THE COST OF TREATING KNOWLEDGE AS A MENTAL STATE[1]

Martin Smith

My concern in this paper is with the claim that knowledge is a mental state – a claim that Williamson places front and centre in *Knowledge and Its Limits*. While I am not by any means convinced that the claim is false, I do think it carries certain costs that have not been widely appreciated. One source of resistance to this claim derives from internalism about the mental – the view, roughly speaking, that one's mental states are determined by one's internal physical state. In order to know that something is the case it is not, in general, enough for one's internal physical state to be a certain way – the wider world must also be a certain way. If we accept that knowledge is a mental state, we must give up internalism. One might think that this is no cost, since much recent work in the philosophy of mind has, in any case, converged on the view that internalism is false. This thought, though, is too quick. As I will argue here, the claim that knowledge is a mental state would take us to a view much *further* from internalism than anything philosophers of mind have converged upon.

Keywords: Knowledge, mental state, internalism, externalism, switching

I. THE OBJECTION FROM INTERNALISM

The claim that knowledge is a mental state is a centrepiece of the knowledge first movement in epistemology – the first claim that one encounters in Williamson's *Knowledge and Its Limits*. While there might be various watered-down ways to interpret this claim, Williamson is clear that it should be taken in a literal, unflinching way. In his view, knowledge is as much a mental state as belief, desire, fear, hope, intention etc. For me to know something – that it's raining outside, that per capita greenhouse gas emissions are higher in Qatar than in any other nation, that the velocipede was invented in 1817 – is just for my mind to be in a particular state. An objection immediately suggests itself: Knowledge, unlike belief, desire, fear, hope, intention etc., is *factive* – if I know that per capita greenhouse gas emissions are higher in Qatar than in any other nation, it has to be *the case* that per capita greenhouse gas emissions are higher in Qatar than in any other nation. But that is not a fact about my state of mind – it involves the wider world.

One way to make this kind of objection more precise is by appealing to *internalism* about the mental. According to internalists, one's mental states are determined by the internal physical state of one's body and brain – any two individuals with the same internal

[1] This paper was presented at the University of Edinburgh in November 2014 and greatly benefitted from the discussion on this occasion. Particular thanks go to Zoe Drayson, Jesper Kallestrup, Clayton Littlejohn, Heather Logue and Aidan McGlynn. I am also grateful to Adam Carter, Emma Gordon, Ben Jarvis and Keith Wilson for very helpful comments on earlier drafts of this paper.

physical state will have the same mental states. I might currently know that per capita greenhouse gas emissions are higher in Qatar than in any other nation, but I could be in the same internal physical state even if that were false, in which case I wouldn't know it. Whether I know something is not, in general, determined by my internal physical state. If my mental states are determined by my internal physical state, as internalists would have it, then knowledge is not a mental state. We might call this the *objection from internalism*[2].

One might think that this objection can be swiftly dismissed – after all, internalism is a view that has largely fallen from favour in the philosophy of mind, succumbing to a number of well known arguments and thought experiments. In 'Is knowing a state of mind?' (1995) and in chapter 2 of *Knowledge and Its Limits*, Williamson argues that, once we have given up the restrictive internalist picture of the mind, and embraced *externalism*, no meaningful impediment remains to treating knowledge as a mental state. As he puts it 'An externalist conception frees one to affirm that knowledge is a mental state' (Williamson, 1995, pp533). In this paper I will argue that treating knowledge as a mental state would, in fact, have us recoil *far further* from internalism than any of the well known arguments and thought experiments take us. Much of this paper will consist of a survey of these arguments and thought experiments, along with a slightly new way of mapping out their purported externalist consequences. The survey will not be especially detailed or deep – just detailed enough to extract the consequences that concern me. I will conclude by returning to the view that knowledge is a mental state, and showing that it takes us, in effect, into uncharted externalist territory[3].

In the end, nothing that I say here will constitute a decisive refutation of the view that knowledge is a mental state. Whether we do decide to treat it as one will ultimately depend on how we weigh the costs and benefits of doing so – and I won't attempt to come to any overall judgment about this. What I do hope to show here is that the costs of treating knowledge as a mental state have, in at least one respect, been seriously underestimated.

Before proceeding it may be worth remarking briefly on what it means to *deny* that knowledge is a mental state. At a minimum, to deny this is to accept that a person's mental states do not, on their own, settle what that person knows – that two individuals with the same mental states might have different knowledge, owing to differences in the worlds they

[2] It is sometimes suggested that internalism should be reformulated in such a way as to include *dualist* theories on which one's mental states are constitutively independent of the external world, but not determined by one's internal physical state. I won't attempt such a reformulation here – but it's worth noting that a dualist of this kind would, I imagine, have just as much reason to resist classifying knowledge as a mental state as one who signs up to the materialist internalism under consideration here.

[3] In 'Is knowing a state of mind? The case against' (2009) Elizabeth Fricker also takes issue with Williamson's use of standard externalist arguments to soften one's resistance to the idea that knowledge is a mental state, arguing that the acceptance of the former and rejection of the latter makes for a perfectly stable, defensible overall view. I agree with this – but aim, in a way, to go further. I will argue that there are in fact very strong reasons for otherwise committed externalists to stop short of granting that knowledge is a mental state.

inhabit. It is sometimes suggested that the alternative to treating knowledge as a mental state is to treat it instead as a metaphysical hybrid or composite of mental and non-mental factors (Williamson, 2000, pp5, section 1.3, Nagel, 2013, pp275, section 2). This, though, seems a very tendentious way of putting things – almost as though anyone who denies that knowledge is a mental state is committed to finding some *analysis* or *factorisation* of knowledge into mental and non-mental components. But the denial that knowledge is a mental state can be perfectly well combined with doubts about whether it is analysable and, indeed, with general doubts about the project of philosophical analysis (Fricker, 2009, section 3, McGlynn, 2014, pp168). Many of Williamson's arguments in the opening chapters of *Knowledge and Its Limits* are directed, first and foremost, against the possibility of analysing or factorising knowledge. In spite of the way that Williamson sometimes presents things, these arguments cannot simply be co-opted to support the view that knowledge is a mental state.

II. NATURAL KIND EXTERNALISM

According to the *causal theory of reference*, the referents of certain proper names and common nouns are determined by the particulars and natural kinds that are causally related to their usage, and not by any ideas or descriptions that users associate with them. The causal theory attracted a number of supporters in the 1970s and 80s due largely to the pioneering work of Kripke (1972) and Putnam (1975), and a corresponding view about the contents of beliefs followed in its wake – in order for one to hold beliefs about certain particulars or natural kinds, it's necessary that one have an appropriate causal relation to them.

In Putnam's classic 'Twin Earth' thought experiment we are introduced to a remote planet, exactly like the Earth, but completely devoid of water – that is, H_2O. Instead of water, a different chemical compound – XYZ – flows in the rivers, fills the lakes and oceans, falls as rain and quenches the thirst of the Twin Earthers. XYZ has all of the same macroscopic properties as water – it's clear, odourless, boils and freezes at 100 and 0^o C at sea level etc. Furthermore, the Twin Earthers, who speak like us, even call it 'water' (and 'eau' and 'mizu' etc.). Finally, we might imagine that it's prior to 1750 and, without access to microscopes and the like, no one on Earth or Twin Earth would be capable of distinguishing these two substances. According to Putnam, while those on Earth and those on Twin Earth might associate exactly the same ideas and descriptions with the term 'water', when the former use the term it refers to H_2O, and when the latter use the term it refers to XYZ.

The lesson here is not just a semantic one, however – there are also implications for the contents of beliefs and other mental states (McGinn, 1977, Burge, 1979, n2, Putnam, 1982, chap 1). When a person on Earth sincerely utters the sentence 'there's water in my flask' he expresses a belief that is true iff there is H_2O in his flask. When his duplicate on Twin Earth utters the same words, he doesn't express a belief about H_2O – neither he, nor anyone on his planet, has ever come into contact with this substance. Rather, his belief is true iff there is XYZ in his flask. These two individuals may be in exactly the same internal physical state (at least if we ignore the fact that the human body contains water). Nevertheless, they hold beliefs with different contents – the former holds beliefs about H_2O and no beliefs about XYZ while the latter holds beliefs about XYZ and no beliefs about H_2O. If we accept this verdict then we are lead to a view we might call *natural kind externalism* – the content of certain beliefs involving natural kind concepts depends on what natural kinds are present in the believer's environment. Similar thought experiments could doubtless be used to motivate similar conclusions about other kinds of mental states – desires, fears, hopes, intentions etc – but I will focus here on beliefs.

There are, of course, a number of objections to natural kind externalism and to the standard verdicts about Twin Earth thought experiments – but I won't explore them. While I am inclined to think that this view is broadly correct, I have no stake in defending it here – or in defending any of the various kinds of externalism I will canvass. My aim, rather, is to show that even if all of these kinds of externalism be simultaneously accepted, the claim that knowledge is a mental state still represents a considerable further step in an externalist direction.

If we accept natural kind externalism, then we will have to abandon the letter of internalism – one's mental states will not be determined exclusively by one's internal physical state. But whether natural kind externalism represents a *major* departure from internalism very much depends on one's motivation for being an internalist in the first place. One might be moved to accept internalism as a result of philosophical theories that one holds – internalism may be entailed, for instance, by a number of once fashionable, reductionist views in the philosophy of mind, such as the identity theory and classical functionalism (see, for instance, Smart, 1959, Armstrong, 1968). If this was one's reason for accepting internalism, then perhaps it would have to be accepted to the letter – one could tolerate no divergence from it. But there are far more pedestrian thoughts that seem to point us in an internalist direction. Here is one: It shouldn't be possible to influence or interfere with a person's state of mind without causally interacting in some way with that person's sense organs or body. If this was the reason one was inclined towards internalism, then one might be quite open to qualifying or compromising the view in various ways.

Suppose a person on Earth was transported, without his knowledge, to Twin Earth. Natural kind externalism does not commit us to the view that this person's beliefs would immediately change. On the contrary, those who have considered such a scenario tend to

agree that the beliefs the exile expresses using the term 'water' would continue to be about H_2O – at least for a while (see, for instance, Burge, 1988, Boghossian, 1989). If he arrives with his flask and says to a thirsty Twin Earther 'there's water in my flask' he would be expressing a true belief about H_2O whereas, if the Twin Earther were to say 'there's water in your flask', he would be expressing a false belief about XYZ. The two would, at this point, be speaking slightly different languages, though neither would be in a position to notice the difference. However, after a substantial period of time has elapsed and the exile has had a good deal of causal interaction with XYZ – drinking it, bathing in it etc. – while his interactions with H_2O recede into the distant past, it seems that he will have become, in all relevant respects, like a Twin Earther himself. Natural kind externalism predicts that, at some point, the contents of his beliefs will shift[4]. This could happen even if the individual remains forever oblivious to his teleportation to Twin Earth.

Let's introduce the term 'switching' to refer to a change in a person's mental states that is not mediated by any change in the person's internal physical state. Typically, when a person's beliefs, desires, fears, hopes, intentions etc. change, this will be the result of some new experience or of the acquisition of new information or of some conscious reasoning or unconscious processing, or of remembering something or forgetting something etc. All of these processes involve very definite changes in a person's sensory systems and brain. Switching is something altogether different – a change in a person's mental states that does not involve any of these processes and does not, in effect, involve the person's *body* at all. If a person's mental states change while his internal physical state remains completely static then this will, of course, count as a case of switching – but so too will cases in which the changes in a person's internal state are incidental or irrelevant to a given mental state change.

Switching strikes us as a peculiar idea – it's strange, perhaps even disconcerting, to think that this is something that could actually happen to us. In the case just described, however, the mental states of the exile do switch – he loses his beliefs about H_2O and acquires beliefs about XYZ. While he may undergo all kinds of changes in his internal physical state after his teleportation, the mental state shift is not due to any of these – they could have unfolded in exactly the same way if he had remained on Earth. Natural kind externalism does, then, allow for the possibility of switching – but the only way this can come about is via the substitution of one environment for another that contains a superficially indistinguishable natural kind, in such a way that an unwitting believer remains oblivious to the change. As we've seen, such switching is *slow* – it takes a substantial period of time, after the substitution, for the switching will take effect. Furthermore, such switching is *proximal* – it involves changes to the natural kinds with which the believer is

[4] Natural kind externalists may disagree about just how soon this shift can occur. Even if it takes some time for the exile's 'water' beliefs to latch on to XYZ, one might hold that they disconnect from H_2O – becoming *indeterminate* in content – shortly after arrival. Whatever view one adopts, the change will not be immediate – requiring, at the very least, that the exile's 'water' beliefs come partly under the causal control of XYZ.

causally interacting during the period of the switch. The final thing to observe about this kind of switching is that it seems exceedingly contrived and farfetched (Warfield, 1992). Given the way the world actually is, the conditions required for such switching are *rarely*, if ever, realised. At worst, then, natural kind externalism, as described here, lands us with the possibility of slow, proximal, rare switching. For an internalist motivated by an instinctive aversion to the idea of switching, the move to natural kind externalism need not seem like a major concession.

II. SOCIAL EXTERNALISM

The Twin Earth thought experiment purports to show that the contents of one's natural kind beliefs can depend upon the nature of the world outside of one's body. A few years after Putnam first described Twin Earth, another famous thought experiment, devised by Tyler Burge, purported to extend this conclusion to virtually *all* beliefs (Burge, 1979, 1986). Burge has us imagine an individual who is unaware that arthritis is a condition of the joints and complains to her doctor of arthritis in her thigh. When she says 'I have arthritis in my thigh' she expresses a (necessarily) false belief. Now consider another individual, with exactly the same history and internal states, who happens to be part of a linguistic community in which the term 'arthritis' is used to refer to a broader class of rheumatoid ailments which happens to include the very complaint she has in her thigh.

According to Burge, when this second individual says to her doctor 'I have arthritis in my thigh' she is not expressing a false belief about *arthritis*, a condition exclusively of the joints – neither she nor anyone in her community has beliefs about arthritis. Rather, she is expressing a true belief about the broader category that her linguistic community recognises. If we accept this verdict, we are lead to a view that we might call *social externalism* – the content of certain beliefs can depend upon facts about how terms are used in the believer's linguistic community.

Unlike the Twin Earth thought experiment, this thought experiment seems to provide a general recipe by which the content of almost any belief could be shown in principle to depend on facts about the wider world. There is nothing unusual about the term 'arthritis' – if the meaning of this term, on the lips of one individual, can be shown to depend on how it is used in her linguistic community then, presumably, the same could be shown about a great many of the terms that we use. When it comes to the meanings of our words, we are not the final arbiters – we are disposed to defer to the common usage of the linguistic community were are a part of. Burge himself takes his externalist conclusion to have a very broad scope and to extend to any belief involving theoretical, observational, natural or artefactual kind concepts (Burge, 1979, section IIb).

Social externalism may allow for a kind of mental state switching as defined in the last section – one could simply imagine a person travelling between the linguistic communities envisaged in Burge's thought experiment. Once again, though, those who have considered such scenarios agree that this person would not immediately lose her beliefs about arthritis upon arriving in the new community. It is only after she has become a fully fledged member of the new community and has clocked up a good deal of causal and linguistic interaction with its members that the contents of these beliefs might eventually shift. And this shift could of course happen without it being due to any change in the individual's internal state and without the individual ever realising that the two communities use the word 'arthritis' in different ways. Once again, the only sort of switching that is in prospect here is switching that is slow, requiring a substantial period of time to elapse, and proximal, requiring changes to the linguistic community with which the believer interacts during this period.

Interestingly, though, switching of this kind would not seem to require farfetched or contrived scenarios to bring about. On the contrary, there need be nothing out of the ordinary about an individual travelling between subtly different linguistic communities, while remaining oblivious to the differences between them. Indeed, as Ludlow (1995) has argued, this is something that may *commonly happen* – as when people move between Britain and the United States without realising that words such as 'chicory', 'jelly', 'pants' etc. mean different things in British and American English. Consider an American who has just arrived in Britain and sincerely says 'Chicory is rich in vitamin A' thereby expressing a belief about *cichorium endivia*. Social externalism may suggest that, once she has lived in Britain for enough time, and become a part of the new linguistic community, this belief may be lost. Her disposition to assert 'Chicory is rich in vitamin A' could still be present, but these words will now express a new belief about *cichorium intybus*.

Ludlow suggests that this kind of switching may even occur when people move across social groups that have their own internal conventions for the use of certain terms (Ludlow, 1995, pp46-49). Philosophers, for instance, use terms like 'realist' and 'pragmatist' in a specialised way that differs from general public usage. Perhaps someone who started socialising more and more with philosophers, while remaining unaware of this difference, would eventually come to express different beliefs with these words. There is, of course, room for reasonable disagreement amongst social externalists as to whether switching can happen quite so easily as this. There may, for instance, be versions of social externalism that place more emphasis on the original linguistic community in which a person acquired the use of a term and which, as a result, make the contents of that person's beliefs far more resilient to these sorts of changes. However we fill in the details, though, social externalism will make switching into a more realistic prospect than natural kind externalism and thus, along one dimension at least, it represents a further step away from internalism.

III DEMONSTRATIVE EXTERNALISM

Natural kind externalism, as noted above, is closely associated with a view about how certain proper names and common nouns refer. A related view about *demonstrative expressions* – expressions such as 'this' and 'that' used to pick out objects being directly perceived – is connected with externalism of a third type. According to Kaplan (1979, 1989), Evans (1982, chap. 6) and others, when one uses a demonstrative expression, one's assertion does not merely have descriptive content – rather, its content may literally *involve* the perceived object being demonstrated. Suppose I see an apple on the table and, pointing towards it, utter the words 'That apple is overripe'. On the present view, if it had been a different apple sitting on the table, then the content of my assertion would have involved a different object and, thus, would have been a different content, even if the new apple looked exactly the same to me and I was in exactly the same internal state.

This may be a view about language but, once again, a view about the contents of beliefs is waiting in the wings (Evans, 1982, chap. 6). As a general methodological rule, we should attribute to people beliefs with the contents of their sincere assertions and hold back from attributing to people beliefs with contents that they would not, or could not, sincerely assert. This forces us to conclude that, in the two scenarios just described, I hold different beliefs, despite being in the same internal state. In the first scenario, I hold a belief about a particular apple. In the second scenario, I hold a belief about a second, different apple and no belief at all about the first apple which I'm in no position to refer to with a demonstrative. We might term this *demonstrative externalism* – the view that the contents of certain beliefs involving demonstrative concepts can depend upon the observed objects being demonstrated.

My internal state could, of course, be the same even if there were *no apple at all* – if I was hallucinating, say. In this case, according to one version of demonstrative externalism, the utterance 'That apple is overripe' would have no content and would fail to express any belief. Although I'd be in the same internal state as someone who holds a genuine belief, I would be suffering from an *illusion of belief* (Evans, 1982, section 2.2, appendix to chap. 6, McDowell, 1986).

Like social externalism, demonstrative externalism may allow for a new kind of mental state switching. For the demonstrative externalist, switching could in principle be accomplished by surreptitiously substituting objects within a person's immediate vicinity. Suppose I am examining an apple on the table before me and am willing to assert things like 'That apple is overripe', 'That apple is a Golden Delicious' etc. Suppose that, while I'm momentarily distracted, someone replaces the apple with another of the exact same size, shape and colour. Even if I don't notice the subterfuge, according to the demonstrative externalist, my beliefs will change – while I remain willing to assert the very same sentences,

they will now express beliefs with new contents. This change will not of course be due to any corresponding change in my internal physical state.

One thing that is remarkable about this kind of switching is that it would appear to take place very *fast*. Demonstrative externalism seems to predict that the content of a demonstrative belief should change the very moment a perceptual, demonstrative connection to a new, substituted object is established. While this kind of switching may not require any farfetched mechanisms per se, it does require that circumstances conspire against a perceiver in a very particular way. For the switching to occur, it must be that a perceived, demonstrated object is substituted for another, indistinguishable object without this impinging in any way on the internal state of the perceiver. While such switching may occur in practice, we would expect it to be a rare event. Finally, the switching envisaged here would still appear to be proximal – after all, the substitution must occur within the range of the believer's perceptual faculties, in order for the objects to be candidates for demonstration.

One clarification is important however – in describing this sort of switching as proximal, I don't mean to suggest that the substitution must necessarily take place close to the believer's body. Suppose I am shown an apple on the other side of the world over Skype and utter those same words – 'That apple is overripe'. One might take the view that this apple is literally being perceived and that it enters into the content of my utterance and my belief. If so, then some appropriate sleight of hand on the other side of the world could elicit a fast mental state switch. This switching would still, however, be categorised as proximal – it is triggered by an event in a part of the world with which, however physically far away it might be, I am in intimate causal contact. The boundaries of a believer's environment, in the relevant sense, aren't drawn at some fixed distance from his body. The relevant notion of an environment is a causal one – it includes those parts of the world with which one is causally interacting in a relatively direct way. I have no interest in trying to specify precisely what a person's environment includes – the notion is not meant to be a precise one. It's enough that there be clear cases of events that occur within a person's environment and clear cases of events that occur outside of it, even if there are a great many borderline cases.

It is interesting to note that, according to Evans, an important condition for an object to feature in the content of a demonstrative utterance is that one be capable of locating that object within egocentric space – the space of possibilities of bodily action (Evans, 1982, sections 6.2-6.4). This condition is not fulfilled in cases of the type just described, leading Evans to speculate that these kinds of cases should, perhaps, be assimilated to cases of 'deferred ostention' in which 'this' and 'that' arguably *are* used as shorthand for a description (Evans, 1982, section 6.2). If this is right then, while I'm Skyping, the apple on the other side of the world is not a part of the content of any of my beliefs. Rather, the belief I express will have a descriptive content and substituting apples won't affect a switch

in my mental states. It may indeed be a general consequence of Evans's view that switching of the present kind can only be brought about by events that take place close to a believer's body – but I won't investigate this further here.

According to demonstrative externalism, as I've defined it, certain beliefs expressed using 'this' and 'that' depend for their content on one's relations to the immediate environment. In addition, demonstrative externalists often endorse a closely related view about certain beliefs expressed using *indexical* expressions such as 'I', 'he', 'she' 'here' etc. Evans (1982, chap. 6, appendix 6) describes a case in which a person and his duplicate on Twin Earth both think about their current location as being 'here'. If both were to sincerely assert 'It's sunny here' then, on Evans's view, they would express beliefs with different contents – contents that involve different locations – even though these locations appear exactly the same and the two individuals are in the same internal state.

Evans also describes a kind of switching case in which a person confined to a hospital bed keeps muttering to himself 'It was so hot here yesterday'. While the person takes himself to be lying still in the dark, his bed is actually in constant motion on silent, well-oiled wheels. When he repeatedly utters this sentence he thinks he is reiterating the same belief over and over – but, according to Evans, he is actually expressing a *new* belief each time, as his location will have changed (Evans, 1982, chap. 6, appendix 6). This sort of switching can also, perhaps, be categorised as fast, proximal and rare. The switching is taken to occur from moment to moment, in response to changes in the person's immediate surroundings and, while this kind of set-up may be practically possible, it would also be exceedingly unusual.

IV DISJUNCTIVISM

Demonstrative externalism predicts that there can be a mental difference between a person who is veridically perceiving the world and one who is hallucinating or taken in by an illusion, even though the two may be in exactly the same internal state. This claim is characteristic of another influential kind of mental state externalism – namely, *disjunctivism* about perception (see, for instance, Hinton, 1967, Snowdon, 1980, McDowell, 1982). According to disjunctivists, seeing that something is the case, feeling (in the sense of touch) that something is the case, hearing (directly, and not in the sense of hearing testimony) that something is the case etc. are all mental states in their own right – different to the mental states of *merely seeming* to see or hear or feel etc. that something is the case, as when one suffers a hallucination or is fooled by an illusion. Further, disjunctivists typically insist that veridical perceptual states have no mental element in common with hallucinations and illusions – no 'highest common factor', as it's sometimes put. The category of 'perceptual

experience', which is supposed to subsume all three types of state, is a heterogeneous or *disjunctive* one.

Relatedly, disjunctivists often hold that, when one sees or feels or hears etc. that something is the case, one's mind reaches right out into the world and literally *involves* external objects and facts – and that this is a fundamental, defining feature of these mental states that is not shared by hallucinations and illusions. When it comes to the externalist character of disjunctivism, however, many of these further commitments are not needed. For externalism to hold, it is enough that there be *some* mental difference between veridical perception, hallucination and illusion, each of which can be associated with the same internal state – it doesn't matter exactly how the difference is explained or how profound it is taken to be.

It is possible, in principle, for a veridical perception to become an indistinguishable hallucination or illusion, or vice versa, without there being any accompanying change in the perceiver's internal state. In such a case, disjunctivism predicts that the perceiver will undergo mental state switching. Suppose I wander into a room and see that there is a blue cube sitting atop a white podium. Suppose that, in an instant, the cube is vaporised and a perfect hologram of the cube projected in its place. Suppose the change does not impact on my sense organs or body in any way. According to the disjunctivist, this event will trigger a mental state switch – I will have gone from seeing that there is a blue cube on a white podium to merely seeming to see that there is a blue cube on a white podium.

The switching would appear to be fast – indeed it would appear to be *instantaneous*. The moment the cube is destroyed, it is no longer the case that there is a blue cube on the podium and I can no longer be said to see that there is a blue cube on the podium. If this kind of example is characteristic of these switching cases, then they will clearly be very rare. In fact, I think it is characteristic – for a veridical perception to seamlessly change into a hallucination or illusion or vice versa is an extraordinary occurrence. Finally, this switching must be proximal, for much the same reason as the switching permitted by demonstrative externalism must be proximal – it requires changes in the objects and states of affairs that one is perceiving.

Natural kind, social and demonstrative externalists all emphasise ways in which the content of a given type of mental state may depend upon external factors. Disjunctivists, however, find a role for external factors in determining the *type* of mental state that bears a given content – the type of attitude one has towards that content. For this reason it is sometimes described as an *attitude externalism* (Williamson, 2000, chap. 2). The view that knowledge is a mental state should also, of course, be characterised as a kind of attitude externalism in this sense. In fact, disjunctivism can provide a certain model for how to think about this view.

The disjunctivist draws a mental contrast between successful states on the one hand and failed or botched states on the other – even though the two kinds of state may be indistinguishable for the one who instantiates them. The successful states include seeing, hearing, feeling etc. that something is the case and the botched states include merely seeming to see or hear or feel etc. that something is the case. One who treats knowledge as a mental state also draws a mental contrast between a successful state and a botched state that may be indistinguishable from the inside. On this view, the successful state is knowing that something is the case and the botched state is merely believing that something is the case. On both views, the difference between the successful states and the botched states does not lie with external, worldly factors – but with mental factors that one is not in a position to detect.

For the disjunctivist, when one sees or hears or feels etc. that something is the case, one's mind reaches right out into the world and involves external objects and facts. If knowledge is a mental state then, presumably, it is also a state in which one's mind reaches right out into the world (indeed much *further* out into the world – though more on this in the next section). One who accepts that knowledge is a mental state might follow the disjunctivist example even further and insist that knowledge and mere belief have no mental element in common. Williamson, though, stops short of this, remaining open to the standard idea that knowledge involves belief and that belief is a mental state (Williamson, 2000, section 1.5). For Williamson, belief can serve, in one sense, as the 'highest common factor' of knowledge and mere belief (though he denies of course, that knowledge can be 'factorised' into belief and further components).

If we do accept the existence of factive mental states such as seeing, hearing and feeling that something is the case, it might seem like a small step to admit one *more* factive mental state – namely, knowing that something is the case. For Williamson, once we embrace factive mental states, knowledge naturally takes its place as the most *general* such state – the factive mental state that one is in when one is in any factive mental state (Williamson, 2000, section 1.4). As I will explore in the next section, focussing on mental state switching offers a rather different perspective on this.

V COUNTING THE COST

I have examined a number of externalist views through the lens of mental state switching. Each of these theories allows for different kinds of switching possibilities, which I have mapped out relative to three dimensions – a speed dimension, a range dimension and a frequency dimension. The results thus far might be presented as follows:

	Speed	Range	Frequency
Natural Kind Externalism	Slow	Proximal	Rare
Social Externalism	Slow	Proximal	Common
Demonstrative Externalism	Fast	Proximal	Rare
Disjunctivism	Fast	Proximal	Rare

How, then, should we fill in the row for the view that knowledge is a mental state? If the arguments of this final section are on the right track, then this view allows for a kind of switching that is unlike anything countenanced so far.

I currently believe, and know, that my local fishmonger has a blue shopfront. Suppose that, while I'm on a trip overseas, the shopfront is restyled and painted red. As soon as I return home and set eyes upon it, I will lose my belief that my local fishmonger has a blue shopfront. This, of course, will be a perfectly standard case of mental state change, in which my beliefs are revised in light of new information. But if knowledge is a mental state, then this is not the first time that the new paint job will have had an effect on my state of mind. Rather, this will have happened days or weeks before, while my body was thousands of miles away and in no causal contact with the fishmonger – for this is the point at which I will have ceased to *know* that my local fishmonger has a blue shopfront. And this change, whenever it happens, won't be due to any change in my internal state, which will continue exactly as it would have if the shop front had not been touched.

The switching that occurs here is fast – the moment the red paint is applied, and it ceases to be true that my local fishmonger has a blue shopfront, I can no longer know it. Furthermore, switching of this kind would presumably be common – the events described are in no way unusual. Already we find ourselves confronting something new – the possibility of fast, common switching. But what is perhaps most striking about the switching described here is that it is *not proximal*. When the switching takes place, it is triggered by events completely *outside* of my environment at that time – events in a part of the world that is both physically remote from my body and in no causal contact with it.

Consider, in addition, the following example discussed by Jennifer Nagel (2013, pp289): If knowledge is a mental state, then on the evening of the 14th of April 1865 a kind of mass mental state switching would have swept the globe. Horrified theatregoers who witnessed Lincoln's assassination might have ceased to believe that Lincoln is president of

the United States – but countless millions more will have ceased to *know* it, including people on the other side of the world to Washington[5].

If knowledge is a mental state, then the range within which switching may be triggered will be effectively limitless. Consider a person stranded in a disabled space probe drifting in deep space billions of miles from the surface of the Earth and with no means of communication. If knowledge is a mental state, then this person can still have influence on the minds of friends, relatives and acquaintances on Earth, and they can still have influence on his. If he shaves his head, for instance, then people on Earth may go from knowing that he has a full head of hair to merely believing it. And this change in their state of mind will presumably be simultaneous with the head shaving, even though it may take *light* several hours to travel the distance between Earth and the probe. The mental state switch will take place long before the *information* that is has taken place could, in principle, reach the Earth.

It may even appear as though treating knowledge as a mental state leads to the possibility of instantaneous action at a distance and thus serves to settle what would usually be regarded as a very controversial empirical issue. But this will only be so if we think of a person's mental state switch as an event spatially located within that person's body or brain – and this is presumably an unwelcome remnant of internalist thinking that should be given up. If we treat knowledge as a mental state, then a person's mental state switch may be better thought of as an event that can take place at any distance from the person's body and brain, or an event that can be spread out over a vast distance, or even an event that has no determinate spatial location whatsoever.

At the very least then we can fill in the final row of the table as follows:

	Speed	**Range**	**Frequency**
Natural Kind Externalism	Slow	Proximal	Rare
Social Externalism	Slow	Proximal	Common
Demonstrative Externalism	Fast	Proximal	Rare
Disjunctivism	Fast	Proximal	Rare
Knowledge is a mental state	Fast	Distal	Common

[5] Since Lincoln did not die until the morning of the 15th of April, one might argue that it was only at *this* point that he strictly ceased to be the president of the United States. I put this potential complication aside for present purposes. There are, in any case, countless further historical examples that we could consider instead of this one.

Uniquely amongst the kinds of externalism I have considered, the view that knowledge is a mental state allows for fast, distal, common mental state switching. But this is not yet the end of the story.

Social externalism, I claimed, would make switching a common occurrence. While I stand by this claim, it is important to keep it in perspective. The kind of switching that social externalism permits may be common relative to the kinds of switching permitted by natural kind and demonstrative externalism, but it is hardly something that we would expect to undergo *on a daily basis*. Indeed, depending on the precise version of social externalism being entertained, switching is something that may happen on only a handful of occasions during a lifetime. But how often does knowledge turn into ignorance?

As I sit writing, it is easy enough to recall examples of this from my own recent past. This morning I believed, and knew, that my bin was on the kerb, having put it there myself. My belief persisted until, some hours later, I saw it by the gate. My knowledge, however, will have been lost some time prior to this when my neighbour, seeing that the rubbish had been collected, wheeled it in. This loss of knowledge, whenever it happened exactly, won't have been accompanied by any change in my internal state. A week ago I believed, and knew, that my one year old daughter had four teeth. That belief persisted until yesterday when I noticed that a fifth tooth had come through. But my knowledge will have been lost some time in the intervening week, without any corresponding change in my internal state. If knowledge is a mental state, then these will both have been bona fide mental state switches.

Furthermore, mass switching events, like the one triggered by Lincoln's assassination, would *themselves* be relatively common occurrences if knowledge were a mental state – an unexpected side-effect of a globalised culture obsessed with celebrity. The table above suggests that treating knowledge as a mental state would lead to common switching – but this, in truth, is to play down the consequences of the view. To do justice to it, what is needed is a further point on the frequency spectrum – *rampant* switching – where none of the other kinds of externalism take us.

The table may also mislead in another respect. To suggest that treating knowledge as a mental state allows for *fast* switching is, once again, to underestimate the consequences of the view. This claim might seem surprising – after all, I've used the term 'fast' to describe even cases in which switching is instantaneous, such as the switching permitted by disjunctivism. Presumably, then, 'fast' must cover the very extreme end of the speed spectrum – how could switching be any speedier than instantaneous?

The idea that knowledge requires a kind of insulation or protection against the possibility of error is one that is widespread in contemporary epistemology. Suppose one believes P via method M at time t. Say that one's belief is *safe* iff it could not easily have been the case that one falsely believed P via M at t. This is sometimes spelled out in terms

of possible worlds: Say that one's belief is safe iff there is no close possible world in which one believes P via M at t and P is false. A number of philosophers, Williamson included, have been attracted to the idea that knowledge requires safety – in order for one to know P at time t, one must safely believe P at time t, in something like the sense just defined (Williamson, 1992, 2000, chap. 5, Sosa, 1999a, 1999b, Pritchard, 2005, chap. 6, Smith, 2009)

With this in mind, return to the case of Lincoln's assassination. Booth, allegedly, decided to wait until a particular line in the play had been spoken before opening fire. Arguably, he could have easily decided to shoot Lincoln a few minutes earlier than he did or, put differently, there are close possible worlds in which he does shoot Lincoln a few minutes earlier than at the actual world. If this is right then, on the evening of April 14th 1865, the belief that Lincoln is president of the US would have ceased to be *safe* at least several minutes before it ceased to be *true*. Several minutes before the assassination, one's belief that Lincoln is president, while still true at the actual world, would be false at some close possible worlds at which it is held via its actual method. If knowledge requires safety, and is a mental state, then *this* is the point at which the mental states of people the world over would have switched – *prior* to the event that would appear responsible for the switch. What we seem to confront here is a case of *backwards* switching, where an event triggers a mental state switch *in the past*.

One might, of course, resist this description of the case. Perhaps what these reflections about safety show is that the assassination itself is *not* what truly triggers the mental state switch – rather other events are responsible, such as Booth approaching Ford's theatre, his crouching outside the presidential box, pistol in hand etc. And these events, we might insist, really do take place before the mental state switch does. The trouble with this redescription, though, is that if Booth had not gone on to assassinate Lincoln, it's not clear that there would have been *any* mental state switching at all, even if these other events had still taken place.

We might consider another case in which there are no obvious precursors to an event that renders a known proposition false. Suppose I decide, on a complete whim, to shave my head, purely to deprive my friends, relatives and acquaintances of the knowledge that I have a full head of hair. If knowledge is a mental state, then my actions will bring about a small spate of mental state switches. But when do these switches take place? Suppose the impulse to shave my head suddenly strikes me at 10am and I act on it immediately. This impulse could, however, have easily struck me at, say, 9:50 or 9:40 instead. As such, there are close possible worlds at which I shave my head at these earlier times. The switch must, then, have already occurred by 9:40 – before my action is performed, and before I even have any inkling of performing it.

Once again, one could insist that there is some causal precursor of my action already simmering away in my brain by 9:40, and that this is what is truly responsible for the mental state switching. But, even if such a precursor really could be identified, it is not something

that I have any awareness of or any control over – and yet it seems that *I* am the one responsible for switching the mental states of my friends and relatives. It seems that the switch is a direct result of my action and my decision, and not of some subconscious activity in my brain before the decision was even made. Although the switch might occur at 9:40, it occurs because of what I do at 10.

One might attempt to avoid these results by joining Neta and Rohrbaugh (2004), Comesaña (2005), Baumann (2014) and some others in denying that knowledge truly requires safety. In order for the present point to stand, though, it is enough that a lack of safety preclude knowledge in the cases just described, whether or not it does so in all possible cases that we might imagine. Indeed, at a bare minimum, all that is really needed to construct backwards switching cases of the present sort is that a lack of safety *sometimes* preclude knowledge of a knowable proposition that can change its truth value over time.

Whether someone can be described as knowing something, at a given time, can depend upon what happens at later times. This, in and of itself, is nothing remarkable – whether a decision can be described as life-changing, whether a coincidence can be described as fortunate, whether a wound can be described as mortal, whether a belief can be described as true, at a particular time, can all depend upon what happens at later times. The observation only becomes remarkable if we insist that knowledge is a mental state. For then, one's *state of mind* at a given time can depend upon what happens at later times – as though one's mind reaches out not only into the external world, but out into the *future* as well[6]. Even after a person has died it may be possible, for a short time at least, to exert an influence, by my actions, on what that person counted as knowing shortly before his death. Again, this need not, in and of itself, be particularly Earth-shattering – but what of the thought that I can still exert an influence, by my actions, on this person's *mental states*?

Treating knowledge as a mental state may not lead to the possibility of instantaneous action at a distance, but it *is* in genuine danger of leading to a possibility that might be considered even more contentious – namely, *backwards causation*. Even if we deny that a mental state switch has any definite spatial location, it seems much harder to deny that it has a definite *temporal* location – that it takes place at a specifiable time. In this case, the preceding examples will genuinely involve effects that temporally precede their causes.

If we accept that knowledge is a mental state, we will have to abandon internalism about the mental and embrace externalism instead. But the binary internalist/externalist

[6] If, as most agree, it is possible to know contingent truths about the future, then treating knowledge as a mental state will lead immediately to the result that the nature of person's mental states, at a given time, can depend on what happens at later times. I currently believe that the sun will come up tomorrow. Whether I currently *know* this depends, amongst other things, on what happens tomorrow. If knowledge is a mental state, then what mental states I'm currently in depends on what happens tomorrow. To obtain this result, we needn't engage in any substantial theorising about knowledge. Some substantial theorising is required though to generate cases of backwards switching, such as those described in the body text.

dichotomy is a very skewed way of dividing up the underlying logical space. There is only one way to be an internalist, but one can be an externalist in many ways. In truth, there is a broad spectrum of possible views, with internalism occupying one extreme end and externalism *per se* covering everything else. Treating knowledge as a mental state doesn't merely nudge us from the internalist end of the spectrum, as many of the best known externalist arguments and thought experiments arguably do – it takes us close to, if not all the way to, the opposite extreme. As I've noted, none of this is to say that knowledge categorically *must not* be treated as a mental state. The picture of the mind that this view foists upon us is one that we could, no doubt, inure ourselves to. But it is not one that should be accepted lightly.

References

Armstong, D. (1968) *A Materialist Theory of Mind* (London: Routledge)

Bauman, P. (2014) 'No luck with knowledge? On a dogma of epistemology' *Philosophy and Phenomenological Research* v89(3), pp523-551

Boghossian, P. (1989) 'Content and self knowledge' *Philosophical Topics* v17, pp5-26

Burge, T. (1979) 'Individualism and the mental' *Midwest Studies in Philosophy* v4, pp73-121

Burge, T. (1986) 'Individualism and psychology' *Philosophical Review* v95, pp3-45

Burge, T. (1988) 'Individualism and self-knowledge' *Journal of Philosophy* v85, pp649-663

Comesaña, J. (2005) 'Unsafe knowledge' *Synthese* v146(3), pp395-404

Evans, G. (1982) *The Varieties of Reference* (Oxford: Clarendon Press)

Fricker, E. (2009) 'Is knowing a state of mind? The case against' in Greenough, P. and Pritchard, D. eds. *Williamson On Knowledge* (Oxford: Oxford University Press)

Hinton, J. (1967) 'Visual experiences' *Mind* v76, pp217-227

Kaplan, D. (1979) 'Dthat' in French, P., Uehling, T. and Wettstein, H. eds. *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press)

Kaplan, D. (1989) 'Demonstratives' in Almog, J., Wettstein, H. and Perry, J. eds. *Themes From Kaplan* (new York: Oxford University Press)

Kripke, S. (1972) *Naming and Necessity* (Oxford: Blackwell)

Ludlow, P. (1995) 'Externalism, self-knowledge and the prevalence of slow-switching' *Analysis* v55, pp45-49

McDowell, J. (1982) 'Criteria, defeasibility and knowledge' *Proceedings of the British Academy* v68, pp455-479

McDowell, J. (1986) 'Singular thoughts and the extent of inner space' in McDowell, J. and Pettit, P. eds. Subject, Thought and Context (Oxford: Clarendon Press)

McGinn, C. (1977) 'Charity, interpretation and belief' *Journal of Philosophy* v74, pp521-534

McGlynn, A. (2014) *Knowledge First?* (Basingstoke: Palgrave Macmillan)

Nagel, J. (2013) 'Knowledge as a mental state' in Gendler, T. and Hawthorne, J. eds. *Oxford Studies in Epistemology Volume 4* (Oxford: Oxford University Press)

Neta, R. and Rohrbaugh, G. (2004) 'Luminosity and the safety of knowledge' *Pacific Philosophical Quarterly* v85(4), pp396-406

Pritchard, D. (2005) *Epistemic Luck* (Oxford: Clarendon Press)

Putnam, H. (1975) 'The meaning of 'meaning'' *Minnesota Studies in the Philosophy of Science* v7, pp131-193

Putnam, H. (1982) *Reason, Truth and History* (Cambridge: Cambridge University Press)

Smart, J. (1959) 'Sensations and brain processes' *Philosophical Review* v68, pp141-156

Smith, M. (2009) 'Transmission failure explained' *Philosophy and Phenomenological Research* v79(1), pp164-189

Snowdon, P. (1980) 'Perception, vision and causation' *Proceedings of the Aristotelian Society* v81, pp175-192

Sosa, E. (1999a) 'How to defeat opposition to Moore' *Philosophical Perspectives* v13, pp137-149

Sosa, E. (1999b) 'How must knowledge be modally related to what is known?' *Philosophical Topics* v26, pp373-384

Warfield, T. (1992) 'Privileged self knowledge and externalism are compatible' *Analysis* v52, pp232-237

Williamson, T. (1992) 'Inexact knowledge' *Mind* v101, pp217-242

Williamson, T. (1995) 'Is knowing a state of mind?' *Mind* v104, pp533-565

Williamson, T. (2000) *Knowledge and Its Limits* (Oxford: Oxford University Press)