

Extended minds and prime mental conditions: probing the parallels

Zoe Drayson

Towards the end of the 1990s, two new forms of externalism about mental states were proposed. Both went beyond the earlier semantic externalism of Putnam and Burge in arguing that environmental factors are relevant to more than just the *contents* of mental states. One was Williamson's (1995, 2000) externalism about the attitudes to mental contents;¹ the other was Clark and Chalmers' (1998) externalism about the realizers of mental states.² These two positions are logically distinct, and have generated entirely separate research programmes. Williamson's externalism is part of his 'knowledge-first' epistemological programme which aims to get past Gettier-style problems for the analysis of knowledge by taking knowledge to be fundamental and unanalysable. Clark and Chalmers' externalism raises issues for philosophy of mind and cognitive science: they suggest that the boundaries of the mind need not map on to the boundaries of the human organism.³

The purpose of this paper is to explore hitherto unconsidered parallels between these two forms of externalism. The argument that Williamson uses to establish the indispensability of externalist propositional attitudes is, I suggest, the same argument used by Clark and Chalmers to establish the indispensability of externally-realized mental states. If one is persuaded of the existence of externalist propositional attitudes by Williamson's argument then, *ceteris paribus*, one should also be persuaded of the existence of externally-realized mental states by Clark and Chalmers' argument – and vice versa.

The first part of the paper describes the two forms of externalism in more details. In the second part, I propose that both Williamson's externalism and Clark and Chalmers' externalism can be

¹ Williamson (1995, 2000) takes Putnam and Burge to have shown that the contents of mental states can depend constitutively on the environment, and criticises attempts to isolate the 'purely mental' core of such externalist mental states. Williamson argues that if the content of a mental state can depend on the external world then so can the attitude to that content: beliefs and other propositional attitudes are genuine mental states constitutively dependent on the external world. (I elaborate his picture further in Section 1.)

² Clark and Chalmers (1998) acknowledge that their brand of 'active' externalism is very different to the 'passive' content externalism of Putnam and Burge, in which the external considerations are distal and historical. They argue that the human organism can be linked with an external entity to create a coupled system in which the relevant external features play an immediate causal role. (I provide further details of their position in Section 1.) Clark has gone on to develop these ideas in later work, but this paper will focus largely on the original (1998) co-authored paper.

³ For a discussion of the different varieties of externalism about the mental and their relation to externalism about justification in epistemology, see Carter, Kallestrup, Palermos and Pritchard (2014).

understood as responses to an issue about psychological methodology that faced previous forms of externalism: any psychological explanation involving an externalist mental state seems to be 'factorable' into internal and external components such that an internalist mental state alone can play the appropriate explanatory role. The third part of the paper shows how Williamson and Clark and Chalmers appeal to the same explanatory virtue of generality to get around this issue.

Williamson argues that knowledge-citing psychological explanations are more general explanations than (and therefore preferable to) merely belief-citing explanations, while Clark and Chalmers argue that their extended notion of belief allows for psychological explanations that are more general than (and therefore preferable to) those explanations involving the non-extended notion of belief. In part four, I explore the significance of these explanatory parallels and their implications.

1. Varieties of externalism about mental states

Externalist approaches to mental states are associated with the content externalism of Putnam (1975) and Burge (1986): what makes these mental states externalist in this sense is that their content is determined by factors external to the individual: when we think about water, for example, whether our thought refers to H₂O or XYZ can depend on the physical or social-linguistic environments. Both Williamson and Clark and Chalmers propose more radical forms of externalism about mental states.

When Putnam (1975) and Burge (1986) argued that the contents of mental states might depend constitutively on the external world, they were often met by attempts to analyse mental states into purely internal and purely external components. Williamson (1995, 2000) thinks that this tendency to expect such an analysis is based on a mistaken metaphysical and conceptual view, and proposes instead that we should conceive of the mind and the external world as dependent variables. On his picture, mental states like belief (and not merely their contents) can depend constitutively on the external world. Williamson further suggests that the mistaken conception of the mind and the external world as independent variables leads us to suppose that knowledge is a hybrid state, a function of a mental variable (belief) and an external variable (truth). If we conceive of mind and world as dependent variables, Williamson argues, then we can understand knowledge as a factive mental state: the fact that it is truth-entailing (and thus depends constitutively on the external world) doesn't make it any less mental. Knowledge, claims Williamson, is not hybrid but 'prime', in the sense that it cannot be decomposed as the conjunction of purely internal and purely external components.

Around the same time as Williamson was proposing his externalist approach to mental states, Clark and Chalmers were developing a different form of externalism about the mental. Clark and Chalmers (1998) argue that the physical realizers of some propositional attitudes need not be located within the brain (or any other part of the body) of the subject. While contemporary cognitive science focuses on the brain, Clark and Chalmers argue that there is nothing privileged about neural mechanisms: if an external entity is linked with the human organism in the right kind of two-way interaction, it can create a coupled system that governs behaviour in the same way as neural mechanisms do. Clark and Chalmers argue that despite the fact that some components of this coupled system are in the external environment beyond the 'skin and skull' boundary of the subject, we should consider the processes it performs as cognitive processes. They go on to claim that when someone is in a particular mental state, such as believing that p , this state can constitutively depend on features of the environment, when those environmental features are appropriately coupled with internal processes.⁴

The respective proposals of Williamson and Clark and Chalmers are both forms of externalism about the mental, broadly considered, but the varieties of externalism they propose appear entirely orthogonal. This appearance is confirmed by the fact that while both proposals have yielded much scholarship, the two externalisms are rarely discussed together. Despite the differences between the two forms of externalism under discussion, I suggest that there is a more general motivation that drives the arguments of both Williamson and Clark and Chalmers: they both want to reject a particular methodological principle of psychology, and they both rely on the same appeal to explanatory virtues to do so.

2. The rejection of psychological internalism

What is psychological internalism?

⁴ As I understand the extended mind argument in Clark and Chalmers (1998), the conclusion is only that mental states can supervene partly on factors external to the human organism: it is not the claim that mental states themselves are external. Such a conclusion would require an existing commitment to identifying mental states with token particulars which is not stated in the paper. Furthermore, Chalmers (2009, xv) is quick to dismiss the idea that the extended mind hypothesis requires any serious theoretical presuppositions: he proposes that it does not even require one to be a physicalist. In discussion, Clark suggests that thinking of mental states as token particulars offers the best explanation of how minds can extend, but I take this not to be a commitment of the original (1998) paper.

One charge often levelled at Putnam and Burge's content externalism was that it was at odds with the following methodological principle of psychological science:

Psychological Internalism (PI): causal explanations in psychology should only invoke internal states of the subject

This principle goes by a number of different names, sometimes with different ways of cashing out exactly what is meant by 'internal': proposals include the claims that internal states are those supervening on the physical state of the subject, those shared by a subject and their physical duplicate, and those states of a subject that are non-relational. Putnam refers to PI as the principle of *methodological individualism*: "no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed" (Putnam 1975, 220). Stich introduces PI as the principle of *psychological autonomy*: "the only properties and relations that may legitimately play a role in explanatory psychological theories are the properties and relations that a subject and its replica will share" (Stich 1978, 575). Kim combines Stich's principle with a metaphysical claim, but only the methodological claim need concern us here: "Internal psychological states [those supervening on the organism's synchronous internal physical state] are the only psychological states that psychological theory needs to invoke in explaining human behavior" (Kim 1982, 59).

It is important to be clear that PI is *not* a metaphysical claim about the nature of mental states: it is a methodological principle concerning the practise of psychological explanation. Of course, there can be metaphysical justifications of PI, such as intuitions about the metaphysics of causation, but there can be justifications of PI based purely on observations of scientific practise.

It is also important to note that PI is *not* a reductionist principle: it is not committed to the idea that to be in a psychological state is to token an internal representation, for example. Mental states (*qua* propositional attitudes) can be understood simply as relations between thinkers and propositions; as such, PI is neutral as to whether this relation should be cashed out in terms of dispositions, behaviour, causal roles, or internal representations, for example. PI merely insists that the relation supervenes on the internal state of the subject at the time.

Traditional problems for PI

Even before we consider the tensions between PI and content externalism, there seem to be obvious counterexamples to PI.

The problem is that some of our propositional attitude terms don't seem to pick out internal psychological states. Factive propositional attitudes such as knowing, remembering, and seeing appear in causal explanations of action, but seem to violate PI: they don't only appeal to internal states of the subject, but also to the truth of the propositional attitude's content. At least in paradigm cases, the truth of the content seems to be a matter external to the subject - regardless of whether we think about the internal/external distinction in terms of duplicates or supervenience. The obvious strategy for dealing with these cases is to claim that each factive propositional attitude is a combination of a non-factive propositional attitude and other factors, such that the non-factive propositional attitude involves only internal states of the subject and plays the appropriate role in psychological explanation. The mental state of knowledge, for example, would be factored into belief, the truth of the belief's content, and some form of justification. Of these factors, only the belief seems to be required for the psychological explanation, and as long as the belief supervenes on only the internal states of the subject then there is no challenge to PI. This 'factoring' approach allows that when we attribute knowledge and other factive mental states in psychological explanations, these propositional attitudes are eliminable in favour of their purely internal factors. As Stich puts it, "what knowledge adds to belief is psychologically irrelevant" (Stich 1978, 574).

Factive mental states are in tension with PI because they make essential reference to environmental factors. For the same reason, Burge and Putnam's content externalism entails that even non-factive mental states are in tensions with PI: when we give a psychological explanation of someone's action in terms of their belief that p , where p is determined by factors external to the individual, we violate PI. Defenders of PI suggest that when my duplicate and I have different belief contents as a result of different environmental factors (beliefs with different 'broad' content), we will share an internally determined belief content (beliefs with the same 'narrow' content). If this is the case, then the factoring approach outlined above can be applied to these scenarios: psychological explanations involving mental states with broad content can be factored into mental states with narrow content, plus the relevant external factors. The broad content is therefore assumed to be an eliminable part of the genuine psychological explanation, in which mental states with narrow content do the causal explanatory work.

In summary, PI claims that causal explanations in psychology should only invoke internal states of the subject: when confronted by explanations that appeal to factors external to the subject, PI suggests that we are not dealing with genuine psychological explanations. The standard defence of PI involves 'factoring' explanations into (a) genuinely psychological internal factors, and (b) non-psychological external factors. In the case of both factive mental states and mental states with broad content, the factoring approach works the same way. First, the propositional attitude is decomposed

into internal and external factors; then an internal factor is isolated that (i) would be shared by a physical duplicate, and (ii) plays the same psychological causal-explanatory role as the original propositional attitude.

New problems for PI

Many externalist approaches to mental states will be in tension with PI, and therefore be candidates for factoring into internal and external components. This is a problem for such proponents of externalism who want their externalist mental states to do explanatory work in psychology. The factoring approach suggests that as long as there are internal mental states that can do the explanatory work, the externalist mental states will be superfluous to the causal explanations of behaviour.

Both Williamson and Clark and Chalmers propose the existence of externalist mental states: Williamson argues that factive propositional attitudes like knowledge (and not merely their contents) constitutively depend on the environment, while still being genuine mental states; Clark and Chalmers argue that propositional attitudes can have physical realizers that extend beyond the bodily boundaries of the subject, while still being genuine mental states. Both Williamson and Clark and Chalmers argue that their respective externalist mental states play a causal role in psychology explanation, thus denying PI. The challenge for them both, I take it, is to show how their respective forms of externalist mental state avoid being factored into genuinely psychological internal states and external non-psychological factors.

Williamson challenges PI by arguing that factive propositional attitudes play a causal role in psychological explanation that cannot be adequately captured by a factoring them into internal and external constituents:

“The internalist cannot substitute ‘believe’ for ‘know’ in the explanation without loss [...] [K]nowing can figure ineliminably in causal explanations. It is causally efficacious in its own right if any mental state is.” (Williamson 2000, 64)

Clark and Chalmers challenge PI by arguing that some propositional attitudes play a causal role in psychological explanation that cannot be adequately captured by factoring them into internal and external constituents:

“The external features in a coupled system play an ineliminable role [...] The external features here are just as causally relevant as typical internal features of the brain.” (Clark and Chalmers 1998, 9)

It seems clear from these passages that Williamson and Clark and Chalmers take the same approach to the challenge of PI and the threat of factoring. Both are claiming that some mental states are externalist in a sense which defies the standard factoring approach, and thus denying that PI is the appropriate methodological principle for all psychological explanation. Furthermore, they both take the same tactic in order to defy the factoring approach: Williamson and Clark and Chalmers propose that their respective externalist mental states are not hybrids of psychological and non-psychological features, and thus can't be factored without explanatory loss. Furthermore, as I demonstrate in the following section, they use the same argument to achieve this end.

3. The appeal to explanatory generality

Williamson and Clark and Chalmers try to avoid the factoring approach that defends PI by claiming that their respective externalist mental states are explanatorily ineliminable in psychological explanation. The similarity does not end there, however, as they both argue for this ineliminability by making the same appeal to same explanatory virtues: both Williamson and Clark and Chalmers emphasise the way in which non-factored psychological explanations can be *better explanations* than those which adhere to PI. Williamson argues that “[t]he primeness of mental states is shown to contribute to their value in the explanation of action” (Williamson 2000, 76) (where prime should be understood as non-composite, not factorable), while Clark and Chalmers argue that “[b]y embracing an active externalism, we allow a more natural explanation of all sorts of actions” (Clark and Chalmers 1998, 9). Both Williamson and Clark and Chalmers argue that their respective brand of externalism provides a better explanation than one which is consistent with PI, on the grounds of explanatory generality.⁵

What is explanatory generality?

The generality of an explanation is the range of situations to which an explanation applies; the breadth of its applicability to different circumstances. If an explanation has enough generality, its *explanandum* rarely occurs in the absence of the *explanans*: the generality of an explanation varies inversely with $P[C | \sim D]$, the probability of C (the *explanandum*) conditional on $\sim D$ (the negation of *explanans*). Consider the question ‘Why does he have food-poisoning?’. The explanation “because his chicken was undercooked” is better than “because his *Poulet a la bretonne*, simmered with navy

⁵ Whether or not these externalist explanations are in fact better explanations than the factored alternatives will be discussed in the final section.

beans, beets and bacon in apple cider, was undercooked”, because the food-poisoning was dependent upon the chicken’s being undercooked rather than on how it was served and what it was accompanied by.

Generality often has to be traded off against accuracy. One explanation has more accuracy than another if its *explanans* better raises the probability of its *explanandum*: the accuracy of an explanation varies with $P[C|D]$, the probability of C (the *explanandum*) conditional on D (the *explanans*). Again, consider the question ‘Why does he have food-poisoning?’. The answer “because his chicken was undercooked” is a better explanation than “it was something he ate”.

Generality and accuracy may not be the only virtues an explanation can have, but they are the most relevant ones for the explanations under consideration here. There are also complex considerations with regard to how one explanatory virtue gets traded off against another, but what matters for our purposes is one key claim about the relationship between accuracy and generality: that if we have two equally accurate explanations, we should (*ceteris paribus*) prefer the more general of the two.⁶ In what follows, I show how the appeal to explanatory generality can be found in Williamson’s argument and in that of Clark and Chalmers.

Williamson: the burglar

Williamson (2000, 62) gives us the example of a burglar who spends all night ransacking a house, risking discovery by staying so long. Williamson claims that the best explanation of the burglar’s actions is that the burglar *knew* there was a diamond in the house. This explanation clearly violates PI: factive propositional attitudes depend not just on the internal state of the subject but also on the state of the external world. A proponent of the factoring approach would argue that PI could be preserved if we factored the explanation into internal psychological factors and external non-psychological factors: they would argue that the burglar’s belief (considered as a genuine psychological state that supervenes on internal states) plays the appropriate explanatory role in the explanation of the burglar’s actions, and that appealing to the burglar’s knowledge adds nothing to the psychological explanation.

Williamson doesn’t think that the factoring approach will work in this case. He proposes that the burglar’s mere belief that there is a diamond in the house doesn’t explain the length and persistence

⁶ What I am calling ‘accuracy’ here is similar to what Williamson refers to as ‘sufficiency’. Molyneux (2007) suggests that Williamson more often argues that explanations involving prime conditions have greater sufficiency, and more rarely uses the argument from generality.

of his search: mere beliefs are subject to revision in a way that knowledge is not. One might argue, however, that these aspects of the burglar's search are not the appropriate *explananda* for a psychological explanation. Once we combine a psychological explanation of the burglar's behaviour in terms of his (internal) belief with the non-psychological fact that the belief is true, we'll have a satisfactory explanation of the burglar's lengthy and persistent search without having to appeal to a mental state that depends on factors external to the individual. Williamson thinks that this way of factoring knowledge into (internal) belief and (external) truth still doesn't account for the burglar's behaviour, on the grounds that even true beliefs lack the stability of knowledge. True beliefs can be based on false premises which can be reconsidered: if the burglar's true belief that the diamond was in the house was based on the prior false belief that the diamond was under the bed (while the diamond was in fact in the cupboard), then upon not finding the diamond under the bed, the burglar would have reconsidered his belief that the diamond was in the house. Notice, however, that this does not demonstrate that knowledge is a prime mental condition. Perhaps we just haven't yet figured out the appropriate way to factor knowledge into psychological and non-psychological components.

The proponent of the factoring approach will insist that there is an appropriate way to factor knowledge: won't there always be a maximally specific knowledge-entailing conjunction of internal and external factors that provide the appropriate explanation? If so, then when Williamson claims that the burglar's knowledge state explains the action, that knowledge state will be entailed by the conjunction of the complete physical specification of his internal state and the complete physical specification of the rest of the world (assuming supervenience physicalism). The proponent of factoring can insist that if we replace 'knowledge' in any *explanans* with this knowledge-entailing conjunction, then we will always be able to factor it into internal and external components. In other words, there will always be an explanation of behaviour that is consistent with PI.

This is where Williamson's appeal to explanatory generality comes in. Williamson argues that while such a conjunction of internal and external specifications would provide as *accurate* an explanation of the burglar's behaviour as the knowledge-based explanation, that explanation would be less *general*. He points out that many features of the maximally specific conjunction will be "quite irrelevant" to whether the burglar's behaviour obtains:

"They will concern physical events that form no part of the causal chain between the agent's initial mental state and the final performance of the action. The agent would have performed the action anyway, even if those features had been different. Their inclusion is a defect in the explanation."
(Williamson 2000, 81-82)

The complete physical specification of the burglar's internal state will make reference to the neurons controlling his eye saccades, for example, while the complete specification of the external environment will make reference to the refraction properties of the diamond. Both of these factors could have been different, it seems, without making any difference to the length and persistence of the burglar's search. This is what motivates Williamson to claim that the knowledge-based explanation will be a better (more general) explanation than one which factors knowledge into internal and external components. The result is the claim that a knowledge-based explanation is a genuinely psychological explanation that violates PI.

Clark and Chalmers: Otto and his notebook

Clark and Chalmers' (1998, 12-13) appeal to explanatory generality is presented less explicitly than Williamson's appeal, but can be found in their thought-experiment concerning Otto, who has memory problems. Otto writes important information in his notebook, which he keeps with him and consults frequently. When Otto wants to go to the museum, he looks in his notebook and reads that the location of the museum is 53rd Street. I take it to be uncontroversial that *after* consulting the notebook, Otto has the occurrent belief that the museum is on 53rd Street. The key step in Clark and Chalmers' argument is their claim that *before* he consulted the notebook, Otto had the *dispositional belief* that the museum is on 53rd Street, and that this dispositional belief is the explanation of why Otto walks to 53rd Street when he wants to go to the museum. But notice that this explanation is inconsistent with PI, because a physical duplicate of Otto who lacked such a notebook would not be attributed the same dispositional belief. Explaining Otto's behaviour in terms of his dispositional belief about the museum location is to appeal to a psychological state that is realized in part by factors external to the subject in the sense that they lie beyond the skin/skull boundary.

The proponent of PI might simply deny that Otto has a dispositional belief about the museum before consulting his notebook. Clark and Chalmers make their case for attributing such a belief to Otto by comparing his case to that of Inga, who has a typical memory. When Inga wants to go to the museum, she consults her memory and heads to 53rd Street. But wouldn't we generally accept that Inga has the dispositional belief that the museum is on 53rd Street before consulting her memory? If we'd explain why Inga, given her desire to go to the museum, headed to 53rd Street in terms of her dispositional belief that the museum was on 53rd Street, then it seems that we should offer a parallel explanation for Otto.

As in the Williamson case, the proponent of the factoring approach will insist that there is a way to preserve PI. They might argue that Otto's dispositional belief that the museum is on 53rd Street is not a genuine psychological state, but rather a hybrid of a wholly internal psychological state (Otto's belief that the museum's location is recorded in his notebook) and a non-psychological fact (that the notebook records the museum as being on 53rd Street). By comparison with the Williamson case, they could argue that wherever Clark and Chalmers propose psychological explanations relying on 'extended' mental states, there will always be an alternative explanation that appeals to the maximally specific conjunction of internal and external conditions.

Clark and Chalmers acknowledge a version of this threat:

"The alternative is to explain Otto's action in terms of his occurrent desire to go to the museum, his standing belief that the Museum is on the location written in the notebook [psychological] , and the accessible fact that the notebook says the Museum is on 53rd Street [non-psychological]" (Clark & Chalmers 1998, 13, my insertions)

Clark and Chalmers' response to the threat of factorisation, I suggest, is the same as Williamson's: the conjunction of internal and external conditions might be equally *accurate* as an explanation, but an explanation that appeals to the (internal) belief that the location is written in the notebook is not sufficiently *general*. To see this, consider that Otto would still have headed to 53rd Street if the address has been circled on a map in his notebook rather than written down, or if the address had been tattooed on his arm, or stored in his neural memory as in Inga's case. This would suggest that an explanation of Otto's behaviour in terms of his belief that the museum is at the location written in the notebook would be less general than an explanation in terms of his belief that the museum is on 53rd Street. Clark and Chalmers' proposal, including its commitment to Otto's 'extended belief', would therefore be a better explanation.

While Clark and Chalmers don't explicitly formulate their argument against factorisation in terms of explanatory generality, the key moves are there. They argue that the alternative explanation offered by a factored approach takes "one step too many": "It is pointlessly complex [...] In an explanation, simplicity is power." (Clark and Chalmers 1998, 13-14)

They propose that by positing a mental state that depends on factors external to the individual, their equally accurate explanation has more generality:

"By using the 'belief' notion in a wider way [...] [t]he notion becomes deeper and more unified, and is more useful in explanation." (Clark and Chalmers 1998, 14)

Clark & Chalmers claim that the extended-belief explanation is a better explanation than one factored into internal psychological states and external non-psychological states. If correct, the extended-belief explanation would be a genuinely psychological explanation that violates PI.

4. The significance of the shared appeal to explanatory generality

I have argued that Williamson and Clark and Chalmers rely on the same appeal to explanatory generality to make their respective cases for externalist mental states. Williamson's argument that knowledge is a prime (i.e. unfactorable) mental condition relies on his claim that knowledge plays an ineliminable role in psychological explanation. Similarly, Clark and Chalmers' argument for attributing to Otto a dispositional belief about the location of the museum (before he consults his notebook) relies on their claim that this dispositional belief plays an ineliminable role in psychological explanation. In both cases, the perceived opponent is the defender of PI who proposes to factor the mental states in question into genuinely psychological internal conditions and non-psychological external conditions. By arguing that the explanatory role of certain mental states cannot be factored into internal and external components without loss of explanatory generality, Williamson and Clark and Chalmers are attempting to avoid the factoring approach and reject PI.

The parallels between the arguments of Williamson and Clark and Chalmers are significant for the following reasons. If the argument from explanatory generality works in Williamson's case to establish that the burglar's knowledge can't be factored into internal and external conditions, then *ceteris paribus* it also works in Clark and Chalmers' case to establish that Otto's dispositional belief can't be factored into internal and external conditions. And vice versa: if the argument from explanatory generality works in Clark and Chalmers' case to establish that Otto's dispositional belief can't be factored into internal and external conditions, then *ceteris paribus* it also works in Williamson's case to establish that the burglar's knowledge can't be factored into internal and external conditions.

Consider first the implications of this claim for Williamson. Williamson wouldn't be able to claim that Otto's dispositional belief is really a hybrid of an internal mental state and external facts about the notebook without denying the argument from explanatory generality that grounds his own claims about the primeness of knowledge. But it's not clear why the location of the realizers of Otto's dispositional belief would be a matter of concern Williamson in the first place. The extended mind argument is aimed at challenging the intuitions of those people who think that mental states are located in the brain of the subject, and Williamson has no such commitments. He has a non-

reductive view of mental states, and thinks of mental states as states that *people are in*, not states that are *in people*: “Subjects are in mental states, not vice versa” (Williamson 2009, 331).

Now consider the implications of the shared appeal to explanatory generality for Clark and Chalmers. They would not be able to claim that the burglar’s knowledge is really a hybrid of an internal mental state and environmental factors without denying the argument from generality that grounds their own claims about Otto’s dispositional belief. This in turn would commit them to the idea that the causal role of knowledge in psychological explanation can’t be reduced to the causal role of belief, whether or not beliefs supervene on the internal states of the subject. Clark and Chalmers would thus have to accept that if Otto *knows* that the museum is on 53rd Street, then his knowledge could play a role in psychological explanation that is distinct from the role played by his merely believing (or believing truly) that the museum is on 53rd Street. This is consistent with Clark and Chalmers’ original use of the Otto thought experiment, which establishes (at most) that psychological states can supervene on feature of the environment external to the human organism.⁷

But since the original Clark and Chalmers (1998) paper, Clark has developed his own work on the extended mind in a more reductionist way. Clark works on the assumption that to be in a mental state is to token a (physically implemented) mental particular: a vehicle of cognition. In fact, Clark even refers to his own position on extended mental states as ‘vehicle externalism’.⁸ The motivation for positing vehicles is the assumption – widespread in cognitive science – that mental contents can only play a causal role via their vehicles. Vehicles are normally thought of as physical or functional states of the brain, but Clark wants to claim that these vehicles can be external to the human organism. Whereas the original (1998) argument for the extended mind concluded that Otto’s believing that *p* need not supervene on his body, Clark’s vehicle externalism concludes that Otto’s notebook (rather than one of Otto’s brain states) is the vehicle of his belief content.⁹ If Clark had to accept Williamson’s argument for the primeness of knowledge, he would have to accept Williamson’s conclusion that the causal role of knowledge-states in psychological explanation is distinct from the causal role of belief-states in psychological explanation. For Clark, this would require that there was a difference between the respective vehicles of belief and knowledge. But the

⁷ For further discussion of whether Otto’s extended beliefs would in fact qualify as knowledge, see Pritchard (2010) and Clark (2015).

⁸ Clark (2005, 1) claims that the extended mind hypothesis is really a claim about extended vehicles, where a vehicle is the material bearer of content that enables a system to possess a contentful mental state. Chalmers (2009, xv), however, is clear that their original argument for the extended mind does not even require a commitment to physicalism.

⁹ In other words, Clark is committed to the idea of token mental states. The existence of token mental states is precisely what Williamson is denying when he writes that “[s]ubjects are in mental states, not vice versa” (Williamson 2009, 331). Williamson thinks that subjects are in types of mental states, and not in virtue of having token mental states, as his (2009) reply to Jackson makes clear.

way in which the causal role of knowledge differs from that of belief, according to Williamson, is (at least partly) in terms of the truth of the content. On vehicle-based accounts of cognition like Clark's, semantic features of the content such as truth-values are causally inert: only physical or functional properties of the vehicles can make a difference. This is the key difference between Williamson's non-reductive approach to the mental and Clark's mechanistic approach to the mental. Presumably anyone others committed to a mechanistic approach to cognition would deny Williamson's claim that knowledge is a prime mental condition, but for Clark there is a particular problem: he has to reject Williamson's argument from explanatory generality in such a way that doesn't also reject his own argument from explanatory generality, upon which the extended mind hypothesis rests.

Furthermore, it is not only the primeness of *knowledge* that Clark would have to worry about, since Williamson thinks that a similar argument from explanatory generality can be used to establish the primeness of other mental states. If mental states in general are non-composite, they *a fortiori* can't be identified with individual components of a cognitive system: mental states are not token particulars or vehicles, either internal or external. On this view, Inga's dispositional belief about the museum is no more internally located than Otto's.

This leaves Clark's vehicle externalism in an interesting place. It looks like philosophers of cognitive science who are committed to the idea of mental states as mechanistic vehicles will have to reject Williamson's claim that knowledge is a mental state. But if doing so involves rejecting the argument from explanatory generality, then they also reject the argument for Otto's extended belief. One solution for proponents of vehicle externalism would be to argue for extended cognition in a way that doesn't rely on the argument from explanatory generality.¹⁰

Possible responses

On the basis of the parallels between Williamson and Clark and Chalmers' reliance on the argument from explanatory generality, I have suggested that if you accept the generality-based argument for one of these two forms of externalism, then *ceteris paribus* you should accept the generality-based argument for the other. To avoid the consequences outlined above, one could focus on the *ceteris paribus* clause to deny that the two arguments are relevantly similar by pointing to premises of

¹⁰ Some of the arguments in the opening stages of Clark and Chalmers (1998) make a more mechanistic case for some form of extended cognition. I've argued elsewhere (Drayson 2010) that the original paper seems to develop two distinct positions: one is vehicle externalism, supported by examples from cognitive science; the other is the extended mind hypothesis, supported by the Otto-Inga thought experiment.

either argument that are not shared by the other, for example.¹¹ But in the absence of any obvious candidates, an alternative response would be to focus on the argument from generality.

It might be suggested that Williamson does not rely on the explanatory generality argument to demonstrate the primeness of knowledge. It is true that he also appeals to metaphysical considerations to argue that we should not expect such concepts to be factorable, but it is not clear what the relationship is between the metaphysical argument and the explanatory argument. And notice that even if Williamson himself takes explanatory claims to be secondary to metaphysical claims, Clark and Chalmers need not see it this way: their own (1998) paper appears to be an argument from explanatory claims to metaphysical claims. Alternatively, one might argue that Williamson's argument from explanatory generality doesn't work in the first place. Perhaps the level of generality provided by Williamson's knowledge-based explanations could be too high to be genuinely explanatory, as Yablo (2003) suggests. And even if Williamson is right to claim that knowledge can't be replaced in an explanation by other conditions without explanatory loss, this wouldn't necessarily make knowledge explanatorily ineliminable: as Molyneux (2007) argues, explanatory loss of one sort might be compensated by another other sort of explanatory gain.

Another possibility is that Clark and Chalmers do not themselves rely on the argument from explanatory generality. Their argument appeals to explanatory simplicity and usefulness rather than *explicitly* to the generality of explanation. I suggest, however, that they need the specific virtue of generality in order to make the strongest case for the extended mind: explanations factored into internal and external components will be able to compete with extended mental states on some readings of simplicity and usefulness.¹² Wikforss (2014), for example, suggests that Clark and Chalmers may be *oversimplifying* the explanation of Otto's behaviour.¹³

While all this suggests that both Williamson and Clark and Chalmers are reliant on the argument from explanatory generality, it's worth considering whether the same notion of generality is at work

¹¹ It might be argued that Clark and Chalmers (1998) assume the truth of functionalism about mental states, which Williamson would deny. But while Clark certainly relies on functionalism in later work, this commitment is not endorsed in the original paper, and is explicitly denied by Chalmers (2009, xv-xvi).

¹² I would also suggest that without the argument from generality, the Otto thought experiment seems to rely more on similarities in function between Otto's belief and Inga's belief, which risks committing the extended mind hypothesis to functionalism. As previously discussed, it is important to Chalmers (2009) at least that the argument is not thus committed.

¹³ Wikforss (2014) suggests that Otto's actions have a higher level of complexity than Inga's, and thus require a higher level of complexity in their explanation.

in both approaches.¹⁴ A related concern is whether the two approaches are really dealing with the same opponent: while both oppose explanations that factor into internal and external constituents, it seems that there are different notions of the internal/external distinction in play. For Clark and Chalmers, the distinction is a locational one, with ‘skin and skull’ of human organism as its boundaries. Clark and Chalmers allow that internally located mental states could have content that was constitutively dependent on the external environment.¹⁵ For Williamson, has no commitment to the location of mental states, these would be externalist mental states.¹⁶

I accept that the more we study the specific details of the respective arguments made by Williamson and Clark and Chalmers, the more differences we will find. As I acknowledged at the beginning, the two approaches are largely orthogonal in many respects. I want to suggest that we can gain a new perspective on the relationship between the two approaches by understanding them as having a shared opponent in the proponent of PI, whom they attempt to defeat by appealing to the same explanatory virtues.

Acknowledgements

This paper was written for the ‘Extended Knowledge’ conference at the University of Edinburgh in April 2015, and subsequently presented at the Australian National University. Many thanks to both audiences for their questions and comments, and also to an anonymous reviewer who provided helpful feedback on the written version.

References

- Burge, Tyler (1986). Individualism and psychology. *Philosophical Review* 95:3–45.
- Carter, Kallestrup, Palermos and Pritchard (2014). Varieties of externalism. *Philosophical Issues* 24 (1):63-109.
- Chalmers, David (2009). Foreword. In Clark, *Supersizing the Mind*, ix-xvi. Oxford University Press.

¹⁴ Kim Sterelny (in discussion) suggests that Clark and Chalmers are appealing to the kind of generality that will specify a natural kind, whereas Williamson is interested in generality considered as the grain of specificity of the explanation. I’m not convinced by this: while Clark’s ‘vehicle externalism’ might be looking for natural kinds, Chalmers (2009) makes clear that the extended mind hypothesis is compatible with even non-naturalist positions. Alternatively, Williamson can be understood as providing natural-kind style explanations. In his (2009) response to Jackson, he claims that his account of externalist causal-psychological explanations fit “good explanatory practice in the natural sciences by identifying significant generalizations” (Williamson 2009, 331).

¹⁵ See Clark and Chalmers (1998, 9) and Chalmers (2009, xvi).

¹⁶ Thanks to Daniel Stoljar for discussion of this point.

- Clark, Andy (2005). Intrinsic content, active memory, and the extended mind. *Analysis* 65 (285):1-11.
- Clark, Andy (2015). What 'extended me' knows. *Synthese*. DOI: 10.1007/s11229-015-0719-z
- Clark, Andy and Chalmers, David J. (1998). The extended mind. *Analysis* 58 (1):7-19.
- Drayson, Zoe (2010). Extended cognition and the metaphysics of mind. *Cognitive Systems Research* 11 (4):367-377.
- Kim, Jaegwon (1982). Psychophysical supervenience. *Philosophical Studies* 41 (January):51-70.
- Molyneux, Bernard (2007). Primeness, internalism, and explanatory generality. *Philosophical Studies* 135(2):255-277.
- Pritchard, Duncan (2010). Cognitive ability and the extended cognition thesis. *Synthese* 175(S1):133-151.
- Putnam, Hilary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.
- Stich, Stephen P. (1978). Autonomous psychology and the belief/desire thesis. *The Monist* 61 (October):573-91.
- Wikforss, Åsa (2014). Extended belief and extended knowledge. *Philosophical Issues* 24 (1):460-481.
- Williamson, Timothy (1995). Is knowing a state of mind? *Mind* 104 (415):533--65.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.
- Williamson, Timothy (2009). Reply to Jackson. In Greenough and Pritchard (eds.) *Williamson on Knowledge*, pp. 330-331. Oxford University Press.
- Yablo, Stephen (2003). Causal relevance. *Philosophical Issues* 13 (1):316-28.