about externalism.

We can contrast the externalist thesis with the thesis that content is strongly locally supervenient.[4] This latter thesis holds that for any object O, if O is in complete non–relational physical state **S** and in mental state **M**, then, necessarily, for any object Q, if Q is in physical state **S**, then Q is in mental state **M** (I use boldface type to indicate that I am talking about a state *type*).

It is natural, perhaps, to take strong local supervenience and externalism to be both mutually exclusive and exhaustive alternatives. Thus, Martin Davies, in the discussion that is our main interest here, in 'Aims and Claims of Externalist Arguments', characterizes externalism as the negation of what he calls, adopting Tyler Burge's terminology, 'individualism'. Individualism he characterizes as equivalent to strong local supervenience, and casts the rest of his discussion in terms of the contrast between strong local supervenience, and externalism:

> ... the constitutive individualist claim entails modal individualist claims —claims of local supervenience.
>
> Given the statement of constitutive individualism, we can assemble a claim of constitutive externalism just by negating it.[5]

However, it is clear from the above characterizations of externalism and strong local supervenience that the negation of strong local su-

[4]The term is due to Davies, 'Perceptual Content and Local Supervenience'.
[5]See page 230 this volume.

pervenience is not equivalent to externalism. Both externalism and strong local supervenience are modal theses about the relation between content properties and other sorts of properties. In the case of externalism, these are relational properties; in the case of strong local supervenience, these are non–relational physical properties. That content properties are not logically fixed by non–relational physical properties does not entail that they are relational properties; equally, that content properties are not relational properties does not entail anything about their connections with non–relational *physical* properties. Externalism and strong local supervenience can both be false without contradiction.

Davies also distinguishes between what he calls 'modal externalism' and 'constitutive externalism'. Modal externalism is the thesis that no internal properties are logically sufficient for mental contents. Constitutive externalism is the thesis that relations between an individual and his environment are constitutive of his contents. Modal externalism is sufficient for the truth of constitutive externalism, and for the falsity of internalism and strong local supervenience. However, Davies suggests that constitutive externalism is *not* incompatible with strong local supervenience.

> As a barely formal point, this failure of entailment [between the negation of modal externalism and the negation of constitutive externalism] is clear enough; but perhaps we should consider a couple of ways in which it might turn out to be impossible to generate the 'Twin Earth' examples that would establish modal externalism. One kind of case would be where there is a necessary connection between the relevant features of the environment E and X's inner constitution, so that a situation with environment E' instead of E is inevitably a situation in which there is no duplicate of X. Another kind of case would be where the fundamental philosophical account of what it is for X to be in mental state S adverts to X's environment, but only in a very general way. The account might speak, for example, of 'whatever environmental feature is related in such–and–such a way to such–and–such an internal state I of X'.[6]

The suggestion here is that constitutive externalism might be true even though modal externalism is false because there might be a necessary connection between the internal properties of an individual and some external property that is constitutive of his contents, so that it is not possible for the external property to fail to obtain while the internal property does.

---

[6]See page 231 this volume.

This is a mistake. For what is an internal property? It is a non–relational property. If a property is non–relational, then it follows from our definition that it is logically possible for it to be instantiated independently of the existence of any other (non–abstract) objects. If an internal property is necessarily linked to an external one, then it is *ipso facto* not an internal, i.e., non–relational, property. An internal property can in principle be instantiated independently of any external property. This means that if strong local supervenience is true, externalism is false *tout court*. Nor is it any help to characterize conditions on contentful states in terms of some unspecified object, for if some object or other is required in order for the putative internal states to be duplicated, then by our definition they are not internal states. Thus, modal and constitutive externalism are not distinct. This is important because it shows that the only way to establish externalism is to show that modal externalism is true.

Thus, we should distinguish the following three positions: externalism, which holds that content properties are relational; internalism, which holds that they are not; and what I will call 'strong individualism', which holds that content properties are strongly locally supervenient. Externalism is incompatible with both internalism and strong individualism. Strong individualism entails internalism, but is not entailed by it.

The importance of distinguishing these three positions is that if we suppose that the alternative to externalism is strong individualism, and we are inclined, as I am, to think that strong individualism is false, then we will be forced to accept externalism. But we are not forced to choose between these options. Externalism can be false without our having to embrace strong individualism. Consequently, to establish externalism, we need to do more than to show that strong local supervenience is false.

Why would one be inclined to miss the possibility of denying externalism without embracing strong local supervenience? I suspect that it is an implicit commitment to reductive naturalism: a commitment to conceptually reducing content properties to other sorts of properties. If one thought that content properties had to be reduced (had to be, in this strong sense, *naturalized*), then if one denied that content properties were relational properties, one would be committed to claiming that they are reducible to non–relational properties, and so one would be committed to strong local supervenience.

There is nothing wrong with this procedure in a context in which it is taken for granted that content can be in this way naturalized. But then we should not suppose that any conclusion we draw can be detached from what it is conditional on; and we should be sensitive to

the possibility that our intuitions are being distorted by this possibly false background assumption. I think it is safe to say that no one has an argument to show that mental concepts are conceptually reducible to other sorts of concepts. The only way to show this would be to produce a successful analysis. No one has succeeded in doing this. As I see it, the reasons for this run deep: our concept of a point view, which is central to our understanding of what it is to have mental states, has no analogue outside the domain of the mental. If this is right, then both the physical externalist and the strong individualist share a common, false assumption.[7]

3 Problems for Externalist Thought Experiments

My remarks so far have been about how best to see the issues between the externalist and the internalist. Now I turn to a consideration of an argument that Martin Davies presents for externalism about perceptual content.[8] The argument, which is admirably laid out, I think helps to raise some general difficulties for attempts to show that content properties are relational properties. (This of course is not the *aim* of the argument.) I will raise two sorts of difficulty. The first is the most fundamental one.

Davies distinguishes between pure input–side theories, teleological theories, and three factor theories, which combine input, output and teleological factors. A pure input–side theory holds that perceptual content varies with the (regular) distal causes of an organism's internal states. A teleological theory holds that what is relevant to a creature's perceptual contents is the evolutionary history of the species of which it is a member. Davies argues against pure input–side theories and against teleological theories and in favor a three factor theory.[9] The decision in favor of the three factor theory is reached on the basis of considering our reactions to thought experiments in which we are asked to judge whether an individual in a counterfactual situation, who is non–relationally physically (or perhaps merely neurally) type identical to an individual in the actual

[7]This is not a problem for the social externalist, however, because the relations which she clams partially determine thought contents are themselves intentional states and events.

[8]Martin Davies, this volume, Section 3, pp. 239-242.

[9]I should note that Davies's commitment to a three–factor theory is tentative, and undertaken in the spirit of aiming to provide examples which give the strongest possible support to externalism.

or in another counterfactual situation, has the same perceptual contents. Let us call such thought experiments 'Twin cases'.

Consider first a pure input–side theory. While input–side factors seem essential to any externalist story, there are serious obstacles in the way of representing them as sufficient. The difficulty here is that if we allow content to vary just in relation to distal causes, without taking into account a creature's behavioral dispositions, it will be possible to describe cases in which the pure input–side theorist must say that a creature's states differ in content between one possible situation and another although the behavior of the creature in the second is intuitively inappropriate for it. In this case, it is natural to say that the creature has simply made a mistake. To take a familiar example, if the internal state of a frog which triggers the extension of its tongue were regularly caused (for the brief lifetime of the frog so unfortunately situated) not by flies but by BBs, the pure input–side theorist would have to say that the frog's perceptual states are about BBs.[10] The frog's behavior, however, is (we want to say) clearly inappropriate for it. It is natural to say that the frog is perceptually representing flies, or perhaps food on the wing, but at any rate not BBs. Thus, information about a creature's behavior, together with an independent conception of its goals, which makes that behavior inappropriate for it, dominates information about the distal causes of its internal states. Thus, a pure input–side theory appears to be inadequate.

That behavior is relevant to our intuitions in these cases apparently shows that no theory which omits output–side factors can be correct. Thus pure input–side theories are incorrect, as are pure teleological theories and two factor theories which combine input–side and teleological elements. But it is not just the creature's behavior in the example above that leads us to overrule information about the distal causes of its states, but the behavior together with an independent conception of the goals appropriate for the creature. The trouble is that the content assigned on the basis of input is not appropriate to the behavior produced, given the goals of the creature. What supplies the goals will be the evolutionary history of the species of which the organism is a member.[11] Thus, the best prospects for a

[10] I leave aside for the sake of argument the familiar difficulties that arise in trying to specify a uniquely relevant cause of the creature's internal representational state, difficulties which I believe can be shown to be insurmountable.

[11] Although I will not pursue the point here, it is, I think, extremely dubious that such talk of goals or functions grounded in facts about the evolutionary histories of species has anything to do with genuine intentionality. One of the

successful externalist theory seem to lie with a three factor theory, one which combines input, output, and teleological factors.

Such considerations dictate Davies's fundamental strategy in describing Twin cases for the purposes of establishing externalism:

> ... begin by considering a hypothetical creature x in possible situation w_1, and then imagine a (brain and central nervous system) duplicate y of x in a different situation w_2 such that:
>
> the distal causes of internal states are different; and
>
> the behavioural consequences of internal states are different; while
>
> there is 'harmony' between distal causes and behavioral consequences (input–output harmony); and (to satisfy teleological intuitions)
>
> this harmony is the product of evolution.[12]

I accept that no two–factor theory that combines just input–side and output–side factors is correct. If this is all the information that we provide in our descriptions of Twin cases, then we do not know enough to say whether the individuals in different situations have different perceptual or mental contents, or even whether they have perceptual or mental contents at all. Bringing in teleological factors, grounded in facts about the evolutionary history of the species of which an individual is a member, crucially adds information about the 'goals' appropriate for the organism or about the (biological) functions of various of its organs. It is an implicit background picture of this sort which I suspect has been driving externalist intuitions about these cases all along. Davies makes this explicit by building it into the description of the counterfactual cases he advances in support of externalism about perceptual content.

The question I now want to pose is whether adding teleological factors to the description of Twin cases in which input–output harmony is maintained across duplicates in different environments is any help to the externalist. I will argue that it is not. My concern will not be with the details of various thought experiments, and so I will not recount them here, but with how much information about a

---

salutary features of the advent of the Darwinian evolutionary theory was that it showed how we could explain *away* the appearance of goal–directedness in nature. The true lesson of the theory of evolution by natural selection is that talk of goals and functions in biology is simply a *façon de parler*. It is ironic that a theory that explains away the appearance of genuine goal–directedness should be invoked as the naturalistic ground for it.

[12] See p. 242 this volume.

creature's evolutionary history we have to add, and with the status of that additional information. When we see how much information we have to add to our description of Twin cases, we will see that intuitions based on the added information do not tell against internalism, as distinct from strong individualism, and so do not tell in favor of externalism.

Let us call a creature whose input and output are in harmony with its environment, in the sense that it survives and propagates in that environment, *well suited* to its environment. In the appropriate kind of counterfactual situation, we know these facts about a creature:

(1) it is well suited to its environment,

(2) its being well suited is a result of natural selection.

Consider an actual individual S with (by and large) veridical perceptual experiences in environment E, with input I, and output O. Suppose that S' is a counterfactual duplicate of S in environment E', with input I' and output O'. Suppose further that (1) and (2) are true of S'.

So far nothing follows about whether S' has different perceptual or mental states from S, or even whether S' has any mental states at all. One is well suited to an environment provided that one survives and propagates in that environment. But there is no contradiction in supposing both that that is true and that it is also true that S' has no mental states at all, even if it is a duplicate of S. Thus, we need to add something to our description of the counterfactual situation. We must add at least that S' has mental states and perceptual experiences. Thus, let us suppose that

(3) if (1) and (2) are true of a creature, then it has perceptual states.

But this is not yet enough. Even if (1), (2), and (3) were true of S', it would still not follow that S' had experiences that were different from those of S. For so far we have no reason to think that its perceptual states represent its environment correctly, and so no reason to think that the difference in environments makes for a difference in perceptual experiences. This would follow only if we added that

(4) if (1) and (2) are true of a creature, and it has perceptual states, then its perceptual states are by and large veridical.

Let us suppose that it does follow from the description of E and E' and (1)–(4) that S' has perceptual experiences which are different from those of S.

One might suppose that adding (4) is unnecessary because it is entailed by (1)–(3). It might be argued, for example, that natural selection guarantees that if an evolved creature had perceptual experiences, those experiences would be by and large veridical. But this would be a mistake. The thought behind such an argument would be that an evolved creature with perceptual experiences would have perceptual experiences as a result of natural selection, that the function of the perceptual experiences would be to provide information about the creature's environment, and that the creature's perceptual faculties would be optimally designed. However, none of this follows from (1)–(3). Not every feature of an organism is guaranteed to have a function for the organism, i.e., to have been selected for because of its contribution to reproductive success, and those that do are not guaranteed to be optimally designed. Thus, it does not follow from the fact that a creature is evolved, and that one of the mechanisms involved was natural selection, and even that it has perceptual experiences, that its perceptual experiences are themselves selected for. If its perceptual experiences are selected for, it still does not follow that they are selected for the purpose of providing information about the creature's environment. Furthermore, even if its perceptual experiences are the result of natural selection, and function to provide information about the creature's environment, it does not follow that its perceptual mechanisms are optimally designed, and so it does not follow that the creature's perceptual experiences are by and large veridical. All that is required is that an organism compete well enough with existing competitors to reproduce. (4) then does not follow from (1)–(3).

But perhaps we should not require this. We are, after all, describing a counterfactual situation. Let us just stipulate that (4) is true. If we stipulate that (4) is true, then in the counterfactual situation, S' will have different perceptual experiences than will S, since I' is different from I, and both S and S' have veridical experiences. Do we now have a thought experiment that establishes externalism? The answer is 'No'. For *stipulating* that (4) is true is not enough. Externalism is the view that relations to an individual's environment are constitutive of the contents of his mental states (perceptual experiences in this case). If (4) is not true necessarily, then the effect of stipulating that it is true in the counterfactual situation is to stipulate that in the counterfactual situation S's perceptual experiences are by and large veridical. This can be stipulated without contradiction. But this would not show that perceptual content properties were relational properties. From a description of an individual's environment, and the assumption that its perceptual experiences are

mostly veridical, one could infer what its contents were. Given this, if we can freely vary the nature of the environment, while keeping internal states fixed and stipulating that the subject's experiences are veridical, we can show that its experiences are not logically determined by its internal states. All of this, an internalist can admit. Stipulating that (4) is true is incompatible only with the claim that an organism's non–relational physical states determine its perceptual contents. Thus, it is incompatible with strong individualism. But as we have seen, strong individualism is not equivalent to internalism. We cannot get more out of this thought experiment, because in *stipulating* that the counterfactual individual's perceptual experiences are mostly veridical, we preclude the possibility of *explaining* their veridicality by appeal to constitutive relations between the individual's environment and the contents of his perceptual experiences. Since their veridicality is sufficient for them to be different in the counterfactual situation and in the actual situation, we need no other explanation for our judgment that the individual's experiences differ in the two situations. Thus, once we have built all of these assumptions into the description of the counterfactual situation, our judgment that the subject has different perceptual contents could not show that externalism is correct. At best it could show that strong individualism is false.

It is here that we see the importance of distinguishing strong individualism from internalism. If the only alternative to externalism were strong individualism, then the mere coherence of stipulating in the counterfactual situation that the subject's perceptual experiences were veridical (or that the subject had no perceptual experiences) would be enough to establish externalism. But since the falsity of strong individualism is compatible with the falsity of externalism, the possibility of varying content while internal properties remain the same is not sufficient to establish externalism. I conclude that adding to our description of the counterfactual situation facts about an individual's evolutionary history is no help to the externalist.

This should not come as a surprise, for evolutionary theory is a contingent scientific theory. It might have been false, and complete confirmation even of its central tenets is, as in the case of all scientific theories, an ideal that is reached only in the limit of scientific inquiry. Whether we have perceptual experiences with contents, however, is neither epistemically nor logically dependent on whether we are evolved beings. If it were, we could infer that evolutionary theory is true from knowing that we have perceptual experiences. We would have a transcendental argument for the truth of evolutionary theory. Unfortunately, the confirmation of evolutionary theory is not so easy.

Further, since externalism is a theory about our ordinary concepts of mental states (no one would deny that it is possible to introduce a concept of content which had the features externalists claim ordinary contents have), it should not appeal to evolutionary theory, since one clearly does not have to understand evolutionary theory to have, and to understand, the concept of a perceptual experience. As a matter of sound method, one should restrict the concepts one appeals to in one's analysans to concepts which could plausibly be available to anyone who had a full command of the concept one is analyzing.

I turn now to the second difficulty I want to raise for the kinds of thought experiment Davies advances. This difficulty can be developed in two parts. In the counterfactual situation we are to imagine, we want, minimally, input/output harmony. To achieve this, we have to imagine that some changes occur which make a difference to what behavior an individual produces in response to different distal causes. We can do this by imagining that in the counterfactual situation the physical laws are different than in the actual situation, or by imagining that the counterfactual individual's body is modified in some way to produce the change in behavior. Davies considers cases of both sorts.[13] Again, we need not recount the particular cases Davies considers. The criticisms I will advance apply to these cases in virtue of the strategies they employ, irrespective of the details.

Let us begin with the first case. In imagining that the laws are different in the counterfactual situation, we are imagining that fundamental physical laws are different, not merely that some derived laws have changed because initial conditions are different. In modifying features of the counterfactual individual's body, we change derived laws by changing initial conditions. So if all changing the laws came to were changing derived laws by changing initial conditions, this first approach would not differ from the second. Therefore, we must suppose that we are changing fundamental physical laws. The difficulty with this is that our individuation of fundamental physical properties depends upon what laws they figure in. If so, then in changing the physical laws in the counterfactual situation, one *ipso facto* changes the descriptions of the non–relational physical states of the individual. Thus, we cannot change the laws between the two situations we compare compatibly with presenting a case that would establish modal externalism.

Let us then imagine a counterfactual individual whose body differs from the actual individual. Davies notes that most strong individualists (who are his target here) will want to maintain that contents

[13]See pp. 243-244 above.

supervene on the non–relational states of the brain and the central nervous system. We can keep *that* the same while changing the body so as to modify behavior appropriately. That seems right. But the result is not enough to establish externalism. For the externalist thesis is that content properties are relational properties. And there is a position in–between strong individualism about neural states and externalism, namely, strong individualism about bodily states. The strong individualist about neural states holds that for any object O, if O is in complete non–relational *neural* state **S** and in mental state **M**, then, necessarily, for any object Q, if Q is in neural state **S**, then Q is in mental state **M**. The strong individualist about bodily states holds that for any object O, if O is in complete non–relational *bodily* state **S** and in mental state **M**, then, necessarily, for any object Q, if Q is in bodily state **S**, then Q is in mental state **M**. Both strong individualism about neural states and strong individualism about bodily states entail strong individualism (the second, indeed, is equivalent to it) and, hence, entail the negation of externalism. But strong individualism about bodily states does not entail strong individualism about neural states. Thus, if to maintain input/output harmony we modify an individual's body, while we might show that strong individualism about neural states is incorrect, we would not show that strong individualism about bodily states is incorrect. But if that is so, then we have not yet established externalism, since strong individualism about bodily states is incompatible with externalism. Thus, even waiving my first objection, we have not yet been presented with a thought experiment that can establish externalism if we require input/output harmony.

There are two strategies one could employ to respond to this difficulty. First, one could try to construct a thought experiment in which we change an individual's environment in a way that maintains input/output harmony without any changes in an individual's bodily states. Here is a very simple case of that kind.[14] Imagine a simple creature in a one–dimensional world, which we can think

[14]Davies gives a case which one might attempt to modify to produce the appropriate conditions, the binaural direction finder. (See above, pp. 246-247.) At first I thought of redescribing this case so that by simply changing the medium in which the binaural direction finder was located, we could change the mapping of external states to internal states and maintain input/output harmony. It turns out not to be so easy to specify in a simple way what would have to be the case for this to work. The case of the creature in the track world described in the text establishes the in–principle possibility for simple creatures. In either case, there would remain an enormous gap between such cases and a case which could be used to genuinely test our intuitions about content.

of as a 'track', equipped with a single light detector and two sound detectors separated in distance from one another. The behavior of the creature consists of moving along the track to the position of occurrences of 'lightning'. These are events which consist of the simultaneous emission of light and sound. The direction of the event is determined by which sound detector detects sound first. The distance of the event is proportional to the product of the speed of sound in the medium through which the sound travels, and the temporal separation of the detection of the light signal and that of the sound signal. We can assume that for practical purposes the light signal arrives with no delay. The velocity of sound in the track world is proportional to the density of the medium through which it travels. The amount of energy the creature spends in moving along the track is proportional to the temporal separation between the receipt of the light signal and the sound signal. The distance it moves along the track is proportional to the product of the energy it expends and the density of the medium. We can call the state the creature goes into upon detecting one or the other sound detector's firing first, and detecting a certain interval between the arrival of the light and sound signals, its perceptual state, one which represents the location of the event. Let us suppose that the events indicate the location of 'food' for the creature, and that this mechanism for finding food is the result of natural selection. And let us suppose that its moving to the location of the event represents input/output harmony. In this simplified case, it is possible to show that we can produce a counterfactual situation in which the creature's internal states remain the same, although its behavior changes and remains in harmony with the distal causes of that behavior. The factors that govern the behavior of the creature are represented in these equations:

(i) $V = c_1 \times p$

(ii) $D_e = c_2 \times I \times V$

(iii) $E = c_3 \times I$

(iv) $D_m = c_4 \times E \times p$

where V is the velocity of sound in the medium, p its density, D_e the distance of the event, I the interval between the detection of the light signal and the detection of the sound signal, E the amount of energy expended by the creature in moving, and D_m the distance the creature moves; c_1–c_4 are constants. The creature's behavior is in harmony with its input provided that $D_e = D_m$. Changes in the density of the medium will affect how distances are mapped

onto perceptual states. Input/output harmony is preserved across variations in the density of the medium provided that $c_1 \times c_2 = c_3 \times c_4$. Thus it is possible in this simple example to produce the right kind of counterfactual situation.

In this kind of case, of course, it would be implausible to attribute any mental states to our creature at all. The challenge for the externalist is to provide a convincing case in which the complexity of the creature and its interactions with its environment approach our own. I am skeptical that this can be done.

The second strategy that the externalist can employ is to relax the requirements Davies imposes on the thought experiments by not requiring that the Twin individuals be physically type identical but only that one keep fixed all internal states that could plausibly be thought to be relevant to the fixing of content. We could, for example, exclude all state types at a level of description which involves concepts which one need not have in order to have the concept of a perceptual experience. I think this strategy is the right strategy, and it is the one I recommend to the externalist. But it puts the additional burden on the externalist of explaining what internal features of an organism are not relevant to fixing its mental contents without begging the question against the strong individualist.

In closing, we can note that these results should be generalizable to arguments for externalism about any sort of content, not just perceptual content. The externalist about propositional attitude content who restricts his attention to input to the organism will face the same pressures to include in his account reference to the organism's behavior; for if the behavior is not in harmony with the assignments made on the basis of the input, we will be pulled to say that these are mistaken. But an input/output theory itself stands in need of some ground for the claim that the behavior displayed is goal–directed. This will drive the externalist to appeal to something like teleological facts, grounded in natural selection. As we saw, merely stipulating that a creature is evolved is insufficient to establish that it has perceptual states or that if it does they are connected at all with its environment; and if the description is appropriately strengthened, the conclusion does not tell against internalism (as opposed to strong individualism). Likewise, once we require input/output harmony in the counterfactual situations we compare, whether we are talking specifically about perceptual content or not, we will be faced with the difficulty of satisfying the two requirements that we have input/output harmony and that we do not change the non–relational physical states of the individuals across the situations we are comparing. These problems, then, face arguments for externalism not just about perceptual content but about any sort of content.