The Internal and External Components of Cognition

Ralph Wedgwood

In his chapter in this volume, Timothy Williamson presents several arguments that seek to cast doubt on the idea that cognition can be factorized into internal and external components. In the first section of my chapter, I shall attempt to evaluate these arguments. My conclusion will be that these arguments establish several highly important points, but in the end these arguments fail to cast any doubt either on the idea that cognitive science should be largely concerned with internal mental processes, or on the idea that cognition can be analysed in terms of the existence of a suitable connection between internal and external components.

In the second and third sections of the chapter, I shall present an argument for the conclusion that cognition involves certain causal processes that are entirely internal — processes in which certain purely internal states and events cause certain other purely internal states and events. There is every reason to think that at least a large part of cognitive science will consist in the study of these purely internal causal processes.

1 An Assessment of Williamson's Arguments

Williamson starts out by assuming, purely for the sake of argument, that there is some distinction that can be drawn between "internal" states of an agent, and "environmental" states. (Environmental states are states that are entirely independent of the intrinsic state of the agent and depend purely on the agent's external environment.) It does not matter for Williamson's

purposes exactly how this distinction is drawn, only that some such distinction can be drawn. (I shall have to return to this point, because it *will* matter for my purposes how exactly this distinction is drawn.)

On the basis of this notion of an "internal" state, Williamson then defines the notions of "broad" and "narrow" state. A state S is "narrow" if and only if "whether an agent is in S at a time *t* depends only on the total internal qualitative state of [the agent] at *t*, so that if one agent in one possible situation is internally an exact duplicate of another agent in another possible situation, the first agent is in S in the first situation if and only if the second agent is in S in the second situation." A mental state is "broad" if and only if it is not "narrow".

In section 1 of his chapter, Williamson rehearses some familiar arguments for the conclusion that practically all the mental states that we ascribe in ordinary folk-psychological discourse are in fact broad. It seems to me that these arguments are eminently plausible. It would be a rash philosopher indeed who sought to challenge Williamson's arguments at this point.

As Williamson points out, there are two main sources of the broadness of these mental states. Many mental states consist in having a certain type of *attitude* towards a certain *content*. Thus, one mental state consists in having the attitude of *belief* towards the content *that Tashkent is the capital of Uzbekistan*; another mental state consists in having the attitude of *hope* towards the content *that Turkey will join the European Union before 2015*; and so on.

In some cases, the broadness of these mental states is guaranteed by the type of *attitude* that they involve. This is the case with the so-called *factive* attitudes, such as the attitude of *knowing*. One cannot know that Tashkent is the capital of Uzbekistan unless Tashkent really is the capital of Uzbekistan: the content of the attitude of knowledge is always a content that is

true. (This is why some philosophers have claimed that the object of knowledge is a fact, rather than merely a true proposition.) Suppose that someone in another possible situation is internally an exact duplicate of me, but in that possible situation the capital of Uzbekistan is not in fact Tashkent at all, but Samarkand instead. In that case, even though in the actual situation, I count as *knowing* that Tashkent is the capital of Uzbekistan, the person in the other possible situation does *not* count as knowing this. Thus, the state of knowing that Tashkent is the capital of Uzbekistan is a broad state, not a narrow state.

The other reason why so many mental states are broad rather than narrow has to do with the *content* of those states. This point can be illustrated by means of the following example. Suppose that Twin-Earth is a planet far away from Earth that is qualitatively extraordinarily similar to Earth in all respects. Consider again my belief that Tashkent is the capital of Uzbekistan. There could surely be someone who is internally an exact duplicate of me but who lives not on Earth, but on Twin-Earth. On this person's lips, the name 'Tashkent' refers not to Tashkent, but to Twin-Tashkent (the corresponding city on Twin-Earth). So the content of the beliefs that that person expresses by utterances of sentences that involve the name 'Tashkent' cannot be the same as the content of the thoughts that I express by using the name. My beliefs are about Tashkent, and the other person's beliefs are about Twin-Tashkent, even though internally we are duplicates of each other. Thus the state of believing or thinking that Tashkent is the capital of Uzbekistan is a broad state, not a narrow state. The considerations that led to this conclusion will generalize to practically all the mental states that we ascribe in ordinary folk-psychological discourse. So it seems that practically all of these ordinary mental states are broad.

In sections 2 and 3 of his chapter, Williamson recapitulates some of the most original

arguments that he developed in a recent groundbreaking book (Williamson 2000). Here he argues for two main points. First, broad mental states play an important explanatory role (especially in explaining the "long term effects" of our mental states). Secondly, many of the mental states that play this important explanatory role are not just broad but *prime*: that is, they are not equivalent to any conjunction of a narrow state and an environmental state.

As I shall argue, Williamson's arguments for these points are entirely sound. Their conclusions are both important and true. Nonetheless, the conclusions of these arguments in fact imply much less than one might at first think. They are quite compatible both with a fairly robust version of "internalism" or "methodological solipsism" (see Fodor 1980) about cognitive science, and with the view that broad mental states can be illuminatingly analysed in terms of the existence of an appropriate sort of connection between internal and external elements. The internalist or methodological solipsist about cognitive science can happily accept all of the arguments that Williamson makes here.

I shall focus on Williamson's arguments for the claim that the factive attitudes (like the attitude of knowing that *p*) play a crucial role in certain causal explanations. If these arguments are sound, then it should be relatively straightforward to see how to adapt them into arguments for the claim that broad mental states of other kinds play a crucial role in causal explanations. The clearest sort of case in which such broad factive attitudes play an important explanatory role is in the explanation of actions that last a significant amount of time and involve the agent's interacting with his environment. To take one of Williamson's examples (1995), suppose that a burglar spends the whole night ransacking a certain house, even though by ransacking the house so thoroughly he runs a higher risk of being caught. Offhand, it seems quite possible that the

explanation of the burglar's acting in this way might have been that he *knew* that the house contained a certain extraordinarily valuable diamond.[1]

As Williamson points out, we would not get such a good explanation of the burglar's behaviour by appealing merely to the fact that he *believed* that the house contained the diamond. This is because one crucial difference between knowing something and believing something is that knowledge involves a *robust* connection to the truth. If you know a proposition *p*, then it is not only the case that *p* is true and you believe *p*, but it is also *not* the case that you might easily encounter some (misleadingly) defeating evidence that would lead you to abandon your belief in *p*.[2] Thus, if the burglar had merely believed that the house contained the diamond, he might easily have discovered evidence in the house that would have led him to abandon his belief that the diamond was there. For example, he might have believed that the diamond was in the house because he believed (i) that if the diamond had not yet fallen into his arch-rival's possession, then it could only be in a certain sugar bowl in the house, and (ii) that the diamond had not yet fallen into his rival's possession. Then he would not have ransacked the house at all: he would have fled the house as soon as he found that the diamond was not in the sugar bowl. Thus, the probability of the burglar's ransacking the house for the whole night given that he knew that it contained the diamond is greater than the probability of his doing so given only that he believed that it contained the diamond.[3]

A second alternative explanation of the burglar's behaviour might appeal, not just to the burglar's belief that the house contained the diamond, but to *all* of the burglar's background beliefs and all of his perceptual experiences during the course of his ransacking the house. It may well be that the probability of the burglar's ransacking the house given that he knew that it

contained the diamond is no higher than the probability of his ransacking the house given that he had all those background beliefs and perceptual experiences. But as Williamson points out, this second alternative explanation has a different defect. It is vastly *less general* than the original explanation. Even if the burglar had had a slightly different set of background beliefs or a slightly different sequence of perceptual experiences, so long as he had still known that the house contained the diamond, he would still have ransacked the house. This second alternative explanation is overloaded with too many specific details; these details are not really necessary to explain why the burglar ransacked the house. Thus, the original explanation, which appealed to the burglar's knowing that the house contained the diamond, seems preferable to this second alternative explanation as well.

It seems to me that this point will in fact generalize to most cases where we are interested in explaining *actions*. According to a plausible philosophy of action, which is due to Al Mele (2000), actions typically involve a causal feedback loop between (i) the agent's perception of what she is doing and (ii) her intentions to move her limbs in such a way as to realize her goals. Typically, the agent is perceptually monitoring her behaviour and its immediate environmental effects, and continually adjusting her behaviour so that its effects are in line with her goals. Moreover, practically all the actions that we are interested in explaining take more than an instant to be performed. (Think of such actions as cooking a meal, washing one's laundry, writing an email message, paying one's bills, and so on.) So the fact that one performs such an action is itself a fact about a complicated interaction between one's desires and perceptions, one's bodily movements, and one's environment. It is only to be expected that if the effect or *explanandum* involves this sort of interaction between an agent and her environment, the cause

or *explanans* will also involve the agent's relation to her environment. Thus, it is not surprising that broad mental states (such as the agent's perceptual knowledge of her immediate environment) will feature in the explanation of most actions.

For these reasons, then, Williamson's main point seems to be correct: an agent's factive attitudes, like the burglar's knowing that the house contains the diamond, may indeed play a crucial role in causally explaining the agent's behaviour. As I mentioned above, there seems to be no reason why this point should not also hold of other broad mental states as well, like my belief that this glass contains water, or that Tashkent is the capital of Uzbekistan. Broad mental states play a crucial explanatory role.

Although this central point seems to me correct, it does not entail the conclusion that Williamson seems to be seeking to support — namely, the conclusion that the "internalist" position that the only mental states that are important *for the purposes of cognitive science* are narrow states is false. Admittedly, once we accept that broad mental states play a crucial role in explaining behaviour, it certainly becomes plausible that these broad mental states will play an important role in certain branches of psychology and social theory. It certainly seems plausible that social psychology and the social sciences (including various forms of anthropology, sociology, economics, and political science) will find it useful appeal to such broad mental states.[4] It also seems overwhelmingly plausible that historians will invoke broad mental states in explaining historical events. But it is not clear that the same point will apply to cognitive science as it is usually understood.

Williamson says at the beginning of his chapter that cognitive science is "the science of cognition", and suggests that cognition is the "process of acquiring, retaining and applying

knowledge". But this is too specific to do justice to the broad array of inquiries that are pursued by cognitive scientists. A better statement of the goal of cognitive science would just be to say that it is to "understand how the mind works."

There are many sorts of ways in which one interpret this goal of "understanding how the mind works". But one sort of understanding that cognitive scientists are often interested in achieving is analogous to the understanding that one would have of a clock if one could identify each of its functional parts (its springs and cogwheels, its pendulum, and so on), and the way in which all these parts interact to bring it about that the clock has a reliable disposition to tell the correct time. As Hobbes (1651, Preface, §3) put it:

> For everything is best understood by its constitutive causes. For as in a watch, or some such small engine, the matter, figure and motion of the wheels cannot well be known, except it be taken insunder and viewed in parts; so to make a more curious search into the rights of states and duties of subjects, it is necessary, (I say, not to take them insunder, but yet that) they be so considered as if they were dissolved …

This sort of understanding of how clocks work is quite different from the understanding of clocks that one would have if one studied the impact of clocks on human society, or the economics of clock production, or the stylistic properties of ornamental clocks (from the standpoint of art history). An analogous understanding of how a computer works would involve an understanding of the structure of its electrical circuits and of the logical structure of its programming code. If this is the sort of understanding that cognitive science is particularly

interested in, that would help to explain why cognitive scientists are so interested in actually trying to *build* machines that can do some of the things that minds can do.

Thus, at least one of the goals of cognitive science will be to explain the *micro-level processes* that are characteristic of the mind. These are processes in which one mental event or state is caused by another mental state or event that precedes it as closely as one mental event can precede another. None of the examples of psychological explanations that Williamson focuses on are explanations of processes of this sort. These micro-level processes are precisely not processes in which one mental state causes "long term effects" by a complicated and extensive interaction between the thinker's mind and his environment. Thus, it is not clear that Williamson's arguments cast much doubt on the idea that the mental states that cognitive science is interested in will very often be narrow states.

Still, someone might think that Williamson's argument that these explanatorily important broad states are typically "prime" states (that is, they are not equivalent to any conjunction of narrow states and environmental states) establishes that broad states cannot be analysed in terms of any relation between narrow mental states and other non-mental factors. (Williamson himself does not claim that his argument establishes that broad states are unanalysable in this way, but some of his readers might think that his argument does show this.) If broad states are unanalysable in this way, then given the importance of broad mental states to the explanation of action, any science of the mind that ignores these broad mental states will, as Williamson puts it, "lose sight of the primary object of its study".

In fact, however, even if broad states are prime, they could still very well be analysable in terms of some relation between narrow states and non-mental factors. This is because a state is

prime just in case it is not equivalent to any *conjunction* of narrow states and environmental states. But obviously there could be an analysis of broad states that does not take the form of a conjunction.

In fact, many of the most promising attempts that philosophers have made on the project of analysing knowledge (to take just the most prominent example of a broad mental state that philosophers have sought to analyse) have not taken the form of conjunctions at all. Instead, they have taken the form of *existential quantifications*.

Thus, for example, at first glance it might seem that Nozick's (1981, chap. 3) analysis of what it is for an agent to know *p* is just a conjunction of a number of conditions. That is, it might seem that Nozick's analysis is this:

An agent knows a proposition *p* if and only if

(1) *p* is true,

(2) the agent believes *p*,

(3) if *p* were not true, the agent would not believe *p*, and

(4) if things were slightly different but *p* were still true, the agent would still believe *p*.

Conditions (3) and (4) are summed up by saying that the agent's belief in *p* "tracks the truth".

On closer inspection, it is clear, however, that when Nozick comes to present the most carefully considered version of his analysis, these four conditions fall within the scope of an existential quantifier. In the final version of his account, Nozick (1981, 179) offers first an analysis of what it is for an agent to know *p via method (or way of believing) M*:

> An agent knows *p* via method M if and only if
>
> (1) *p* is true,
>
> (2) the agent believes *p* via M,
>
> (3) if *p* were not true, and the agent still used M to arrive at a belief about whether (or not) *p* is true, the agent would not believe *p* via M, and
>
> (4) if things were slightly different, but *p* were still true, and the agent still used M to arrive at a belief about whether *p* is true, the agent would still believe *p* via M.

Then Nozick (1981, 182) uses this notion of knowing *p* via method M to define what it is for an agent to know *p simpliciter*:

> An agent knows *p* if and only if *there is some method M* such that (a) the agent knows *p* via M, and (b) if there are any other methods M_1 via which the agent believes *p* but does not know *p*, then these methods are "outweighed" by M.

Ignoring some of these complications, we may say that according to Nozick's analysis, for an agent to know *p* is just for there to be some method M such that M tracks the truth, and the agent believes *p* via M.

What has this to do with Williamson's claim that knowing *p* is a "prime" state? Let us assume — just for the sake of argument — that the state of believing *p* via a particular method M is a narrow state; and let us also assume that the state of being in a situation in which method M tracks the truth is an environmental state. Still, Nozick's analysis will guarantee that the state of

knowing *p* is not equivalent to the *conjunction* of any pair of narrow and environmental states of this sort. The reason for this is that there are usually many methods that one could use to arrive at a belief about whether (or not) *p* is true, and for almost all such methods, there are possible situations in which they track the truth, and other possible situations in which they do not.

Consider a simple model in which there are just two relevant methods, M_1 and M_2, and two relevant possible situations S_1 and S_2. Suppose that in both situations, S_1 and S_2, both method M_1 and method M_2 will lead one to believe *p*. However, in situation S_1, method M_1 tracks the truth while method M_2 does not; and in situation S_2, method M_2 tracks the truth while method M_1 does not.

Then, given Nozick's analysis, knowing *p* will not be equivalent to the conjunction of believing *p* via M_1 and M_1's tracking the truth — since one might know *p* even if one were not in this conjunctive state, if one believed *p* via M_2 in situation S_2, in which M_2 tracks the truth. Similarly, knowing *p* is also not equivalent to the conjunction of believing *p* via M_2 and M_2's tracking the truth — for one might know *p* even if one were not in that conjunctive state, if one believed *p* via M_1 in situation S_1, in which M_1 tracks the truth. Moreover, knowing *p* is not equivalent to the conjunction of believing *p* via either M_1 or M_2 in a situation in which either M_1 or M_2 tracks the truth, since one might be in that conjunctive state even if one did *not* know *p* — if one believed *p* via M_1 in S_2, in which M_1 does not track the truth (or if one believed *p* via M_2 in S_1, in which M_2 does not track the truth). And finally, knowing *p* is obviously not equivalent to the conjunction of believing *p* via either M_1 or M_2 in a situation in which *both* M_1 and M_2 track the truth, or to the conjunction of believing *p* via *both* M_1 or M_2 in a situation in which both M_1 and M_2 track the truth.

So it seems that Nozick's analysis of knowledge implies that knowing *p* is not equivalent to any such conjunction at all. At best, it is equivalent to an open-ended disjunction of conjunctions 'Either: believing *p* via M_1 while M_1 tracks the truth; or believing *p* via M_2 while M_2 tracks the truth; or …' But we are assuming here, for the sake of argument, that the state of believing *p* via method M is a narrow state, and being in a situation in which M tracks the truth is an environmental state. According to Nozick's analysis, it is states of this sort that determine whether or not one knows *p*. So, if knowing *p* were a "composite" state according to Nozick's analysis, then knowing *p* would surely be equivalent to a conjunction of a narrow state and an environmental state of this sort. Since, as we have seen, according to Nozick's analysis, knowing *p* is not equivalent to any such conjunction, we should conclude that according to this analysis, knowing *p* is prime, rather than composite. But of course, if Nozick's analysis of knowledge is correct,[5] then knowledge can be analysed by using such notions as "tracking the truth" and "believing *p* via method M" and the like.

Thus, Williamson's plausible claim that many broad mental states are prime, rather than composite, does not by itself show that the broad state of knowing *p* cannot be analysed along Nozick's lines (even if we assume that the state of believing *p* via method M is an entirely narrow state). So the claim that these broad states are prime does not show that these broad states cannot be analysed in terms of the existence of narrow states of certain kinds connected in appropriate ways to various non-mental environmental factors. If broad mental states could be analysed in this way, then it would not be justified to complain that an approach to cognitive science that did not explicitly mention these broad mental states would "lose sight of the primary object of its study". Thus, the claim that these broad mental states are prime does not support this

complaint.

2 The Argument from Hallucination

Even if the implications of Williamson's arguments are limited in the ways that I outlined in the previous section, this is hardly yet an argument in favour of any sort of internalism. In particular, it has not even been shown that there are any narrow mental states at all. Some readers may doubt whether there are any narrow mental states, given that — as Williamson argues, and as I have conceded — almost all (if not absolutely all) of the mental states that we ascribe in ordinary folk-psychological discourse are broad. In this section, I shall argue for the existence of narrow mental states. I shall do this by deploying a generalized version of the well-known "argument from hallucination" (see for example Mark Johnston 2004).

As I shall argue, the argument from hallucination is in effect a general type of argument; there are many specific arguments that are examples of this general type. I shall give two examples of the argument later in this section. The characteristic structure of these arguments is as follows. First, each of these arguments describes a pair of cases, which are externally highly dissimilar, but internally very similar. (For example, such a pair might consist of: (i) a case in which one genuinely perceives one's immediate environment, and (ii) a case in which one has a hallucination that one would not be able to distinguish from such a genuine perception of one's environment.) Then, the argument will try to make it plausible that there must be a mental state that is present in both cases. Since this mental state is present in both cases despite the enormous difference in the external features of the two cases, this makes it plausible that this mental state is

a narrow state.

Many philosophers have tried to criticize such arguments from hallucination.[6] But it seems to me that such criticisms at best undermine certain incautious formulations of these arguments. For example, some incautious formulations of the argument try to conclude, not just that there is a mental state that is present in both of the two cases that the argument focuses on, but that this is a mental state of a very special kind, with a very special object (such as a "sense datum") or a special sort of content (such as a special "narrow content" different from the sort of content that ordinary mental states have). As I shall formulate it, the argument from hallucination does not itself try to establish any of these further claims: its conclusion is simply that there is a mental state that is present in both of the two cases, neither more nor less.

Of course, if there is a mental state that is present in both of these two cases, it is natural to ask further questions about this mental state: What sort of mental state is this? And what is the relation between this mental state, which is present in both these two cases, and those mental states that are present in one but not the other of these two cases? However, there is a wide range of answers that could be given to these further questions. While it would indeed be an objection to the argument from hallucination if there were *no* plausible answer that could be given to those further questions, the argument itself is not tied to any specific answer to those further questions.

The first of the two examples of the argument from hallucination that I shall present here starts with a pair of cases that consists of (i) a genuine perception and (ii) a hallucination. (One of the differences between these two cases is that a perception is a factive state: if one perceives that *p* is the case, then *p* is the case: for example, if you see that the window is broken, then the window must indeed be broken.) Let us take the pair of cases that Mark Johnston (2004, 122)

invokes in his statement of the argument from hallucination. You are undergoing brain surgery, while quite conscious, under local anaesthetic. The surgeon "applies electrical stimulation to a well-chosen point on your visual cortex. As a result, you hallucinate dimly illuminated spotlights in a ceiling above you. … As it happens, there really are spotlights in the ceiling at precisely the places where you hallucinate lights." Then: "the surgeon stops stimulating your brain. You now genuinely see the dimly lit spotlights in the ceiling. From your vantage point there on the operating table these dim lights are indistinguishable from the dim lights you were hallucinating. The transition from … hallucination to … veridical perception could be experientially seamless. Try as you might, you would not notice any difference, however closely you attend to your visual experience."[7]

What does it mean to say that "from your vantage point", the dim lights that you see in the ceiling are "indistinguishable from the dim lights you were hallucinating"? It seems to mean this: you lack any reliable ability to respond to the hallucination by forming different beliefs and judgments from the beliefs and judgments that you would form in response to the genuine perception. And the reason why this is the case seems to be that in each of these two cases, you are disposed to form almost exactly the same beliefs and judgments — that is, the same beliefs (and the same doubts and uncertainties) about what is going on in your environment, about your own mental states, and so on.[8]

What can explain this remarkable fact that these two cases are so extraordinarily similar with respect to the beliefs and judgments that you are disposed to form in those cases? One plausible explanation is that there is a mental state that is present in both of these two cases, and it is this common mental state that disposes you to form those beliefs and judgments. As I noted

above, I do not have to take a definite stand on the further question of what exactly this common mental state is. Many different answers to this further question are possible. For example, one possible answer is that in this pair of cases, the mental state that is common to both cases might be an *experience as of there being dimly illuminated lights in a ceiling above you*.

Some philosophers deny that there is any common mental state. According to these philosophers the two cases involve fundamentally different mental states — in the one case a hallucination, and in the other a genuine perception; all that these cases have in common is that both cases involve the *disjunction* of these two mental states — that is, they both involve the disjunctive state of *either* hallucinating spotlights in a ceiling *or* seeing spotlights in the ceiling.[9] However, this "disjunctivist" response clearly fails to provide any explanation of something that surely cries out for explanation — namely, how it can be that these two cases are so similar with respect to the beliefs and judgments that one is disposed to form in those cases. After all, *any* two cases in an agent's mental life, no matter how dissimilar these cases may be from each other, will both involve the disjunction of some mental state involved in the first case and some mental state involved in the second. For example, consider one case in which I am in excruciating agony, and another in which I am listening to some beautiful music. These two cases have in common that they both involve the disjunctive state of *either* being in excruciating agony *or* listening to some beautiful music. But that the two cases have this much in common would hardly explain any other similarity that they might have (such as a striking similarity in the beliefs and judgments that one is disposed to form in those cases). Disjunctivism does not begin to engage seriously with the explanatory problem that is raised by the argument from hallucination.

The argument from hallucination can be generalized to other cases as well. In particular, it also be applied to two cases where your mental states differ in content. There are several different theories about what determines the reference of terms like our term ‘water’ and of the concepts that they express. According to most of these theories, such terms refer to the natural kind that actually causes the thinker (or members of the thinker’s community) to use the term in the thinker’s normal environment. Now suppose that you are transported from Earth to Twin-Earth in your sleep, and that you then remain on Twin-Earth for the rest of your life. At some point, it will be Twin-Earth, rather than Earth, that counts as your normal environment, and it will be a community on Twin-Earth, rather than any community on Earth, that counts as your community. At that point, then, your terms and concepts switch from referring to the objects and kinds of Earth to referring to the objects and kinds of Twin-Earth. But it is striking that you do not notice any switch in the content of your thoughts. This change seems to leave everything else about your mental states and dispositions unchanged. But that is an extraordinary fact. How can the contents of all your thoughts change so thoroughly and yet leave so much intact? You might even move back and forth between Earth and Twin-Earth several times, in which case the contents of your thoughts might change back and forth several times. How is it possible for such repeated cognitive revolutions to escape your attention?

The best explanation of this, it seems to me, is that there is a mental state that is common to both the Earth case and the Twin-Earth case. In saying that there is a “mental state” present in both cases, I just mean that there is a mental *property* that you have in both cases. I am not requiring that this mental property should take the form of standing in a definite mental relation to a particular content. Again, I do not need to take a definite stand on the further question of

what exactly this common mental property is. But one plausible answer to this further question may be that the common mental state is a state such as that of *believing a content of such-and-such a type*. Even if there is no such thing as "narrow content" — that is, even if all intentional contents depend on the thinker's relations to her environment — there may still be narrow *types* of content. That is, it may be that purely internal facts about the thinker are enough to determine that she is indeed believing a content *of such-and-such a type*, even though it is not enough to determine precisely *which* content of this type she is believing. (For example, for a content to be of such a narrow type might be for the content to be composed in such-and-such a way out of *concepts* of such-and-such types — such as concepts that have such-and-such basic conceptual roles. But it does not matter for my purposes exactly how these narrow types of content are defined — only that there are such narrow types of content.)

I shall suppose then that the argument from hallucination succeeds in showing that there is a mental state that is common to both cases in all these pairs of cases. But does it really show that these common mental states are *narrow* states? As I noted at the beginning of section 1, there is some initial unclarity about how exactly we should draw the boundary between internal states and external states. I suggest that we can use these pairs of cases that the argument from hallucination appeals to — the pair consisting of the case of genuine perception and the case of hallucination, the pair consisting of the case on Earth and the case on Twin-Earth, and so on — in order to clarify where this boundary between the internal and the external should be drawn. Admittedly, I have not given a precise account of what all these pairs of cases have in common. Giving such an account, it seems to me, would to require much further investigation (possibly including empirical psychological investigation); and I shall not try to anticipate the results of

such an investigation here. But to fix ideas, here is a suggestion that seems plausible, at least on first inspection: in each of these pairs of cases, the broad states are uncontroversially different between the two cases, but if the thinker shifts from one case to the other and back again, she will not notice any change; and the reason for this seems to be because all the thinker's mental dispositions are unaffected by the difference between the two cases (except of course the thinker's dispositions with respect to the broad mental states that differ between the two cases).

At all events, once we have a grasp on what these pairs of cases have in common, then we can just *stipulate* that the states that are present in both cases in all these pairs of cases all count as "internal states".[10] If a state is present in all these cases, despite the enormous difference of environmental states between all these cases, this makes it reasonable to call these states "internal states"; and a state that supervenes on these internal states is what I am calling a "narrow state".

At least when the notion of a "narrow state" is understood in this way, it seems to me that it is indeed plausible that the argument from hallucination provides a strong reason to accept the conclusion that there are indeed narrow mental states. As I noted above, this conclusion does not depend on the correctness of any particular answers to the further questions about what sort of states these narrow mental states are, or what their relation is to the broad mental states that are also present in these cases. But to fix ideas, it may be helpful for me to suggest some possible answers to these further questions. In answer to the first of these further questions, I have already suggested that these narrow states consist in standing in *non-factive mental relations* towards certain *narrow types of content*. For example, such narrow states would include: having an experience with a content of such-and-such a type; having a belief with a content of such-and-

such a type; and so on.

What about the second of these further questions? What is the relation between broad states and narrow states? For example, what is the relationship between the broad state of knowing *p* and the narrow state of believing a content of such-and-such a type (where the content *p* is in fact of such-and-such a type)? It seems plausible that the relationship is one of one-way strict implication: necessarily, if one is in the broad state of knowing *p*, then one is in the narrow state of believing a content of such-and-such a type; but the converse does not hold. This makes it plausible that the relationship is that of a *determinate* to a *determinable*, as the property of being scarlet is a determinate of the determinable property of being red, and the property of being an equilateral triangle is a determinate of the determinable property of being a triangle. Thus, for example, the relation of knowing is a determinate of the determinable relation of believing; the content *p* is a determinate of such-and-such a determinable narrow type of content; and the state of knowing *p* is a determinate of the determinable property of believing a content of such-and-such a narrow type.

3 Internal Causal Processes

Even if there are indeed narrow mental states, as I argued in the previous section, perhaps we cannot expect that the causal explanation of such narrow mental states will themselves appeal only to such narrow mental states. Perhaps broad states will always be needed to explain such narrow states. If so, then there does not seem to be much prospect for a form of cognitive science that focuses exclusively on narrow states. In this section, I shall argue that this is not so. There is

a large class of narrow mental states, and a certain sort of causal explanation of these narrow states, such that causal explanations of this sort will explain narrow states purely on the basis of other narrow mental states. The causal processes that are described by these explanations are what I shall call “internal causal processes”.

First, the explanations that I am concerned with are explanations of cases in which someone forms or revises their attitudes for a reason. For example, these would include explanations of why a certain agent forms a certain new belief or intention, or revises an old belief or intention, on an occasion on which the agent forms or revises her attitudes in this way for a reason. In a very broad sense, then, these are all explanations of pieces of *reasoning*. The piece of reasoning in question may be either theoretical reasoning (the upshot of which is that the agent forms or revises her beliefs), or practical reasoning (the upshot of which is that the agent forms or revises her intentions about what to do), or any other kind of reasoning that there may be. What this class of explanations excludes, then, are explanations of cases where an agent comes to have a mental state, but not for any reason — such as cases where an agent comes to feel thirsty, or to have a certain sensory experience (on the assumption that these are not mental states that the agent comes to have for a reason).

Secondly, the explanations that I am concerned with are explanations that seek to break down a process of reasoning into its *basic steps*. (As Hobbes would say, we are trying to understand the mental process’s “constitutive causes”.) A basic step of this sort would be a mental process that cannot itself be analysed, at the relevant level of psychological explanation, into any other mental sub-processes at all. Thus, suppose that there is basic step that leads from one’s having a sensory experience as of *p*’s being the case to one’s coming to believe *p*. Then

one's having this experience is (at least part of) the *proximate psychological explanation* of one's coming to hold this belief. There are *no intervening steps*, between the experience and the belief, that can be captured at the relevant level of psychological explanation.

In this section, I shall argue that in a case of this kind, if the *explanandum* consists of the fact that the agent acquires (or ceases to have) a narrow mental state, then the proximate explanation will always also consist in some fact about the agent's narrow mental states.[11] I shall argue for this in two stages. First, I shall argue that in any case of this kind, the proximate psychological explanation of an agent's acquiring a *mental* state is always some fact about that agent's *mental* states. Then I shall argue that when the mental state in question is a *narrow* state, then the proximate explanation of the agent's acquiring that state is always a *narrow* mental state of the agent.

In arguing for the first point, I am not denying that it is *ever* correct to explain the fact that an agent acquires a mental state through reasoning on the basis of something other than a fact about the agent's mental states. For example, the fact that I come to believe that Fermat's last theorem is true could surely be explained by the fact that I have been *told by a reliable informant* that Fermat's last theorem is true — even though the fact that I have been told by a reliable informant that Fermat's last theorem is true is not a fact about my mental states. This explanation may be quite correct. It just does not identify the *proximate psychological explanation* of my coming to believe that Fermat's last theorem is true.

Intuitively, it seems, if this is a correct explanation, there must also be a more detailed correct explanation, in which my coming to believe that the theorem is true is not directly explained by my being told by a reliable informant that Fermat's last theorem is true, but is

instead explained by some intervening fact about my mental states. For example, perhaps my coming to believe that Fermat's last theorem is true is explained by my having the *belief* that I have been told by a reliable informant that the theorem is true; and my having this belief (that I have been told by a reliable informant that the theorem is true) is itself explained by my having an *experience* as of someone (whom I take to be a reliable informant) telling me that the theorem is true.

Suppose that I claim that an agent's acquiring a certain belief is explained by a certain external fact that is not a fact about that agent's mental states; and suppose that the context does nothing to make it clear how there could be any more detailed correct explanation in which the link between that external fact and the acquisition of that belief is mediated by any intervening facts about the thinker's mental states. For example, suppose that I say, "I once lived in Edinburgh, so George W. Bush believes that I once lived in Edinburgh". It would be natural for you to reply, "But how does Bush know anything about you at all? Did you meet him and talk about your life? Did he have you investigated you by the CIA? Or what?" In asking these questions, you seem to reveal that you would not accept this explanation unless it is plausible to you that this link, between the fact that I once lived in Edinburgh and Bush's believing that I once lived in Edinburgh, is mediated by intervening facts about Bush's mental states.

In general, then, if an agent acquires a mental state through reasoning, the proximate psychological explanation of her acquiring this mental state on this occasion will be some fact about her mental states. In fact, it is plausible that this is one of the distinctive features of reasoning — the process of forming or revising one's mental states for a reason — in contrast to mental processes of other kinds: reasoning involves some change in one's beliefs or intentions or

other attitudes the proximate explanation of which is some other fact about the reasoner's mental states.

So far, I have only argued that the proximate explanation of an agent's acquiring a mental state through reasoning must involve some fact about the agent's *mental states*. I have not yet argued that if the *explanandum* consists in the fact that the agent acquires a certain *narrow* mental state through reasoning, the *explanans* must also consist in a fact about the agent's narrow mental states as well. Ironically, my argument will rely on the very same principle that Williamson relied on to defend the causal efficacy of the state of knowing *p*: if the *explanandum* consists of the fact that the agent acquired a certain narrow mental state, we will achieve a *more general* explanation by appealing to another fact about the agent's narrow mental states than by appealing to a fact about the agent's broad states.

In this second stage of the argument of this section, I shall rely on the idea that I suggested at the end of the previous section, that the relation between a broad mental state and the corresponding narrow state is the relation of a *determinate* to a *determinable*. Thus, for example, the broad state of knowing *p* is a determinate of the determinable narrow state of believing a content of such-and-such a type (where *p* is a content of the relevant type).

If narrow states are related to broad states as determinables to determinates, then it is plausible that whenever one is in a narrow state, one is also in some more determinate broad state. For example, whenever one believes a content of narrow type T, one either knows *p* or falsely believes *q* (where *p* and *q* are both contents of type T) or has some other broad state of this kind. Suppose that in fact one knows *p*. Thus, the event of one's coming to believe a content of type T occurs at exactly the same place and time as the event of one's coming to know *p*.

Some philosophers will want to conclude that these events are in fact identical. But I have been assuming that entities that enter into explanatory relations, either as the thing that gets explained (the *explanandum*) or as the thing that does the explaining (the *explanans*), are *facts* rather than events. It surely is plausible that even if the event of one's coming to believe a content of type T occurs at exactly the same time and place as the event of one's coming to know *p*, the fact that one comes to believe a content of type T is not the *same fact* as the fact that one comes to know *p*. After all, even though in fact both of these facts obtain, it could easily happen that the first fact obtains (you come to believe a content of type T) but the second fact does not (this belief does not count as a case of knowing *p*). Since they are distinct facts, I shall assume that they may have distinct explanations.

So, consider a case in which the *explanandum* — the fact that we are trying to explain — is the fact that an agent acquires a certain narrow mental state through reasoning. Specifically, suppose that this *explanandum* is the fact that the agent acquires a belief in a content of a certain narrow type T_1. Now consider two rival explanations of this fact. According to the first of these explanations, the agent acquires this narrow mental state because she is in a certain antecedent *broad* mental state — say, the state of knowing a certain propositional content *p*. According to the second explanation, she acquires this narrow mental state because she is in certain antecedent narrow state — where this narrow state is in fact a *determinable* narrow state of which the broad state cited in the first explanation is a determinate. Thus, if the broad state cited in the first explanation is the state of knowing *p*, the narrow state cited in the second explanation might be the state of believing a content of type T_2 — where the propositional content *p* is a content of type T_2, and knowing is a type of believing.

Now it seems quite possible that the fact that the agent is in the narrow mental state that is cited in the second explanation will be just as close to being causally sufficient for the *explanandum* as the fact that she is in the broad mental state that is cited in the first explanation. The probability that the agent will acquire a belief in a content of type T_1 is just as high given that she is in the antecedent narrow state of believing a content of type T_2 as the probability that she will acquire such a belief given that she is in the antecedent broad state of knowing *p*.

However, the second explanation will obviously be *more general* than the first. Consider a case in which you are *not* in the broad state that is cited in the first explanation, but you are still in the narrow state that is cited in the second explanation (which is a determinable of which the broad state cited in the first explanation is a determinate). Surely you would still acquire the narrow mental state (such as the belief in a content of type T_1), which is the fact that both explanations sought to explain. After all, the argument for postulating such narrow states in the first place — the argument from hallucination — was precisely that such narrow states were needed to explain certain striking similarities in the short-term mental effects of certain pairs of cases. Thus, I argued that there must be a narrow mental state present both in the case of hallucination and in the case of genuine perception because both cases had such similar short-term mental effects (so that it was possible for an agent to shift from the case of hallucination to the case of genuine perception without noticing any difference at all). Similarly, I argued that there must be a narrow mental state present both in the Earth case and in the Twin-Earth case to explain the striking similarities in the mental causal effects of the two cases (and to explain why one would not notice the contents of one's thoughts change as one is transported back and forth between Earth and Twin-Earth in one's sleep).

It is plausible that other things equal, we should prefer the more general of two explanations that otherwise count as equally good explanations of the same effect, from the same temporal distance. This point is especially plausible if the fact cited as the *explanans* in the more general explanation is a determinable of which the fact cited as the *explanans* in the less general explanation is a determinate. Here is a simple illustration of this point. Suppose that we want to explain why a certain code-protected door opened for the hero. One explanation that we could give would be to say that the door opened because the hero drew an equilateral triangle with each side measuring three inches, using her right index finger. A second explanation that we could give would be to say that the door opened because she drew a triangle. Now suppose that in fact *any* triangle drawn on the code-pad would have succeeded in opening the door. In that case, the second explanation is a better explanation, because it is more general than the first.[12]

For these reasons, then, it seems highly plausible that the proximate psychological explanation of these cases in which an agent acquires a narrow mental state through reasoning it itself always a fact about the agent's narrow mental states. This is not to say that broad states *never* play a role in psychological explanations. As we saw when I endorsed Williamson's arguments in section 1, knowledge does seem to play such a role in the explanation of certain *actions*. Here, however, the *explanandum* — such as the burglar's ransacking the house that contains the diamond — consists in an agent's interacting with his environment in a certain way. It is only to be expected that the *explanans* — the burglar's knowing that the house contains the diamond — will also consist in the agent's standing in a certain relation to his environment. This does not show that such broad states will figure in the explanation of the fact that the agent acquires a *narrow* mental state (such as the fact that a thinker comes to believe a content of type

T_1 at time *t*). A narrow mental fact of this sort is surely more likely to have a correspondingly narrow mental explanation. In general, the overall effect of the principle about explanation that I am appealing to here is that in any correct explanation there must be a certain sort of *proportionality* between the *explanandum* and the *explanans*. The *explanans* must be sufficient in the circumstances to produce the *explanandum*; but it also must not contain any irrelevant elements that could be stripped away without making it any less sufficient to produce the *explanandum*. This proportionality principle makes it plausible that narrow mental states will be particularly well placed to explain other narrow mental states, and broad mental states will be particularly well placed to explain other broad mental states (as well as actions, which like broad mental states also depend, in part, on the agent's relations to her environment).

For these reasons, then, it seems plausible to me that there are correct explanations in which both fact that is explained (the *explanandum*) and the fact that does the explaining (the *explanans*) are facts about the agent's narrow mental states. At least one branch of cognitive science could be devoted to ascertaining precisely which explanations of this sort are correct. This branch of cognitive science would be a theory of the nature of this kind of internal mental processes, which would be largely independent of the agent's relationship to her wider environment.

Moreover, as we saw in considering Nozick's definition of knowledge, Williamson's arguments do not show that broad states, like the state of knowing *p*, cannot be analysed in terms of the existence of appropriate connections between narrow states, and the non-mental objects, properties and kinds in the agent's environment. It may be that some such analysis can be given of all such broad mental states.[13] If that is the case, then a correct account of all the causal

relations, actual and counterfactual, among both the agent's narrow states and the non-mental objects, properties and kinds in the agent's environment, will in fact state all the same facts as a correct account that overtly appeals to broad mental states. In that case, a form of cognitive science that focused purely on explaining narrow mental states in terms of other narrow states (when supplemented by an account of the causal relations between these narrow mental states and various non-mental features in the agent's environment) would be able to capture everything that can be stated in terms of such broad states; and a cognitive science of this kind could not fairly be accused of "losing sight of the primary object of its study".

I cannot undertake to settle the question here of whether all broad mental states can be analysed in such terms. Even if broad mental states cannot be analysed in this way, a form of cognitive science that restricts itself to studying purely internal cognitive processes would still be investigating some pervasive and genuinely cognitive phenomena. But if broad mental states can be analysed in this way, then such a form of cognitive science could truly claim to be seeking the answer to the question of "how the mind works".[14]

Notes

1. See also Williamson (2000, 60–64 and 75–88). Of course, if we can give a non-circular definition of knowledge in terms of other folk-psychological notions — for example, if knowledge can be defined as a rational belief that is in a certain sense "reliable", as I believe — then knowledge would not play an *indispensable* role in any of these explanations. I shall touch on this question in the last section.

2. This point is due to Harman (1973, 143-44); for further discussion, see Wedgwood (2002a).

3.We would also not get such a good explanation of the burglar's behaviour by appealing to the fact that the burglar *truly believed* that the diamond was in the house. Even if the diamond was in the house (say, hidden inside the grand piano), the burglar might have believed that it was in the house only because he believed that it was in the sugar bowl, in which case he would still not

have ransacked the house for the whole night.

4. For example, see Diamond (1997, 143): “an entire field of science, termed ethnobiology, studies peoples’ knowledge of the wild plants and animals in their environment”.

5. Of course, this is a big ‘if’. Almost no one thinks that Nozick’s analysis is exactly right as it stands. But I actually think that an analysis that is at least a little like Nozick’s does succeed; see Wedgwood (2002a).

6. For such criticisms, see Dancy (1995) and McDowell (1994).

7. Johnston actually focuses on three cases: a hallucination whose content is false or non-veridical, a veridical hallucination, and a genuine perception. It seems to me however that this additional sophistication is not strictly necessary for the argument.

8. I say “*almost* exactly the same beliefs and judgments” because strictly speaking demonstrative judgments (such as the judgment that *those lights there* are dim) will be different in the two cases, as we can see from the fact that such demonstrative judgments will have different truth conditions in the two cases.

9. This is the view of a school of thought known as “disjunctivism”. For some canonical statements of this disjunctivist view, see Hinton (1973), Snowdon (1981), and McDowell (1994). For criticism of some of the arguments that are used to support this disjunctive view, see Millar (1996).

10. One question that requires further investigation is whether these “internal” states supervene on intrinsic features of the agent’s brain or whether their supervenience base must include something about the wider environment. It may be, for example, that a brain in a vat that had *never* been connected to a body that was capable of acting in a normal environment could not have any mental states at all. If so, then these “internal” states will not supervene on intrinsic features of the agent’s brain, but only on a *slightly* wider supervenience basis, which might include certain highly general and unspecific features of the agent’s environment. Nonetheless, the supervenience basis for these internal states would presumably be much narrower than the broad states that Williamson focuses on.

11. The argument that I give here is a generalization of an argument that I gave elsewhere (Wedgwood 2002b).

12. I owe this example to the editor of this volume. For some further discussion of this principle about why we should under certain circumstances prefer the more general causal explanations, see Yablo (1992a, 413–23, 1992b, and 1997).

13. Indeed, I believe that it can be argued that there would be something wildly puzzling and mysterious about broad mental states if they were not analysable in some such way. But there is not enough space to present this argument here.

14. This chapter was written with the support of a Research Leave award from the UK Arts and Humanities Research Board, for which I should like to express my gratitude. I should also like to thank Alex Byrne, Timothy Williamson, and Rob Stainton, the editor of this volume, for very helpful comments on an earlier draft.

References

Dancy, Jonathan (1995). "Arguments from Illusion", *Philosophical Quarterly* 45: 421–38.

Diamond, Jared (1997). *Guns, Germs and Steel: The Fates of Human Societies* (New York: W. W. Norton).

Fodor, Jerry (1980). "Methodological Solipsism Considered as a Research Strategy in Cognitive Science," *Behavioral and Brain Sciences*, 3: 63-109.

Harman, Gilbert (1973). *Thought* (Princeton, NJ: Princeton University Press).

Hinton, J. M. (1973). *Experiences* (Oxford: Clarendon Press).

Hobbes, Thomas (1651). *De Cive*, English edition ("*Philosophical Rudiments concerning Government and Society*"). (London: R. Royston).

Johnston, Mark (2004). "The Obscure Object of Hallucination", *Philosophical Studies* 120: 113–183.

McDowell, John (1994). *Mind and World* (Cambridge, Mass.: Harvard University Press).

———— (1995). "Knowledge and the Internal", *Philosophy and Phenomenological Research* 55: 877–93.

Mele, Al (2000). "Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance", *Philosophical Perspectives* 14: 279-300.

Millar, Alan (1996). "The Idea of Experience", *Proceedings of the Aristotelian Society* 96: 75–90.

Nozick, Robert (1981). *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press).

Putnam, Hilary (1975). "The Meaning of 'Meaning'", reprinted in his *Mind, Language and Reality: Philosophical Papers Vol. 2* (Cambridge: Cambridge University Press).

Snowdon, P. F. (1981). “Perception, Vision and Causation”, *Proceedings of the Aristotelian Society* 81: 175-192.

Wedgwood, Ralph (2002a). “The Aim of Belief”, *Philosophical Perspectives* 16: 267-297

———— (2002b). “Internalism Explained”, *Philosophy and Phenomenological Research* 65: 349-369

Williamson, Timothy (1995). “Is knowing a state of mind?”, *Mind* 104: 533-565.

———— (2000). *Knowledge and Its Limits* (Oxford: Clarendon Press).

Yablo, Stephen (1992a). “Cause and Essence”, *Synthese* 93: 403–49.

———— (1992b). “Mental Causation”, *Philosophical Review* 101: 245–80.

———— (1997). “Wide Causation”, *Philosophical Perspectives* 11: 251–81.