For the *Routledge Handbook of Theories of Luck*, (ed.) I.M. Church. London: Routledge (forthcoming).

Epistemic Luck and the Extended Mind

J. Adam Carter

University of Glasgow

adam.carter@glasgow.ac.uk

Abstract: Contemporary debates about epistemic luck and its relation to knowledge have traditionally proceeded against a tacit background commitment to *cognitive internalism*, the thesis that cognitive processes play out inside the head. In particular, *safety-based* approaches (e.g., Pritchard 2005; 2007; Luper-Foy 1984; Sainsbury 1997; Sosa 1999; Williamson 2000) reveal this commitment by taking for granted a traditional internalist construal of what I call the *cognitive fixedness thesis*—viz., the thesis that the cognitive process that is being employed in the actual world is always 'held fixed' when we go out to nearby possible worlds to assess whether the target belief is lucky in a way that is incompatible with knowledge. However, for those inclined to replace cognitive internalism with the *extended mind thesis* (e.g., Clark and Chalmers 1998), a very different, 'active externalist' version of the cognitive fixedness thesis becomes the relevant one for the purposes of assessing a belief's safety. The aim here will be to develop this point in a way that draws out some of the important ramifications it has for how we think about safety, luck and knowledge.

1. Epistemic luck, knowledge and cognitive internalism

Most all contemporary epistemologists accept the following anti-luck platitude:

> **Anti-luck platitude**: For all *S*, *p*, if *S* knows a proposition, *p*, then *S*'s belief that *p* is not (in some to-be-specified sense) 'true by luck'.

Something like the anti-luck platitude was arguably the key take-away lesson from Gettier's (1963) counterexamples, which showed (*contra* the JTB theory) that a belief could be justified and true, and yet, true by luck in a way that seems intuitively incompatible with knowledge[1].

But the platitude needs some sharpening. In what sense, exactly, does knowledge require that a belief not be true by luck? We know of some very chancy events *that* they occurred, and so the anti-luck platitude must be unpacked in a way that is reconcilable with this fact[2]. Also, we sometimes know things (improbable or not) on the basis of evidence that we could have very easily not have come across. The detective who just so happens to catch a piece of very compelling evidence

blowing in the wind can come to *know* that a suspect is guilty—even when it was just a matter of luck that the evidence was acquired in the first place[3].

According to Duncan Pritchard (2005), the way forward is to model the crux of the anti-luck platitude modally, in terms of a *safety condition*. This strategy, while popular, is not only way to do things[4]. But it's the strategy I'll focus on here. Doing so will, I hope, most clearly help to reveal how our views about the bounds of cognition can effect our thinking about epistemic luck.

First though, some quick review. Pritchard's safety strategy takes as a starting point a *modal account of lucky events* generally speaking[5]. This account says that if any event is lucky, then it is an event that occurs in the actual world but which does not occur in a wide class of the nearest possible worlds where the relevant conditions for that event are the same as in the actual world. A fair lottery win counts as lucky on the modal account because in a wide class of the nearest possible worlds where the initial conditions (i.e., you buy a ticket) are the same as in the actual world, you lose.

One kind of event is true belief formation. On the modal account, *S*'s true belief that *p* is lucky in a way that is incompatible with knowledge (what Pritchard calls *veritically lucky*) if and only if *S*'s belief that *p* is true in the actual world α but false in nearly all nearby possible worlds in which *S* forms the belief in the same manner as in α[6].

A belief is defined as 'safe' when the truth of the belief is not (veritically) lucky in this way–viz. when in the nearest possible worlds in which *S* continues to form her belief about the target proposition in the same way as in the actual world, it continues to be true. We can now unpack the anti-luck platitude in terms of the notion of safety: if *S* knows a proposition, *p*, then *S*'s belief can't be 'true by luck' in the sense that *S*'s belief that *p* must be safe; put colloquially, if *S* knows that *p*, then *S couldn't easily have been incorrect* given the relevant way that *S* in fact forms her belief about whether *p*.

So far, so good. But what should count as the relevant way the individual forms the belief in the actual world? This matters. It effects what we will end up 'holding fixed' about the process employed in the actual world, when going out to nearby worlds and assessing whether the belief is true.

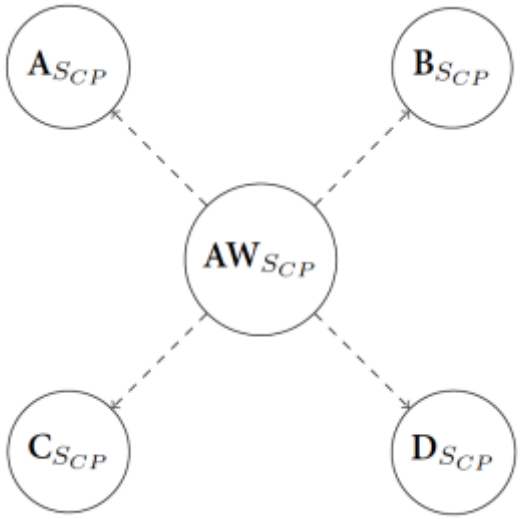
According to a rather naive way of thinking about this, the way you form a belief in the actual world, for the purposes of assessing the safety of a belief, is just a matter of whatever evidence for that belief you have in your possession in the actual world[7]. On reflection, though, this is not enough to hold fixed when looking out to other worlds. After all, the mental processes that one deploys in evaluating one's evidence could vary in reliability (perhaps dramatically[8]) across cases where the evidence one has remains exactly same. Accordingly, if on the safety account we don't at least hold fixed the mental processes one employs in the actual world in forming her belief, when moving out to possible worlds, we'd would no way of accounting for why (for example), for two individuals who have the same evidence, the person who haphazardly evaluates that evidence but luckily draws the right conclusion fails to know, while the person who reads the evidence unimpeachably and draws the right conclusion does not.

The foregoing suggests a plausible corollary to a safety condition on knowledge—call it the *cognitive fixedness thesis*.

> **Cognitive fixedness thesis**: For all *S*, *p*, and cognitive process ϕ, if ϕ is a cognitive process that *S* employs in the actual world in forming her *p*-belief, then, in evaluating the safety of

> *S*'s *p*-belief, ϕ must be held fixed when we go out to near-by possible worlds to assess whether *S*'s *p*-belief remains true.

For example, in the following diagram, suppose *AW* is the actual world, and let *A*, *B*, *C*, and *D* be close nearby worlds to the actual world *AW*. Cognitive fixedness tells us that if an individual *S* at *AW* employs a cognitive process (*CP*) in her belief that *p* in *AW*, then when we look to *A*, *B*, *C* and *D* to assess whether *S*'s belief remains true in these nearby worlds, we must also suppose that *CP* is the cognitive process that *S* employs in *A*, *B*, *C* and *D*, rather than any other cognitive process.



Crucially, *without* cognitive fixedness assumed to be in play, the safety account would be open to allowing cases where the cognitive process that's *actually* used by an agent in forming the target belief is treated as not part of the relevant way she forms her belief in the actual world, and consequently, the account would end up generate implausible results—e.g., that whether one was drunk when drawing an inference could potentially not matter for determining whether her belief is safe.

Interestingly, there are potentially conflicting ways to interpret the cognitive fixedness thesis, depending on how we think about what actually *counts* as the cognitive process one employs. The received view on this matter, *cognitive internalism* (e.g., Adams and Aizawa 2001; 2008; 2010) says the following:

> **Cognitive internalism**: An individual's mind is (in short) in her head; cognitive processes (e.g., memory, inference, introspection, etc.) are exclusively intracranial processes, which play out inside the head.

The cognitive fixedness thesis, paired with a background commitment to cognitive internalism, gives us a more explicit version of the cognitive fixedness thesis, according to which:

> **Cognitive fixedness thesis (internalism)**: For all *S*, *p*, and *intracranial cognitive process* ϕ_I, if ϕ_I is a cognitive process that *S* employs in the actual world in forming her *p*-belief, then, in evaluating the safety of *S*'s *p*-belief, ϕ_I must be held fixed when we go out to near-by possible worlds to assess whether *S*'s *p*-belief remains true.

Cognitive internalism, however, has become increasingly controversial in recent philosophy of mind and cognitive science[9]. In the next section, I will briefly sketch a more inclusive picture of a cognitive process—one that owes in a large part to work by Clark and Chalmers (1998)—and on this basis, consider an 'extended mind friendly' interpretation of the cognitive fixedness thesis and show how this thesis interfaces in surprising ways with knowledge, luck and safety.

2. Safety, cognitive fixedness and active externalism

Consider at this point the following case, due to Clark and Chalmers (1998):

> Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory.

Upon considering such a case, one might be inclined to think something along the following lines: 'It's almost as if Otto's memory is in his notebook, not his head!' After all, Otto in the above case is using the notebook to store and retrieve information, which is what we ordinarily use our biomemory to do.

What proponents of the extended mind thesis insist is that Otto's memory really is outside his head—*literally*—viz., that his memory process criss-crosses the boundaries between Otto's brain and the the world, so as to include the notebook as part of the memory process. As Clark and Chalmers put it, in terms of a parity insight:

> **Parity Principle**: If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is part of the cognitive process. (Clark & Chalmers 1998: 8)

Clark and Chalmers' rationale for this inclusive way of thinking about what to include as part of a cognitive process is aimed at safeguarding against bioprejudice—viz., giving undue consideration to things like material constitution and location when delineating cognitive processes. As they put it:

> […][T]he notebook plays for Otto the same role that memory plays for Inga; the information in the notebook functions just like the information [stored in Inga's biological memory] constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin[10] (Clark & Chalmers 1998, 13).

As Clark and Chalmers appraise the situation, Otto's memory process is a transcranial process that includes his notebook; and furthermore, Otto's memories are stored in his notebook no less than our memories are stored in biomemory. And just as, in ordinary circumstances, we are credited with non-occurrent beliefs in virtue of information that is stored in biomemory, Otto is credited with non-occurrent beliefs in virtue of information that is stored in his extended memory, i.e., the notebook. In this respect, Otto's mind is literally extended; it supervenes on parts of the world outside his skin and skull.

Clark and Chalmers' embracing of the extended mind is of course at tension with cognitive internalism. Call the denial of cognitive internalism *cognitive externalism*—viz., the denial of the claim

that (in short) necessarily, cognitive processes have intracranial material realisers. Clark and Chalmers' extended mind thesis entails cognitive externalism, though this entailment is asymmetrical[11].

For our purposes, what matters presently is this: if we think about what literally *counts* as part of a cognitive process in an externalist rather than an internalist fashion, then the familiar internalist unpacking of the cognitive fixedness thesis is no longer viable. Rather, what will be needed is an externalist gloss of the cognitive fixedness thesis, according to which:

> **Cognitive fixedness thesis (externalism)**: For all *S*, *p*, and *transcranial cognitive process* ϕ_T, if ϕ_T is a cognitive process that *S* employs in the actual world in forming her *p*-belief, then, in evaluating the safety of *S*'s *p*-belief, ϕ_T must be held fixed when we go out to near-by possible worlds to assess whether *S*'s *p*-belief remains true.

To get a feel for how the externalist cognitive fixedness thesis will have a bearing on which beliefs count as safe, consider the following variation on Clark and Chalmers' case of Otto, due to Carter (2013):

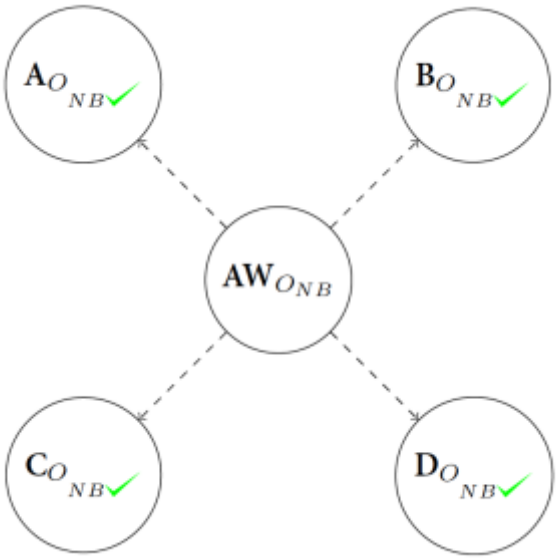
> *Notebook Jokester*: Otto consults his notebook to determine when his doctor's appointment was today, and finds the correct time, noon, written in the book. Unbeknownst to Otto, his notebook had been stolen by a jokester, who fudged with the times of Otto's other appointments that day, changing them all back an hour. The jokester, however, overlooked the doctor's appointment, leaving the original and correct time intact[12].

Against a background commitment to cognitive internalism, the state and qualities of Otto's notebook (e.g., what is written in the book, whether certain pages are missing) are the sort of thing that could vary across possible worlds along with other aspects of Otto's environment. After all, by the internalist cognitive fixedness thesis, what we are obligated to hold fixed across worlds is just the intracranial cognitive process Otto employs in the actual world, a process that does not include the qualities of the notebook he consults. The safety theorist operating with an internalist fixedness principle accordingly can allow that in some near-by worlds, the jokester changes the entry in Otto's notebook that corresponds with his doctor's appointment. And this, indeed, seems like a *prima facie* good outcome. After all, by supposing that in some near-by worlds the jokester changes the date of doctor's appointment (rather than overlooking just this one entry) we can straightforwardly account for the strong intuition in *Notebook Jokester* that the unsafety of Otto's belief undermines his knowledge; his belief could easily have been incorrect because the jokester could easily have tinkered with, rather than left alone, the doctor's appointment entry.

Things are interestingly different in *Notebook Jokester* for the 'extended mind' safety theorist who operates with a Clark/Chalmers-style externalist rather than internalist cognitive fixedness principle. Because the state of Otto's notebook is literally the state of Otto's memory, given the extended mind thesis, the state of Otto's notebook *must* be exported to all possible worlds, when—in evaluating the safety of Otto's belief—we assess whether his target belief remains true in these worlds.

But once *this* point is appreciated, a strange result materialises. Since by stipulation the jokester leaves intact Otto's correct entry specifying his doctor's appointment, the *correct entry* for the doctor's appointment is (as part of the state of the notebook, viz., the state of Otto's memory) itself exported to all worlds for the purposes of evaluating the safety of the target belief in the actual world. But this means Otto's belief in the actual world is safe! After all, in the nearest worlds (i.e., in the diagram

below, *A*, *B*, *C*, and *D*) where Otto (O) consults his notebook as he does in the actual world, he is consulting a notebook (NB) with the *correct* time written for his doctor's appointment. And those are worlds where Otto believes truly the time of his appointment—viz., he looks right at the correct entry.[13].



I noted that this is a strange result because Otto's belief is *obviously* unsafe—structurally very similar to a barn facade case. *Ex hypothesi*, in *Notebook Jokester*, the jokester really could have easily tinkered with the entry for Otto's doctor's appointment just as the jokester did with the other entries in the notebook, which would have led Otto to believe incorrectly what time the doctor's appointment was. And yet, with the externalist cognitive fixedness principle in play, the safety theorist seems forced into the awkward position of saying otherwise. As with any puzzle, something has to be rejected, but it's not clear what. To appreciate why, consider three potential salient options and their respective drawbacks:

> Option 1. *Reject the extended mind thesis*. Rationale: Otto's belief is clearly unsafe. But, since safety theorists who embrace EMT should accept the extended cognitive fixedness principle, and this principle generates the results that Otto's belief is not unsafe, the extended mind thesis should go. A drawback is that this will plausibly be the last option that one already friendly to the extended mind thesis would be inclined to pursue. But more importantly, this route seems to sidestep rather than really engage with the puzzle.
>
> Option 2. *Reject that EMT proponents should embrace the externalist cognitive fixedness principle*. Rationale: this allows one to hold on to the extended mind thesis while at the same time avoiding the impalatable result that Otto's belief is safe. A drawback, though, is that it would be unprincipled at best for the safety theorist who embraces EMT to advert to the internalist version of the cognitive fixedness principle just to achieve the result that that Otto's belief comes out unsafe. After all, the internalist version of the cognitive fixedness thesis simply makes explicit that what should be held fixed (when evaluating for safety) under the description of a 'cognitive process' are exclusively intracranial cognitive processes. But

embracing this line, for the EMT theorist, is tantamount to purporting to embrace a position while theorising in a way that suggests otherwise.

Option 3. *Deny that Otto's belief in Notebook Jokester is unsafe*. Rationale: like Option 1, this is a quick way out of the puzzle. However, this line is really a nonstarter; at least, to pursue this line, one would need to defend the position that Otto couldn't easily have been wrong, given the relevant way he formed the belief, while maintaining that (for example) in paradigmatic instances of unsafe belief which seem closely structurally similar—viz., barn facade cases—the protagonists have unsafe beliefs.

3. 'Extended' epistemic luck: new philosophical problems

In this section, I want to suggest why cases like *Notebook Jokester* raise some additional perplexities for traditional thinking about knowledge, luck and safety, beyond the puzzle sketched in §2. In particular, in §3.1, I'll show how such 'extended luck' challenge the familiar intervening/environmental luck distinction; in §3.2 I raise and reply to an objection to the epistemic significance of such cases as raised in recent work by Benjamin Jarvis (2015), and in §2.3, I close by outlining a further kind of problem brought about by the interfacing of epistemic luck and the extended mind, which I call the retroindication problem.

3.1 Taxonomising epistemic luck

In the mainstream literature on epistemic luck, there is a distinction due to Pritchard (2005) between two *kinds* of knowledge-undermining (i.e., veritic) epistemic luck: *intervening* and *environmental*. This distinction is useful in marking two structurally different kinds of Gettier cases[14]. 'Intervening luck' undermines knowledge when the unsafety of the target belief is due to luck intervening 'between the agent and the fact' in a way that is not present in cases where the unsafety of a target belief is simply a matter of her being in an environment where she easily could have been mistaken. For example, in Gettier's (1963) famous original cases, there is a disconnect between the justification the agent has for the target belief and what causes the belief to be true, a disconnect that is then regained by luck[15]. By contrast, in barn-facade-style cases (e.g., Ginet 1975; Goldman 1976), the unsafety of the target belief does not owe to any such disconnect. For instance, in Ginet's (1975) original case, nothing goes ostensibly awry. The hero in question forms a perceptually grounded belief that there is a barn in front of him, having looked directly at a genuine barn. The source of the unsafety of the belief is just down to the individual being in the environment she is in, one in which she is surrounded by fakes. In nearby worlds, she looks at a fake rather than a real barn and believes falsely.

Question: in *Notebook Jokester*, is the truth of Otto's unsafe belief about his doctor's appointment a matter of intervening or environmental luck? As it turns out, neither answer is wholly satisfying.

In one respect, the answer seems to be *environmental*. Consider that in the barn facade case—the paradigmatic environmental luck case—the hero looks at an actual barn in a circumstance under which he easily could have looked at a fake and believed falsely. And Otto's situation in *Notebook Jokester* is arguably very similar. Because the jokester overlooks Otto's doctor's appointment belief, leaving it intact and accurate, this particular piece of information is much like the 'real barn' in the barn facade case—viz., with some abstraction: one looks to something accurate which is surrounded by what is inaccurate and which one easily could have looked to. The setback, however, with this diagnosis, is this: while the safety theorist adverting to a cognitive internalist cognitive fixedness thesis can unproblematically help herself to it, the proponent of EMT can not. After all, if you are

an EMT theorist, it's hard to see why Otto should be described as in a bad *environment*. Unlike barns, the entries in his notebook—no matter what their state—are internal to his cognitive life.

The alternative of course is to think of the case instead as one where the unsafety of Otto's belief is more akin to a case of intervening luck. However, this description is problematic as well; it's not the case that there is a disconnect between the justification the agent has for the target belief and what causes the belief to be true, as is characteristic of intervening luck cases. In this respect, unsafe beliefs, in extended mind cases, would appear to require some kind of modification for a taxonomy of kinds of knowledge-undermining luck proposed initially against a tacit commitment to cognitive internalism.

3.2 Jarvis's (2015) bio-jokester

Consider the following 'bio-twist' on *Notebook Jokester*, due to Benjamin Jarvis (2015, 468):

> **Bio-Jokester**: Otto* (without Alzheimer's) has a normally functioning biological memory, which he consults to determine when his doctor's appointment was today, and ends the correct time, noon, is what he (at least seems) to remember. Unbeknownst to Otto*, his memory has been systematically altered by a jokester, who used pharmaceuticals and subliminal suggestion to plant false memories about his other appointments that day, making Otto* inclined to believe that they are an hour earlier than he used to believe. The jokester, however, overlooked the doctor's appointment, leaving the original and correct time registered in Otto*'s memory.

With reference to this creative 'intracranial twist' on the case, Jarvis casts doubt on whether cases like *Notebok Jokester* point to any special epistemological significance of extended mind cases, *per se*. In particular, Jarvis questions whether the fact that *Notebook Jokester* is an extended mind case is itself relevant either to the kind of puzzle proposed at the end of §2 or to the taxonomical problem concerning environemntal and intervening luck canvassed §3.1. He reasons, with reference to his above *Bio-Jokester* case, that cases that do not feature radically extended cognition (i.e., by the lights of the extended mind thesis) should generate the same kinds of epistemological issues. This is because Jarvis says that he doesn't see 'any reason to think that there should be a difference in verdict between JOKESTER and BIO-JOKESTER' (2015, 468).

While I am sympathetic to some extent with Jarvis' worry, I want to explain why I think the thrust of the worry can be resisted. Let me now be clear about what I grant and do not grant. I grant that that Jarvis's case appears to have as much import as *Notebook Jokester* does for our theorising about the environmental/intervening distinction, as per §3.1. In short, and for reasons that mirror what was noted about *Notebook Jokester*, if the target belief is unsafe, in Jarvis's *Bio-Jokester* case, it does not fit neatly in either the environmental nor the intervening category. This is, I think, additional evidence for supposing that views that advert to this distinction (for instance, as it plays an important role in recent debates about knowledge-how and understanding[16]) will need to clarify exactly how it is to be drawn.

However, I want to now pose a dilemma to Jarvis's diagnosis of *Bio-Jokester*. The dilemma, in short, is this: Firstly, there is reason to reject Jarvis's conclusion that the cases are in fact epistemically symmetrical, because there are reasons to suppose that Otto's belief in *Notebook Jokester* is unsafe (this was, to be clear, something that I argued that the EMT theorist who adverts to an externalist fixedness principle can't account for) where as Otto*'s belief in *Bio-Jokester* is (despite initial appearances) very plausibly *actually* safe. However—and here is the crux of the dilemma—if one

were to adjust the details of the *Bio-Jokester* case so as to control for this epistemic difference, then *Bio-Jokester* comes apart from *Notebook Jokester* in a diferent epistemic respect—viz., it will (more so than *Noteook Jokewster*) be plausibly diagnosed as a case of an unsafe belief due to intervening luck.

In order to defend this dilemma, it's important to see why, despite initial appearances, Otto*'s belief in *Bio-Jokester* is plausibly actually safe and in this respect differs from *Notebook Jokester*, whose belief ought to be diagnosed as unsafe. Consider that Otto*'s belief in his doctor's appointment in *Bio-Jokester* is unsafe only if Otto really would (given the description of his belief forming in the actual world) believe falsely in nearby worlds. But it's this point that is dubious in a way that it's not in *Notebook Jokester*. Remember that, in *Bio-Jokester*, it is a feature of the case that Otto* is storing the information relevant to organising his life in biomemory. Operating under the assumption that Otto* is appropriately epistemically vigilant, he will—like other individuals in normal circumstances—rely on *metacognitive virtues* (see, for instance, Morton 2004; Sosa 2015), in light of which Otto will be sensitive to certain kinds of abberations and incoherences. For instance, we can assume that in the default case, if Otto* attempts to retrieve one of the memories that was altered by pharmaceuticals and subliminal suggestion, he will be (upon locating the memory) be sensitive to how this fails to cohere with his other beliefs; normally functioning metacognitive faculties will plausibly 'flag' such beliefs, so that the individual does not simply automatically endorse it. This fact militates against the 'unsafety' verdict for the doctor's appointment belief in *Bio-Jokester*; even if Otto*'s doctor's appointment belief is affected by the jokester in near-by worlds, it's not clear that (with properly functioning metacognitive virtues) in thoser worlds if Otto * retrieved that entry, he would actually endorse its content, and thereby, believe falsely. Whereas, the same is not the case in *Notebook Jokester*, where, if the content of the doctor's appointment entry was altered by a jokester, it's hard to see how Otto would be in any position to flag this in a relevantly analogous way that would prevent him from automatically endorsing its content.

Of course—and here is the rest of the dilemma—one might suppose this disanalogy could be controlled for. Simply revise the case of *Bio-Jokester* so that the pharmaceuticals and subliminal inception have further deleterious effects on Otto*'s memory—viz., effects that undermine not just some of his diary entries. Rather, we could suppose the pharmaceuticals and subliminal messages disable his meta-cognitive virtues so that Otto* is in no position anymore to spot incoherences between the beliefs which are tampered with by the jokester and other beliefs he holds.

This move attains the result that Otto*'s belief in *Bio-Jokester* will be unsafe. But it gets this result at the cost of raising a further disanology: as soon as we revise *Bio-Jokester* in the fashion just described, *Bio-Jokester* (unlike *Notebook Jokester*) begins to look like a pretty clear case of intervening epistemic luck. Otto* after all would be using a now-thoroughly defective cognitive process, one riddled pretty comprehensively with pharmaceuticals and sublimation, which happens to issue a true belief on a particular occasion. In light of this dilemma, I think *Notebook Jokester* continues to raise philosophical problems for the safety theorist which can't so easily be mimicked by intracranial analogue cases.

3.2 Luck, fixedness and the retroindication problem

I want to now briefly sketch as final philosophical issue relevant to anti-luck epistemology—what I call the retroindication problem—which is uniquely raised by extended luck cases. The details of the retroindication problem might be easiest to appreciate by running a 'high-tech' variation on *Notebook Jokester*; call this 'Glitchy iCloud':

> *Glitchy iCloud*: Otto** consults his iPhone's Apple Calendar to determine when his doctor's appointment was today, and finds the correct time, noon, written in the the online diary. Unbeknownst to Otto**, an iCloud glitch has affected all of Otto**'s Apple Calendar entries; however, a further glitch which counteracts the original glitch prevents the original glitch from tampering with just the doctor's appointment entry.

Provided Otto's notebook is part of his extended cognitive process in *Notebook Jokester*, so is Otto**'s iPhone in *Glitchy iCloud*—and indeed, high-tech formulations of Otto are standard fare in the extended cognition literature[17]. With reference to the extended cognitive fixedness principle, which safety theorists inclined to the extended mind thesis should embrace, the extended cognitive process which Otto** employs in the actual world—viz., a process that includes the iPhone—must be held fixed in all worlds when we assess whether Otto**'s belief continues to be true. The retroindication problem is that, even if we hold this fixed, certain *other* things could happen in such worlds which, if these things were to happen, they would cause Otto**'s relationship with the iPhone to *no longer count as an extended cognitive process anymore*, by the lights of extended cognition, thus retroindicating that the process that includes the iPhone should not be held fixed in such worlds. This is a problem for evaluating safety that is unique to extended mind cases.

To make the retroindication problem more concrete, just add some further detail to *Glitchy iCloud*. Conider that, for the purpose of evaluating the safety of Otto**'s belief in the actual (*Glitchy iCloud*) world, some worlds will be worlds where the following is true: the glitch (which we hold fixed, as is, as part of the state of Otto**'s mind) will cause a media campaign against Apple, one which has as a consequence that Otto** (in such worlds) is constantly exposed to news stories and testimony according to which he is told *not* to trust any information on an Apple device. In such worlds, Otto** subjects all entries on his iPhone calendar to intense critical scrutiny and does not automatically endorse the information stored in the calendar.

But in worlds where Otto** fundamentally mistrusts his iPhone, Otto**'s interaction with the it *no longer qualifies* (at least, by Clark's light's) as an extended cognitive process; Otto** will be, in short, not relying on the device anymore in a way that is analogous to the way we rely on biomemory in the intracranial case. But if this is right, then it retroindicates that Otto**'s extended process (from the actual world) should *not* (and contrary to the externalist cognitive fixedness thesis) be assumed to be the cognitive process that he employs at such worlds; the entries in the online calendar, in such worlds, are no longer part of a *cognitive* process.

The retroindication problem, thus, calls into question the coherence of maintaining the externalist cognitive fixedness principle, by revealing how what counts (by the lights of EMT) as an extended cognitive process in the actual world may not be exportable to all worlds *as* an extended cognitive process. This is a further issue, generated by extended mind cases, that I think deserves attention by philosophers interested in the anti-luck ramifications of radically extended cognition.

4. Concluding remarks

The connections between the extended mind thesis, in the philosophy of mind and cognitive science, and mainstream epistemology, are still in early days[18]. Here my aim has been to raise what I think are some of the more interesting philosophical issues the extended mind thesis poses for the specific area of epistemology concerned with the relationship between knowledge and luck. I've argued that on at least one natural way of thinking about this relationship, in terms of safety, the

extended mind thesis generates some new perplexities. In particular, I've argued that EMT proponents ought to, when assessing the safety of a target belief, embrace a version of the cognitive fixedness thesis that comports with the EMT conception of what should be held fixed under the description of a cognitive process in the actual world.

However, I've shown that once a suitably 'externalist' fixedness principle is in play, it's hard to maintain intuitive judgments about safety, and furthermore, it becomes less clear how we should distinguish between environmental and intervening epistemic luck, a distinction that is much more straightforward in cases where what's fixed under the description of a cognitive process is the intracranial cognitive process employed in the actual world. I've considered and responded to some challenges to the significance of the kind of puzzles I've raised, and I concluded by raising yet another problem—the retroindication problem—which I think extended-mind-friendly safety theorists must grapple with. I have not attempted to solve the problems I've raised here. I am, as of yet, not convinced what the right way forward will be to address them. My more modest aim is to show that extended mind cases have important and interesting import for how we theorise about knowledge and luck, and to make explicit where some of these points of interest lie.

References

Adams, F and Aizawa K (2001). "The Bounds of Cognition." *Philosophical Psychology* 14 (1): 43–64.

——— (2008) *The Bounds of Cognition*. Blackwell.

——— (2010) "Defending the Bounds of Cognition." In *The Extended Mind*, edited by Richard Menary. MIT Press.

Bondy, P. and Pritchard, D. (forthcoming) "Propositional Epistemic Luck, Epistemic Risk, and Epistemic Justification." *Synthese*, 1–10. doi:10.1007/s11229-016-1262-2.

Broncano-Berrocal, F. (2015) "Luck as Risk and the Lack of Control Account of Luck." *Metaphilosophy* 46 (1): 1–25.

——— (2016) "Epistemic Luck." *Internet Encyclopedia of Philosophy*.

Carter J.A. (2013) "Extended Cognition and Epistemic Luck." *Synthese* 190 (18): 4201–14.

Carter J.A., and Broncano-Berrocal, F. (2017) "Epistemic Luck." In *Routledge Encyclopedia of Philosophy*, edited by Tim Crane, 1–15. Routledge.

Carter, J. Adam, and Peterson, M. (2017) "The Modal Account of Luck Revisited." *Synthese* 194(6): 2175-2184.

Carter, J. Adam, and Pritchard, D. (2015a) "Knowledge-How and Cognitive Achievement." *Philosophy and Phenomenological Research* 91 (1): 181–99.

——— (2015b) "Knowledge-How and Epistemic Luck." *Noûs* 49 (3): 440–53.

Carter, J.A. Gordon, E.C. and Palermos, S.O. (2015) "Extended Emotion." *Philosophical Psychology* 29 (2): 198–217.

Carter, J.A, Kallestrup, J, Palermos, S.O and Pritchard, D. (2014) "Varieties of Externalism." *Philosophical Issues* 24 (1): 63–109.

Church, I.M. (2013) "Getting Lucky with Gettier." *European Journal of Philosophy* 21 (1): 37–49.

Clark, A. (2003) *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford University Press.

——— (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Clark, A. and Chalmers, D. (1998) "The Extended Mind." *Analysis* 58 (1): 7–19.

Coffmann, E.J. (2007) "Thinking About Luck." *Synthese* 158 (3): 385–98.

Engel, M. (1992) "Is Epistemic Luck Compatible with Knowledge?" *Southern Journal of Philosophy* 30 (2): 59–75.

Gettier, E. (1963) "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–23.

Ginet, C. (1975) *Knowledge, Perception and Memory*. Dordrecht: D. Reidel Publishing Company.

Goldman, A. (1976) "Discrimination and Perceptual Knowledge." *Journal of Philosophy* 73 (20): 771–91.

Greco, J. (2004) "A Different Sort of Contextualism." *Erkenntnis* 61 (2-3): 383–400.

Hales, S.D. (2014) "Why Every Theory of Luck Is Wrong." *Noûs*. doi: 10.1111/nous.12076

Hutchins, E. (1995) *Cognition in the Wild*. MIT press.

Jarvis, B. (2015) "Epistemology and Radically Extended Cognition." *Episteme* 12 (4): 459–78.

Lackey, J. (2008). "What Luck Is Not." *Australasian Journal of Philosophy* 86 (2): 255–67.

Luper-Foy, S. (1984) "The Epistemic Predicament: Knowledge, Nozickian Tracking, and Scepticism." *Australasian Journal of Philosophy* 62 (1): 26–49.

Menary, R. (2006) "Attacking the Bounds of Cognition." *Philosophical Psychology* 19 (3): 329–44.

Moore, G. E. (1993) *G.E. Moore: Selected Writings*. Routledge.

Morton, A. (2004) "Epistemic Virtues, Metavirtues, and Computational Complexity." *Noûs* 38 (3): 481–502.

Palermos, S.O. (2011) "Belief-Forming Processes, Extended." *Review of Philosophy and Psychology* 2 (4). Springer: 741–65.

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

——— (2007) "Anti-Luck Epistemology." *Synthese* 158 (3). Springer: 277–97.

——— (2010) "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175 (1). Springer: 133–51.

——— (2013) "There Can Not Be Lucky Knowledge." In *Contemporary Debates in Epistemology*, edited by Matthias Steup and John Turri, 152–64. Blackwell Oxford.

——— (2014) "The Modal Account of Luck." *Metaphilosophy* 45 (4-5): 594–619.

——— (2015a) "Anti-Luck Epistemology and the Gettier Problem." *Philosophical Studies* 172 (1): 93–111.

——— (2015b) "Risk." *Metaphilosophy* 46 (3): 436–61.

——— (Forthcoming) "Epistemic Risk." *Journal of Philosophy*.

Riggs, W. (2009) "Luck, Knowledge, and Control." In *Epistemic Value*, edited by Pritchard, Haddock, and Millar, 204–21. Oxford: Oxford University Press.

Sainsbury, R.M. (1997) "Easy Possibilities." *Philosophy and Phenomenological Research* 57 (4): 907–19.

Sosa, E. (1999) "How to Defeat Opposition to Moore." *Noûs* 33 (13): 141–53.

——— (2015) *Judgment and Agency*. Oxford: Oxford University Press.

Sutton, J. (2010) "Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process." In *The Extended Mind*, edited by Richard Menary. Cambridge: MIT Press.

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

Wilson, R.A. (2000) *The Mind Beyond Itself*. New York: Oxford University Press.

Zagzebski, L. (1994) "The Inescapability of Gettier Problems." *The Philosophical Quarterly* 44 (174): 65–73.

Zimmerman, M.J. (1987) "Luck and Moral Responsibility." *Ethics* 97 (2): 374–86.

NOTES

[1] For discussion on this point, see Church (2013), Pritchard (2015a), Carter and Broncano-Berrocal (2017, sec. 1), and Broncano-Berrocal (2016, sec. 1c).

[2] Consider one very liberal reading of 'true by luck': *S*'s belief that *p* is true by luck iff it's a matter of luck *that p is true*. The anti-luck platitude, unpacked in this way, says that if *S* knows a proposition, then it is not it's a matter of luck *that p is true*. But this is obviously false given that we know all sorts of improbable propositions. Let *p* be the proposition *I was just dealt a royal flush in Texas Hold 'Em poker!* Suppose *p* is true, and thus that you were dealt a hand of cards the probability of which your being dealt was just 0.0032%, a frequency of 1 in 4,324 hands. After the cards are all dealt and the royal flush sits clearly on the table, you obviously *know* it's a royal flush. It's improbability or disconnection from your skill is by the by.

[3] See here Engel (1992, 66–72), Pritchard (2005, 136) and Greco (2004, 398).

[4] The modal account is not uncontroversial. For example, according to the lack-of-control account of luck, an event is lucky for a given agent just when it is significantly enough beyond that agent's control (see, for example, Broncano-Berrocal 2015; Coffmann 2007; Riggs 2009; Zimmerman 1987). While the modal account will be helpful for illustrative purposes, in light of the points that will be made here about the cognitive internalism and the extended mind thesis, similar conclusions to those I'll draw in §3 could be reframed in terms of the lack-of-control account of luck. For some recent scepticism about either of these accounts of luck, see Lackey (2008) and Hales (2014).

[5] E.g., Pritchard (2014). For a recent modification of this account in light of some objections raised to previous formulations, see Carter and Peterson (forthcoming). See also (Pritchard 2015b; Forthcoming) for presentations of Pritchard's recent transition from anti-luck epistemology to anti-risk epistemology.

[6] See also Pritchard (2013).

[7] For a more detailed discussion of evidence and luck, in connection with the propositional/doxastic justification distinction, see Bondy and Pritchard (forthcoming).

[8] Sherlock, for example, could evaluate evidence against Moriarty in normal conditions, or he could evaluate this evidence while under the influence of crystal DNT, a drug that causes generally reliable but believable hallucinations. If the hallucination-infused evaluation of the evidence led Sherlock to believe Moriarty was guilty, then (unlike in the normal good case) Sherlock would be in no better a position, epistemically, than the Duke of Devonshire who, in a case noted by Moore (1993, 189), 'once dreamt that he was speaking in the House of Lords and, when he woke up, found that he was speaking in the House of Lords'.

[9] For a sample of some recent challenges to cognitive internalism, see for example Clark and Chalmers (1998), Clark (2008), Sutton (2010), Hutchins (1995), Palermos (2011), Menary (2006) and Wilson (2000).

[10] Clark (2008) grants that, just as in the case of ordinary biological memory, the availability and portability of the resource of information should be crucial (see for discussion also Carter, Gordon, and Palermos 2015). Accordingly, Clark provides a set of functionalist criteria that must be satisfied by non-biological candidates for inclusion into an individual's mind: 1) 'That the resource be reliably available and typically invoked'. 2) 'That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny […] It should be deemed about as trustworthy as something retrieved clearly from biological memory'. 3) 'That information contained in the resource should be easily accessible as and when required' Clark (2008, 46).

[11] Not all strategies of resisting cognitive internalism involve endorsing the full extended mind thesis. For example, the embedded cognition thesis is arguably more conservative than the extended mind thesis while nonetheless parting ways with cognitive internalism. For a taxonomy of various varieties of cognitive (and epistemic) externalism and their relationship to one another, see Carter et al. (2014).

[12] Note that we can abstract from this case to form a kind of recipe for generating similar cases, including 'high tech' versions of the case. For instance, just suppose Otto is storing his information not in a notebook but in his Apple Calendar on his iPhone; further, suppose a computer glitch affects all of his entries but, due to a further glitch which counteracts the original glitch, prevents the original glitch from tampering with the doctor's appointment entry. I consider a case of this sort in §3.3.

[13] After all, in most worlds where Otto consults his notebook, replete with the correct entry, he believes truly. This much is part and parcel with the thought that Otto is simply a vigilant user of his notebook (as is necessary to meet

Clark's glue and trust conditions—see for example Pritchard (2010)), one who by stipulation is using his notebook in a way that is analogous to the way we store and retrieve information in biomemory.

[14] See, for example, Pritchard (2015a), Carter and Broncano-Berrocal (2017, sec. 1), and Broncano-Berrocal (2016, sec. 1c). The distinction has theoretical importance in debates about understanding and knowledge-how, standings which have been argued to be compatible with environmental but not epistemic luck, unlike propositional knowledge which is incompatible with both varieties. To the extent that this is right, it would count against proposals which identify understanding and knowledge-how with knowledge-that. See, for example, Carter and Pritchard (2015b) for a presentation of this kind of argument.

[15] In Gettier's Smith/Jones case, the justification Smith has is for the proposition 'The man who will get the job has 10 coins in his pocket' is justification Smith has for the proposition that Jones will get the job and has 10 coins in his pocket'. But what makes the target proposition true has nothing to do with Jones whatsoever. It's that Smith got the job and had 10 coins in his pocket. Note that this regained disconnect that typifies intervening luck should not be conflated with the more general phenomenon described by Zagzebski (1994) as 'double luck', which is general enough to subsume both intervening and environmental cases. After all, it is bad luck that protagonist in a barn facade case happens to be in an epistemically unhospitable environment but good luck that perception issues a true belief on the particular occasion that it does. But this does not mean that the source of the justification for the target proposition and what causes it to be true are disconnected in the way we find it to be in intervening luck cases.

[16] For an overview of how the intervening/environmental luck distinction has import for these debates, see for example [Carter and Pritchard (2015a); Carter and Pritchard (2015b);].

[17] See, for example, Clark (2003) for a range of such cases.

[18] For a forthcoming volume of essays on the relationship between active externalist approaches in the philosophy of mind and cognitive science and mainstream epistemology, see Carter et al. (forthcoming).

ACKNOWLEDGEMENTS

I am especially grateful to Ian Church for helpful comments on a previous version of this paper. I'm also thankful to Andy Clark, Bolesław Czarnecki, Emma C. Gordon, Benjamin Jarvis, Jesper Kallestrup, Duncan Pritchard and Orestis Palermos for many fruitful discussions about issues at the intersection of epistemology and the extended cognition and mind theses.

BIOGRAPHICAL NOTE

J. Adam Carter is a lecturer in philosophy at the University of Glasgow, working mainly in epistemology. His work has appeared in (among other places) *Noûs*, *Philosophy and Phenomenological Research*, *Analysis*, *Philosophical Studies* and the *Australasian Journal of Philosophy*. Carter's first monograph, *Metaepistemology and Relativism*, was published in 2016 with Palgrave MacMillan