

Revised 30 June 2020. For K. Loughheed & J. Matheson (Eds.) *Epistemic Autonomy* (Routledge, *forthcoming*)

Epistemic Autonomy and Externalism

J. Adam Carter

University of Glasgow

adam.carter@glasgow.ac.uk

The philosophical significance of attitudinal autonomy—viz., the autonomy of attitudes such as beliefs—is widely discussed in the literature on moral responsibility and free will. Within this literature, a key debate centres around the following question: is the kind of attitudinal autonomy that's relevant to moral responsibility at a given time determined entirely by a subject's present mental structure at that time? Internalists say 'yes', externalists say 'no'. In this essay, I motivate a kind of distinctly *epistemic attitudinal autonomy*, attitudinal autonomy that is relevant to knowledge. I argue that regardless of whether we are externalists or internalists about the kind of attitudinal autonomy that is relevant for moral responsibility, we should be externalists about the kind of autonomy that a belief must have to qualify as knowledge.

1. Suppose Prometheus is tied to a ship, bound by ropes so that he can't move an inch. Prometheus is considerably less autonomous than he would be were his ropes cut. But even in this predicament—where he is completely unable to physically act or affect his environment—there remains a sense in which Prometheus is autonomous in a way he would *not* be were he not only physically bound but also drugged and hypnotised.

This difference is a useful reference point for distinguishing between two broad types of personal autonomy: *outward-directed autonomy*—which is what Prometheus lacks in virtue of being tied up—and *inward-directed autonomy*, which is what he retains even when shackled (but *not* when drugged or hypnotised).¹

The species of personal autonomy that this chapter will have as its focus is exclusively *inward-directed* personal autonomy—viz., autonomy of the *mind*. But the focus will be much more narrow than this. For one thing, there are two main

¹For further discussion of this distinction, see Mele (2001, 144–6). As Mele notes, 'inward-directed' autonomy is also elsewhere described as 'psychological autonomy' (2001, 138).

ways of thinking about *structure* of the property that inward-directed autonomy picks out. Here we can distinguish between autonomy as (i) a global property of persons, taken as a whole, and as (ii) a property of particular attitudes of a person, such as beliefs and desires. My interest will be the latter—viz., in *attitudinal autonomy*.²

Discussions of attitudinal autonomy are almost entirely restricted to the literature on moral responsibility.³ In such debates, the central focus has been how to spell out the kind of attitudinal autonomy that *matters* for the purposes of moral, rather than epistemic, evaluations. I suggest that there is also an interesting kind of attitudinal autonomy—what I’m calling *epistemic attitudinal autonomy* (hereafter: *epistemic autonomy*)—that matters for knowledge.

Here is the plan for what follows. §2 articulates a case for thinking that (i) there is a distinctive kind of epistemic autonomy, autonomy of *belief*, that matters for propositional knowledge (§2.1); and (ii) that this kind of attitudinal autonomy is importantly different from the kind that matters for moral responsibility (§2.2). §3 shows what an internalist account of epistemic autonomy would look like and argues that any such account faces intractable problems. §4 considers two broad ways to be an externalist about epistemic autonomy: *counterfactual externalism* and *history-sensitive externalism*. The former is shown to have its own problems, whereas a version of the latter offers much more promise.⁴

2.1 Let’s take as a starting point Keith Lehrer’s (1990) classic case of Mr. Truetemp:

TRUETEMP: Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp’s head so that the very tip of the device, no larger than the head of pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn,

²For the present purposes, I’ll remain neutral about what the relationship is between attitudinal autonomy and global autonomy, viz., the kind of autonomy that applies to a person as a whole. One natural way of thinking about this relationship is as follows: one’s global autonomy is a matter of one’s attitudinal autonomy. So, one is autonomous as a person just in case one’s relevant attitudes are autonomous attitudes. See, however, Young (1980) for criticism of this picture.

³For some representative discussions, see Mele (2001, 2003); Fischer (2011); Fischer and Ravizza (2000); Weimer (2009); Cyr (2019); Cuypers (2006); and Levy (2011).

⁴For a more detailed treatment of the case for history-sensitive externalism about epistemic autonomy, see Carter (2020, Ch. 2).

sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? (1990, 162–3)

The predominant view in mainstream epistemology is that Mr. Truetemp does *not* know that the temperature is 104 degrees in the above case, despite the reliability⁵ of his belief-forming process.⁶ There's little consensus, though, as to why.⁷ Lehrer himself, along with William Alston (1988), both reason along the following lines: Truetemp lacks knowledge of the relevant temperature belief because (i) knowledge requires epistemic justification; and (ii) there is at least some additional justified belief that Truetemp lacks but which he'd need in order to be epistemically justified in believing that it is 104 degrees. For Lehrer, that 'extra' belief is a 'metabelief' about the reliability of the temperature-implant based process; for Alston, the extra belief just needs to be a belief that could serve as a good reason for believing that it's 104 degrees (a reason Truetemp presently lacks). If Truetemp had those beliefs, and appropriately based his temperature belief on them, then he'd be a knower.

While something like this might look plausible as a diagnosis of Lehrer's *original* version of the case, it's not hard to think up variations on the case where nei-

⁵Although the reliability of the process is, arguably, built in to the description of the case—and as such is not a point of dispute—one creative line of dissent is due to Jennifer Nagel (2016). Nagel's contention is that while the thermometer is reliable enough, it is a separate matter whether Truetemp's *belief-forming* process is reliable, and on this point, her assessment is that it is not, as it involves a kind of unreflective endorsement that is generally unreliable. For a critique of Nagel's reading of the case, see Goldman (2016). Note that the matter of whether Nagel's line would work on variations of Lehrer's case remains open (e.g., where the details of the implantation are varied) even if it were granted that her 'unreliability' diagnosis is a correct assessment of the original case.

⁶See however, Beebe (2004) for a dissenting view here.

⁷It's worth noting that Lehrer's case is often discussed together with Bonjour's (1980) case of Norman the Clairvoyant. These cases are, as Sosa (1991) calls them, 'metaincoherence' cases, in that they both involve a lack of coherence between the source of the reliability of the process and the subject's own view of this source. Despite this commonality between the cases, there are also epistemic asymmetries which make the TrueTemp case comparatively more useful in the context of discussing epistemic autonomy.

ther the Lehrer nor Alston lines would work. For example, suppose we run a twist on the case that holds everything fixed except the following: what the scientists implant is more sophisticated than the tempucomp: call it the ‘Tempucomp_{Deluxe}’. The Tempucomp_{Deluxe} not only compels⁸ Truetemp to believe the target proposition (i.e., that it’s 104 degrees), but it *also* compels him to believe a further proposition ‘X’, where we can fill in ‘X’ with either a Lehrer-style metabelief or an Alston-style reason. Then—through a powerful form of hypnotism—the Tempucomp_{Deluxe} closes the circle by causing Truetemp to base his belief on the relevant reason.⁹

On a Lehrer- or Alston-style diagnosis, it looks as though Mr. Truetemp is now in the clear, knowledge-wise. And yet, he’s surely *not*. Here’s why he’s not: if there is something epistemically objectionable (in the sense of being knowledge-incompatible) about Mr. Truetemp original’s belief, there should, by parity of reasoning, be something epistemically objectionable about the new beliefs which he’s acquired in just the same way (e.g., compulsion by the implanted mechanism). In short, if one doesn’t know the temperature on the basis of a gadget-compelled belief, then neither does one know the temperature on the basis of a gadget-compelled belief one is compelled to support on the basis of equally gadget-compelled beliefs.

An entirely different story for why Mr. TrueTemp fails to know—one that looks initially as though it might fare better in the Tempucomp_{Deluxe} version of the case—is due to John Greco (2010) and Duncan Pritchard (2010, 2012). The Greco-Pritchard diagnosis of the original Truetemp goes as follows: (i) knowledge must derive from *cognitive ability* in the sense that the correctness of a *known* true belief must be because of the manifestation of a cognitive ability¹⁰; (ii) Truetemp’s true belief derives from a reliable process but *not* from any cognitive ability of his; (iii) Therefore, Truetemp doesn’t know. The rationale for (ii) is, in short, that knowledge generating cognitive abilities must be appropriately *integrated* into a thinker’s wider cognitive architecture. What constitutes the

⁸We are continuing to hold fixed the the compelled belief results from the sophisticated causal mechanism Lehrer describes.

⁹Suppose this is either a causal basis or a doxastic basis—or a combination of both—depending on what kind of account of the epistemic basing relation one favours. For discussion, see Korcz (2019) and Carter and Bondy (2019).

¹⁰Note that Greco and Pritchard disagree about the extent to which the correctness of a true belief must be creditable to ability in order for one to know. Greco’s line is that it must be *primarily* creditable to cognitive ability. Pritchard, along with Lackey (2007), take this to be too strong given that the condition seems not to be satisfied in cases of testimonial knowledge acquisition in friendly epistemic environments. Pritchard’s preferred formulation is accordingly weaker than Greco’s; for Pritchard the correctness of a known belief must be at least *significantly* creditable to cognitive ability, even if not primarily creditable. For discussion, see Pritchard (2012).

right kind of cognitive integration is a complicated issue.¹¹ But here's one notable idea, due to Greco: integrated dispositions are at least sensitive to the operation of other belief-forming dispositions. The disposition Truetemp has to form temperature beliefs, however, is plausibly not sensitive to other dispositions he has for forming beliefs (2010, 150–2). It's just controlled entirely by the mechanism.

It looks, *prima facie*, like the Pritchard-Greco line might work not just as a diagnosis for why Truetemp lacks knowledge, but also as one that carries over to Tempucomp_{Deluxe}. After all, in that case, when Truetemp believes correctly, it doesn't seem to be down to any properly integrated cognitive ability that he has, and this is so even if (as is stipulated on that version of the case) he has some additional beliefs that stand in support of the target proposition.

But notice that we can just pull the same trick again! Just imagine a further twist on the case, where Truetemp has installed an even more impressive device—viz., a 'Tempucomp_{SUPER-Deluxe}'—one that significantly rewires his cognitive architecture in such a way that he now is compelled not only to believe truly what the temperature is (104 degrees) on the basis of a reliable disposition, but also—thanks to the Tempucomp_{SUPER-Deluxe}—that reliable disposition has been 'auto-integrated' by the device. In this version of the case, we can think of Truetemp now having an ability, albeit, one that he is being compelled by the device to exercise in such a way as to end up with the belief he does.

It's not hard to see a general pattern emerging here. When a thinker is caused to believe a proposition (even if reliably) in a manner such that the acquisition of the belief 'bypasses' the thinker's own exercise of her cognitive faculties (as it does in all variations of the TrueTemp case considered), then two things seem to follow. First, (i) the intuition that the thinker lacks knowledge, viz., as in the original Truetemp case, carries over; and second, (ii) that a viable explanation for *why* this knowledge fails to be present can't simply point to some epistemic (roughly: truth-linked) condition on knowledge which is said not to be satisfied. A *different* kind of explanation is needed. At least, this seems to be the lesson from the foregoing discussion.¹²

Let's now take a step back. At this juncture, there are broadly three paths available:

¹¹For some representative discussions, see Carter and Kallestrup (2019); Palermos (2011, 2014); Menary (2007); Andrada (2019); Clark (2015); Carter and Kallestrup (2017); and Carter and Palermos (2015b, 2015a).

¹²The reason for (ii) is that, plausibly, for any proposed epistemic condition we might appeal to in an attempt to explain why the no-knowledge intuition in (i) holds, we can simply imagine a variation of that case where the following conjunction applies: that proposed epistemic condition is satisfied *and* it's the case that the acquisition of the target belief bypasses the thinker's own exercise of her cognitive faculties.

- *Option 1*: (a) Grant that one lacks knowledge when the acquisition of the target belief bypasses the thinker’s own exercise of her cognitive faculties; (b) Insist that there is some plausible epistemic condition (e.g., some kind of epistemic justification condition) on knowledge that can’t *possibly* be satisfied by further and even fancier iterations of the ‘Tempucomp_{SUPER-Deluxe}’.¹³
- *Option 2*: Bite the bullet and *deny* that one lacks knowledge when the acquisition of the target belief bypasses the thinker’s own exercise of her cognitive faculties (e.g., as in the original Truetemp case).
- *Option 3*: (a) Grant that one lacks knowledge when the acquisition of the target belief bypasses the thinker’s own exercise of her cognitive faculties; but (b) *deny* that there is some plausible epistemic condition (e.g., some kind of epistemic justification condition) on knowledge that can’t *possibly* be satisfied by further and even fancier iterations of the ‘Tempucomp_{SUPER-Deluxe}’.

The prevailing literature has generally gone, *de facto*, for Option 1 or, in rare cases, Option 2.¹⁴ But what the pattern we’ve seen in this section suggests is that Option 3—which thus far has remained unexplored—might be the most plausible avenue for those who don’t want to bite the bullet and maintain that Truetemp knows.

But if Option 3 *is* right way to go, then an interesting implication follows for the theory of knowledge: it looks like knowledge requires not just epistemically justified belief, but (in some sense to be articulated) *epistemically autonomous* belief, where the the relevant kind of ‘epistemic autonomy’ must be such that, with reference to it, we could explain why ‘Tempucomp_{SUPER-Deluxe}’-style cases *in principle* are not cases of knowledge.

I’m going to assume from here on out that Option 3 is worth exploring.¹⁵ And so the question driving the rest of the chapter will be:

¹³Also in logical space is the idea that when the acquisition of the target belief bypasses the thinker’s own exercise of her cognitive faculties, she thereby acquires a defeater for that belief, and this explains why she lacks knowledge. Such a defeater would best be understood not as a psychological or mental state defeater (given that there is no such defeater *ex hypothesi*), but as a *normative* defeater. A normative defeat diagnosis would seem to occupy a mixed space between Option 1 and Option 3. I am setting aside this option for consideration here because there is no straightforward way to articulate this diagnosis in terms of a proposition Truetemp both ought to believe and is such that it would count against the target proposition. However, for a more complete story on how defeaters interface with an epistemic autonomy condition on knowledge, see Carter (2020 Ch. 3).

¹⁴See Beebe (2004). For an alternative diagnosis that does not fall neatly into any of these categories, see Kaplan (2018), whose position is that the matter of whether Truetemp knows lacks any methodological import, and that the theorist should simply remain agnostic on the point.

¹⁵For a detailed defence of pursuing this strategy, see Carter (2020 Ch. 1).

Guiding Question: In what sense, exactly, does propositional knowledge require epistemically autonomous belief? Put another way: what should an epistemic autonomy condition on propositional-knowledge-apt belief look like?

2.2. The most interesting fault line for answering the above question is—as we’ll shortly see—one between *internalist* and *externalist* approaches to the epistemic autonomy of beliefs. But before digging in to this issue, it’s worth briefly addressing the following question: “Won’t an account of the kind of attitudinal autonomy that matters for moral responsibility also work just fine as an account of the kind of attitudinal autonomy (of beliefs specifically) that matters for knowledge?”

The answer to this question is ‘no’. A quick and easy way to see why these issues come apart will be to focus on how they clearly do so in cases of *self-arrangement*—viz., where one’s lack of attitudinal autonomy at a later time is intentionally pre-arranged by one at a previous time.

Consider two versions of an incapacitated driving case:

- *Version 1*: Someone forces you to ingest a potent cocktail of hallucinogenic drugs, puts you behind the wheel of a moving vehicle, and then—driving this vehicle under the influence of the drugs—you cause a wreck.
- *Version 2*: Everything is the same except you *chose* to take the hallucinogenic drugs.

The prevailing thinking in the moral responsibility literature is that while your attitudes (e.g., beliefs, desires, perceptions, emotions) lack the kind of attitudinal autonomy that matters for moral responsibility in Version 1, this isn’t the case in Version 2. And this is so even though your attitudes are equally at the mercy of the strong drugs in both cases when you cause the wreck. What is said to make the difference is *self-arrangement*;¹⁶ in Version 2, you (unimpeded by any drugs) chose to take the drugs that would later have the effects on you that they did. So these attitudes, no matter the effect the drugs have on them, remain autonomous *in the sense that matters for moral responsibility* later on (e.g., when you’re behind the wheel) because these effects are self-arranged.

So does self-arrangement make a difference when it comes to whether a belief is autonomous in the way that matters for *knowledge*, as it does when what’s at issue is the kind of attitudinal autonomy that matters for moral responsibility? To get a clearer grip on this, let’s compare self-arrangement versions of the drunk

¹⁶See, for discussion, Mele (2001, 176), Fischer and Ravizza (2000, 50), and Carter (2018).

driving case we've already considered (Version 1 and Version 2) with the original (non-self-arrangement) and self-arrangement variations on our three tempu-comp cases.

	No self-arrangement	Self-arrangement
Drunk Driving	No Moral Responsibility (V1)	Moral Responsibility (V2)
Tempucomp	No Knowledge	?
Tempucomp _{Deluxe}	No Knowledge	?
Tempucomp _{SUPER-Deluxe}	No Knowledge	?

If self-arrangement really made a difference with respect to whether a belief is autonomous in the way that matters for knowledge in a way that is analogous to the difference it makes (*vis-à-vis* attitudinal autonomy) in the case of moral responsibility, then we should expect that—in self-arrangement versions of our tempu-comp cases—there would be, by parity of reasoning, no epistemic-autonomy related reason for withholding knowledge (just as there is no autonomy related reason for withholding moral responsibility).

But this is the wrong result. The import of self-arrangement—once we shift focus from moral responsibility to knowledge—turns out to be clearly disanalogous. Just consider: while the matter of whether you took the drugs or had them forced upon you *obviously* bears on whether you are morally responsible later for being in whatever state you're in (e.g., incapacitated and thus dangerous behind the wheel), it's not at all evident that our verdict on Truetemp, *qua* candidate knower, should change in the slightest if we added to the Truetemp backstory that Truetemp willingly at some point in the past paid a superscientist to experiment on him. The addition of this kind of historical fact seems entirely irrelevant to whether he counts as knowing, at a later time, when affected as he is by what the scientist does to him.¹⁷

Summing up, then: (i) propositional knowledge requires not just epistemically justified belief, but *epistemically autonomous* belief (*conclusion of §2.1*); and (ii) the *kind* of attitudinal autonomy (*viz.*, of beliefs) that matters for knowledge is different from the kind of attitudinal autonomy that matters for moral responsibility (*conclusion of §2.2*)—which means that to get an account of it in view, we

¹⁷This is not to say that it would be irrelevant to whether Truetemp knows if in addition to self-arranging the procedure (e.g., by paying a superscientist), Truetemp also came to know the ins and outs of how the device works. (For related discussion on this point, see Carter 2018, 2020, Ch. 2). Rather, the idea here is just that the relevance of the fact of prior self-arrangement is clear when what's at issue is moral responsibility, and not when what's at issue is whether one possesses knowledge.

need to look beyond the kinds of accounts already available in the literature on moral responsibility.

3. What is the *nature* of the kind of epistemic autonomy a belief must have in order to qualify as knowledge? Here's one potential answer:

Internalism about epistemic autonomy (IEA): The knowledge-relevant (viz., epistemic) autonomy of a belief at a time, T, is determined entirely by the subject's present mental structure at T.

If IEA is true, then (in slogan form) psychological twins at a time do not differ in the epistemic autonomy of their beliefs at that time.¹⁸

In the literature on internalism about epistemic justification, psychological twins cases are often used in the service of supporting an internalist view of epistemic justification—viz., one on which psychological twins at a time do not differ with respect to the epistemic justification they have for their beliefs at that time. This idea is, at any rate, at the heart of the New Evil Demon thought experiment against epistemic externalism.¹⁹

It would be natural to expect that, if drawing our attention to psychological twins cases is a move that is supposed to lend intuitive support for internalism about epistemic justification, then so it would as well for internalism about epistemic autonomy. However, the opposite seems to be true.

Consider the following case, which is an epistemic twist on a case used by Mele (2001, 145) to argue against internalism about attitudinal autonomy.

PSYCHOLOGICAL TWINS: Ann and Beth are psychological twins. They are identically mentally constituted. Both believe that Cicero's scribe was named Tiro. Ann believes this because she read it in a book. Beth believes it because scientists want her to be psychologically identical to Ann, and so they brainwash her until her psychology – as it pertains to all matters of Roman history – matches Ann's exactly.

If IEA is true, then Ann—who is a paradigmatic knower—satisfies an autonomous belief condition on knowledge if and only if Beth does. But Beth looks

¹⁸Examples of internalism about epistemic autonomy would be (epistemic variations on) internalist views of attitudinal autonomy, such as those defended by Frankfurt (1988) and Dworkin (1981).

¹⁹See, e.g., Cohen (1984) and Lehrer and Cohen (1983) for presentations of the problem, and Littlejohn (2009) for an overview.

sure to fail any plausible construal of an autonomous belief condition on knowledge; she is effectively brainwashed. Beth is, at best, a Truetemp.

There are two important take-away points here. First, internalism about epistemic autonomy is false.²⁰ But, secondly, it's worth emphasising that this verdict is neither implied by, nor implies, a denial of internalism about epistemic justification. After all, it's possible for all that's been said that Beth and Ann are equally epistemically justified in believing that Cicero's scribe is Tiro (in virtue of their matching psychology post-brainwashing) even while they differ with respect to the kind of epistemic autonomy that matters for knowing.²¹

4. On the assumption here on out that internalism about epistemic autonomy is a non-starter, the focus will now be squarely on epistemic autonomy and externalism. According to externalism about epistemic autonomy:

Externalism about epistemic autonomy (EEA): It's not the case that the knowledge-relevant (*viz.*, epistemic) autonomy of a belief at a time, T, is determined entirely by the subject's present mental structure at T.

EEA is a minimal conception of externalism about epistemic autonomy in that it consists merely in the denial of IEA. (Compare: a minimal conception of externalism about epistemic justification consists just in a denial of internalism about epistemic justification.²²)

²⁰At least, (IEA), which is a 'time-slice' thesis is false. There could potentially be other ways of carving up the internalism/externalism divide, *vis-à-vis* epistemic autonomy, such that PSYCHOLOGICAL TWINS would not count against all internalist views, even if it counts against (IEA). See, for example, the discussion later in this paper of 'counterfactual externalism', which could (on some possible ways of carving up the landscape) qualify as an internalist view.

²¹Additionally, a denial of internalism about epistemic autonomy is compatible (given the conclusions from §2.2) with various ways of thinking about the kind of attitudinal autonomy that matters for moral responsibility. Given that these have been shown to be fundamentally different issues, it is at least possible to be, for example, an externalist about the kind of attitudinal autonomy that matters for knowledge while being an internalist about the kind of attitudinal autonomy that matters for moral responsibility. That said—and it is beyond the scope of what I can do to argue for this here—it's worth noting that Mele's original version of the PSYCHOLOGICAL TWINS case is (e.g., 2001, 145), independently, militates against going for an internalist view about attitudinal autonomy relevant to moral responsibility. For related discussion, see Weimer (2009) and Carter and Pritchard (2020 Ch. 2).

²²Of course, just as there are different versions of internalism about epistemic justification (e.g., accessibilism vs. mentalism), you can have distinct minimal conceptions of epistemic externalism that consist in denials of different formulations of epistemic internalism. For discussion, see Carter et al. (2014).

Substantive externalist theses in epistemology involve not only a denial of internalism, but also a positive thesis about what it is in virtue of which something possesses positive epistemic status when it does. For example: *process reliabilists* about epistemic justification (e.g., Goldman 1979, 1999) submit—in addition to simply denying internalism—that what makes a belief epistemically justified is the reliability of the process that issues the belief.

This section will consider two substantive forms of externalism about epistemic autonomy: *counterfactual externalism* (§4.1) and *history-sensitive externalism* (§4.2).

4.1. One way to be a substantive externalist about the kind of attitudinal autonomy that matters for moral responsibility is to take a ‘counterfactual’ approach of the sort that has been defended in various places by John Christman (e.g., 1991, 2007). According to Christman, the question of whether a given belief, P, is autonomous in the way that matters for moral responsibility is one we can settle by asking how the subject, in the here and now, would respond to P, *if she were to critically reflect on P in light of (an accurate description of) its origins*.

This is an externalist account because what *makes* the relevant belief autonomous can be something external to the subject’s psychology. (After all, we might at any given time be clueless as to how we *would* respond *were* we to critically reflect on a given attitude; for the counterfactual externalist, all that matters is how, in fact, we would respond under appropriate conditions of reflection.)

With a bit more precision, the proposal is as follows:

Counterfactual externalism about moral-responsibility relevant attitudinal autonomy: Necessary and sufficient for an attitude’s autonomy, at a time, t, is that (i) the agent is able to adequately reflect on the attitude P at t; where (ii) ‘adequate reflection’ requires the possession of a representation of an alternative to that attitude, Q, one that would be necessary for the subject to realistically imagine an alternative to P; and (iii) were the agent at t to adequately reflect on P in light of a ‘minimally adequate account’ of P’s developmental history, she would not be alienated from it.²³

With reference to this kind of account, Christman diagnoses why, for instance, certain traits one has would *not* be autonomous (in the sense relevant to moral responsibility) even if one has never in fact repudiated them and indeed even if one positively endorses them as their own. He gives the following case to lend support to this idea:

²³I borrow this succinct summation of Christman’s view from Weimer (2009, 186).

PIANO: We can imagine a person who finds out she has been severely abused in her childhood by someone who is responsible for several of the proclivities and skills she has developed. In coming to grips with the memories of the abuse, she has repudiated many of those proclivities. But she doesn't realize or remember that this person also taught her to play the piano, which she still loves to do and does well. If she were told that her piano-playing was also rooted in her time with this abuser, she would feel alienated from that part of herself also; but in her ignorance she plays on contentedly. Does this person count as autonomous? On my view she does not, since, were she to reflect on the trait in light of its origins she would be alienated from it (2007, 22–23).

There are some well-known worries for this kind of proposal—to both the necessity and sufficiency of satisfying the counterfactual condition.²⁴ These objections won't concern us here.

Rather, what will be of interest is what an 'epistemic twist' on a Christman-style counterfactual externalism might look like, and whether it could offer a plausible account of epistemic autonomy (regardless of whether it is viable or not as an account of the kind of attitudinal autonomy relevant to moral responsibility.)

With this in mind, consider the following variation on the view:

Counterfactual externalism about epistemic autonomy: Necessary and sufficient for a belief's epistemic autonomy, at a time, *t*, is that (i) the agent is able to adequately reflect on the belief *P* at *t*; where (ii) 'adequate reflection' requires the possession of a representation of an alternative to that attitude, *Q*, one that would be necessary for the subject to realistically imagine an alternative to *P*; and (iii) were the agent at *t* to adequately reflect on *P* in light of a 'minimally adequate account' of *P*'s developmental history, she would not be alienated from it.

For a case (broadly analogous to the PIANO case) that might intuitively motivate this kind of view, consider the following case of 'therapy Truetemp':

THERAPY TRUETEMP: Suppose that Mr. Truetemp²⁵ is taken to a therapist, to help him make sense of why he thinks so often,

²⁴See Weimer (2009), and particularly, his discussion of the case of Dora and Cara (Weimer 2009, 190).

²⁵Mr. Truetemp from the original case—suppose—though any of the versions from §2.1 will do.

seemingly inexplicably, about the temperature. Through this therapy, Truetemp uncovers suppressed memories of the operation. After some further digging, he comes to believe that many (but not all) of his temperature beliefs are the result of this operation and, as a result, feels alienated from them—such that he desires to give these beliefs up. But, *unbeknownst* to him, the belief he has presently that it's 104 is among those sourced in the operation; since he's unaware of this, he does not in fact feel alienated in any way from this belief.

Counterfactual externalism about epistemic autonomy generates the *prima facie* intuitive result here that, in THERAPY TRUETEMP, the mere fact that Mr. Truetemp doesn't *actually* appreciate the origin of the target belief (that it is 104 degrees) and repudiate it on that basis should count for naught if in fact he *would have* done so were he to have reflected on it while appreciating its origins.

Nonetheless, counterfactual externalism about epistemic autonomy both doubly misses the mark—satisfying this kind of counterfactual condition is neither necessary nor sufficient for epistemic autonomy.

Let's consider the sufficiency leg of the view first. Suppose Truetemp reflects on a belief of his about the temperature (e.g., that it was 84 degrees yesterday at noon) which was *in fact* not the result of the Tempucomp at all, but instead a result of trusting an expert meteorologist. Now, suppose Truetemp—because he was treated badly by a meteorologist as a child and has discovered the meteorologist to be the source of his belief—feels alienated from it. Counterfactual externalism about epistemic autonomy generates the implausible result that Truetemp's belief that it was 84 degrees yesterday at noon *lacks* the kind of autonomy that matters for knowledge.

The necessity leg of the view also runs into problems. Firstly, the view generates the wrong result in the case of mathematical knowledge. All mathematical knowledge (as well as logical knowledge) looks like it's going to be ruled as not epistemically autonomous with reference to counterfactual externalism simply because it will fail the 'adequate reflection' clause, and regardless of the way one came to possess that knowledge in the first place. In paradigmatic cases of mathematical knowledge, we will not be in a position to 'realistically imagine an alternative' to the known belief given the necessity of mathematical truths.²⁶ And what goes for mathematical truths plausibly also goes for analytic truths. (When reflecting on your belief that bachelors are unmarried, can you 'realistically imagine' an alternative?)

²⁶For a recent discussion of the relationship between mathematics as a subject matter and necessity truths, see Yli-Vakkuri and Hawthorne (2018).

A second problem for the necessity leg of counterfactual externalism about epistemic autonomy concerns the alienation condition that is central to it. Suppose Truetemp is a transhumanist—deeply influenced by the thinking of Ray Kurzweil²⁷. Truetemp, having discovered the origins of his Tempucomp-generated belief that is 104 degrees, does *not* feel alienated from it in any way, owing to his philosophical view that he is (in the words of Andy Clark 2003) a ‘natural born cyborg’. It’s implausible that whether Truetemp’s belief that it’s 104 degrees has the kind of autonomy that matters for whether he *knows* that it’s 104 degrees depends in any way whatsoever on whether he aligns himself philosophically with Kurzweil’s and Clark’s transhumanist thinking or, instead, rejects transhumanism for bioconservativism.²⁸

4.2. Here’s a summary of where we’ve got to:

- (a) Propositional knowledge requires not just epistemically justified belief, but *epistemically autonomous* belief (*conclusion of §2.1*).²⁹
- (b) The *kind* of attitudinal autonomy (viz., of beliefs) that matters for knowledge is different from the kind of attitudinal autonomy that matters for moral responsibility (*conclusion of §2.2*).
- (c) We should reject *internalism* about knowledge-relevant (i.e., epistemic) autonomy, the view that the epistemic autonomy of a belief at a time, T, is determined entirely by the subject’s present mental structure at T. (*conclusion from §3*).
- (d) *Counterfactual externalism* is implausible as a substantive form of externalism about epistemic autonomy (*conclusion from §4.1*).

In this section, we’ll look at a substantive form of externalism about epistemic autonomy that has real promise.

The view I want to now defend takes as a starting point two conditions that form the backbone of Mele’s (2001) *history-sensitive externalism* about the kind of attitudinal autonomy that matters for moral responsibility. These two key conditions are:

²⁷See, e.g., Kurzweil (2005).

²⁸For some representative statements of bioconservativism, which takes a normative stance against embracing any sort of cyborg-style future for humanity, see Sandel (2009) and Kass (2004).

²⁹A separate question, of course, is whether and to what extent epistemic justification requires epistemically autonomous belief. It’s beyond the scope of what I can do here to take this up in detail. However, there is good reason to think that epistemically justified belief—and not merely knowledge—also requires epistemically autonomous belief.

- a **bypass condition**—viz., a condition pertaining to whether the attitude in question was acquired in a way that ‘bypassed’ the subject’s cognitive faculties.
- an **unsheddability condition**—viz., a condition pertaining to whether the subject is able to give up, or at least attenuate the strength of, the relevant attitude.

According to Mele, an attitude is autonomous in the way that matters for moral responsibility only if it has a certain kind of *history* (and, regardless of whether one is aware of that history). In particular, the attitude has to have a history that is free from *compulsion*. And an attitude has a compulsion-free history only if it’s *not* the case that the acquisition of the attitude satisfies both the bypass and unsheddability conditions, spelled out in a particular way.

I will set aside entirely whether Mele’s history-sensitive externalism is a viable account of attitudinal autonomy relevant to moral responsibility. Rather, what I want to show is that *versions of* these two key conditions can be used to frame a very plausible externalist account of *epistemic* autonomy.

With this in mind, let’s get a simple version of the view on the table.

History-sensitive externalism about epistemic autonomy (HSEEA):

S’s belief that p is epistemically autonomous (viz., autonomous way that is necessary for propositional knowledge) at a time, t, if and only if p has a compulsion-free history at t; and this is a history it has if and only if it’s not the case that S came to acquire her belief that p in a way that: (i) bypasses S’s cognitive faculties, and (ii) the bypassing of such faculties issues in S’s being unable to shed P.

HSEEA looks strong from the start. For one thing, it looks, at least *prima facie*, as though it deals nicely with a range of cases we’ve considered so far. Beginning with PSYCHOLOGICAL TWINS: whereas internalism can’t explain why Ann’s belief is epistemically autonomous but Beth’s is not, HSEEA has a simple answer: Beth’s belief has a compulsion history and Ann’s doesn’t.

Likewise, HSEEA is not threatened by the kinds of cases that made trouble for the necessity and sufficiency legs of counterfactual externalism, and this is because the matter of how one *would* reflect on a given belief one has acquired is irrelevant to epistemic autonomy on HSEEA. Granted, the counterfactual condition of counterfactual externalism looked initially like it would be essential to get the right result in cases like THERAPY TRUETEMP, the epistemic analogue to Christman’s PIANO case. *However*, HSEEA gets the right result in that case as well. It follows from HSEEA that Therapy Truetemp’s belief that it’s 104 degrees is *not* epistemically autonomous. This is not, on HSEEA, because he would

have repudiated it had he reflected on it in light of an accurate description of its origins. Rather, it's because that belief has a compulsion history, with reference to the bypass and unsheddability clauses of HSEEA.

And even more, notice that HSEEA looks well-suited to deal not only with the original Truetemp case, but also with the kinds of 'epistemically improved' variations on the case we considered in §2.1 which were used to motivate an epistemic autonomy condition on propositional knowledge in the first place. What these cases exploited after all was the fact that (in imagined scenarios) the epistemic justification-related credentials of a belief could be attained in ways that bypass the subject's exercise of any of her cognitive faculties. A bypass condition of the sort that features in HSEEA is exactly the kind of condition that can rule out these cases *in principle* as cases where knowledge is present.

Question: given the work a bypass condition is able to do, isn't the addition of an unsheddability condition theoretically superfluous?

The answer is that it's not superfluous. To see why, consider a further twist on the Truetemp case, one where sheddability is present.

TRUETEMP-SHEDDABLE: In the original version of the case, it's not said explicitly, but it is implied, that Mr. TrueTemp can't easily shed this belief (by eradicating it or attenuating its strength). He's at any rate stuck with it. Let's now suppose that, on the present variation of the case, this is explicitly not so. Mr. TrueTemp can easily shed the belief, by simply judging the content to be false (e.g., in light of other things he believes) or otherwise attenuating its strength. Finally, let's suppose he elects *not* to revise this belief in any way, despite having the power to, after subjecting it to (non-compelled) rational scrutiny, including scrutiny by which he comes to find out that the mechanism he's using is a reliable one.

It's not at all evident that Truetemp doesn't know that it's 104 degrees in TRUETEMP SHEDDABLE, even though the history of the belief includes the fact that it was acquired in a way that clearly bypassed Truetemp's cognitive faculties. In Truetemp sheddable (unlike in the original case, presumably) the power to revise or give up the belief is entirely Truetemp's, and the fact that he has *not* revised this particular belief (given the new evidence he has) has nothing whatsoever to do with the workings of the tempucomp.

The situation would be different if, unlike in TRUETEMP SHEDDABLE, we make explicit that whatever beliefs are caused by the tempucomp are unrevisable in any way, such that *even if* Truetemp acquired new evidence against such beliefs, this evidence would have no sway for him.

In light of the above, the initial presentation of HSEEA will include an unsheddability condition along with a bypass condition.

That said—unfortunately—HSEEA can not quite stand up to scrutiny without further refinements. In what follows, I want to describe three problems for HSEEA. None requires that we abandon the core HSEEA-style view, but these three problems require three distinct improvements on HSEEA.

The first problem is a kind of *preemption* problem which goes as follows: you can't 'bypass' a cognitive faculty that isn't there to bypass. However, presumably, if a tempucomp was installed in infancy prior to the development of any cognitive faculties, then provided a tempucomp-issued belief caused at such a point was unsheddable, it looks like it should fail an epistemic autonomy condition on knowledge no less than in a case where the cognitive faculties of a thinker were first developed *and then* bypassed by the tempucomp.

Fortunately, there is an easy fix here, which will be to tweak the bypass component so that it is a disjunctive condition that includes either bypassing or *preempting* cognitive faculties. Let's say that the subject comes to possess a belief in a way that preempts one's cognitive faculties if and only if the process that issues the belief lacks the opportunity (trivially) to bypass the subject's cognitive faculties. The resulting tweaked version of HSEEA is as follows:

History-sensitive externalism about epistemic autonomy (HSEEA)*:

S's belief that p is epistemically autonomous (viz., autonomous way that is necessary for propositional knowledge) at a time, t, if and only if p has a compulsion-free history at t; and this is a history it has if and only if it's not the case that S came to acquire her belief that p in a way that: (i) bypasses **or preempts** S's cognitive faculties, and (ii) the bypassing **or preemption** of such faculties issues in S's being unable to shed P.

In short, HSEEA* can deal with 'infant hardwiring' cases whereas the original HSEEA can not. So far so good.

But there remain more substantive challenges. The second that I want to consider—the *opportunity problem*—concerns the 'cognitive faculty' dimension of the bypass clause. To get a feel for this problem, let's consider two possible variations on the Truetemp backstory:

- *Variation 1*: The neurosurgeons, prior to implementing the tempucomp, are allowing Truetemp to play a role in the customisation of the device prior to fitting—by giving Truetemp an option to prearrange exactly which beliefs (at a later time) will be compelled by the device. Truetemp, however,

is very drunk, and agrees (capable only of muddled thought) to each belief the neurosurgeons propose that the tempupcomp will later compel him to endorse.

- *Variation 2*: The neurosurgeons, prior to implementing the tempucomp, are allowing Truetemp to play a role in the customisation of the device prior to fitting—by giving Truetemp an option to prearrange exactly which beliefs (at a later time) will be compelled by the device. Truetemp is sober and clear-headed; however—in a room with no books or computer—he is deprived of any information on the basis of which to assess, for any given belief the neurosurgeons propose the tempucomp will later compel him to endorse, whether it would be reasonable to agree to have it later so compelled. He accordingly makes each decision arbitrarily, by simply flipping a coin.

Without further qualification, it looks as though the bypass clause is *not* satisfied in either Variation 1 or Variation 2. But if that's right, then HSEEA* implies, counterintuitively, that—whereas Truetemp in the original case lacks knowledge because his belief is not epistemically autonomous—this is *not* the case if the backstories in either Variation 1 or Variation 2 are added.

Bizarrely, then, what HSEEA* appears to imply that your cognitive faculties haven't been 'bypassed' (in the way that matters for epistemic autonomy of a given belief) so long they were *in any manner whatsoever* exercised in a way that made a difference to the acquisition of the target belief. A more plausible picture would hold that Truetemp—even though he trivially exercised his faculties (while in poor shape, and poorly situated)—in the course of acquiring the target belief(s), his doing so made no difference with respect to the epistemic autonomy of those beliefs. Put another way, a better formulation of the bypass clause in HSEEA* will allow us to say that Truetemp is in effectively the same position, *vis-à-vis* the epistemic autonomy of the beliefs he acquires in Variations 1 and 2, as he is in the original case.

A strategy for a fix, which will get Variations 1 and 2 right, is to incorporate the following idea: to replace the generic 'cognitive faculties' with 'cognitive *competences*', where—and following a well-known idea in epistemology due to Ernest Sosa (2007, 2015)—a cognitive or epistemic *competence* is not merely a disposition to reliably attain epistemic ends (e.g., true beliefs), but a disposition to do so when (i) in *proper shape*; and (ii) *properly situated*.³⁰ By way of analogy: it doesn't

³⁰Sosa articulates the three components of a competence as involving a 'seat, shape, and situation' component—viz., a 'triple S' structure. A *complete* competence incorporates all three elements. For example, you exercise your complete competence to drive a car only if you actually are awake and alert (proper shape) and on normal roads (proper situation). See, e.g., Sosa (2010).

count against one's competence to drive a car if one would drive off the road when drugged or placed on abnormally slick roads. Likewise: it doesn't count against, e.g., your perceptual, reasoning and memory-related competences if, in exercising them, you gain false beliefs when when mentally compromised (e.g., improper shape) or improperly situated (e.g., in a room with no possibility to acquire the kind of evidence that would normally bear on whether to accept the target proposition.)

Whereas the acquisition of the relevant beliefs acquired in Variations 1 and 2 above do *not* bypass Truetemp's cognitive faculties, they *do* bypass his relevant cognitive *competences* in these cases. Thus, what Variations 1 and 2 suggest is a transition from HSEEA* to HSEEA**:

History-sensitive externalism about epistemic autonomy (HSEEA):**

S's belief that p is epistemically autonomous (viz., autonomous way that is necessary for propositional knowledge) at a time, t, if and only if p has a compulsion-free history at t; and this is a history it has if and only if it's not the case that S came to acquire her belief that p in a way that: (i) bypasses or preempts S's cognitive **competences**, and (ii) the bypassing or preemption of such **competences** issues in S's being unable to shed P.³¹

While HSEEA** can handle all the cases we've discussed thus far, there remains one further important area where the condition needs a refinement, and which this section will conclude by discussing. We can call this third issue the *remote unsheddability* problem.

The crux of the remote unsheddability problem is illustrated in cases like the following. Stipulate that Truetemp's acquisition of some belief, X, bypassed his cognitive competences, as per HSEEA**. But now suppose Truetemp's belief that X is not, strictly speaking, unsheddable, simply because it is *remotely* possible that it be shedded, even though it would not be shedded through any normal course of competent inquiry.

Remote sheddability cases like the above reveal the need for an amendment to the unsheddability clause in HSEEA** in order to avoid the unwanted result that almost no belief, any any circumstance, will be unsheddable.

The natural fix here is to block an unrestricted modal reading of 'unsheddability' by restricting the relevant class of worlds to nearby worlds—thus replacing 'unable to shed' with unable to '*easily enough*' shed, as follows:

³¹While this position might seem like an 'internalist' view on some ways of cutting up internalism/externalism about epistemic autonomy, it comes out externalist on my proposed dividing lines in light of being at odds with (IEA). See fn. 22 for further discussion.

History-sensitive externalism about epistemic autonomy (HSEEA)^{*}:**

S's belief that p is epistemically autonomous (viz., autonomous way that is necessary for propositional knowledge) at a time, t, if and only if p has a compulsion-free history at t; and this is a history it has if and only if it's not the case that S came to acquire her belief that p in a way that: (i) bypasses or preempts S's cognitive competences, and (ii) the bypassing or preemption of such competences issues in S's being unable to **easily enough** shed P.

HSEEA^{***}, unlike HSEEA^{**}, can make sense of the fact that—for example—Truetemp's belief couldn't change from (i) lacking the kind of epistemic autonomy that's necessary for knowledge, to (ii) possessing that kind of autonomy, simply were we to alter the belief's sheddability profile such that there is a far-off world (e.g., one where he is struck just right by a bolt of lightning, remapping his cognitive architecture) in which the belief could be shed.

5. Let's recap. Although a flat-footed history-sensitive externalist account of epistemic autonomy—i.e., HSEEA—outperforms its internalist and counterfactual externalist competitors, HSEEA nonetheless faced three distinct kinds of problems discussed in §4.2: (i) the *preemption problem*, (ii) the *opportunity problem*, and (iii) the *remote unsheddability problem*.

I've discussed these problems separately because each forces a different *kind* of revision to HSEEA. The version we finished with, HSEEA^{***}, reflects all three revisions to the original HSEEA account, and so has the resources to handle all three kinds of problems.

HSEEA^{***} might well benefit from further technical refinements, given that the view cobbles together different working parts. The aim here is not to make those further refinements, but rather, to have at least made enough of the key refinements to illustrate what the view has the power to do in comparison with internalist and externalist alternatives.³²

References

Alston, William P. 1988. "An Internalist Externalism." *Synthese* 74 (3): 265–83.

³²Thanks to Jonathan Matheson and Kirk Lougheed for helpful feedback. This paper was written as part of the Leverhulme-funded 'A Virtue Epistemology of Trust' (#RPG-2019-302) project, which is hosted by the University of Glasgow's COGITO Epistemology Research Centre, and I'm grateful to the Leverhulme Trust for supporting this research.

- Andrada, Gloria. 2019. "Mind the Notebook." *Synthese*, 1–20. <https://doi.org/10.1007/s11229-019-02365-9>.
- Beebe, James R. 2004. "Reliabilism, Truetemp and New Perceptual Faculties." *Synthese* 140 (3): 307–29.
- BonJour, Laurence. 1980. "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5: 53–73.
- Carter, J. Adam. 2018. "Virtue Epistemology, Enhancement, and Control." *Metaphilosophy* 49 (3): 283–304.
- . 2020. *The Future of Knowing: Knowledge, Radical Enhancement, and Epistemic Autonomy*.
- Carter, J. Adam, and Patrick Bondy. 2019. *Well-Founded Belief: New Essays on the Epistemic Basing Relation*. Routledge.
- Carter, J. Adam, and Jesper Kallestrup. 2017. "Extended Circularity." In *Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard. Oxford: Oxford University Press.
- . 2019. "Varieties of Cognitive Integration." *Noûs*. <https://doi.org/https://doi.org/10.1111/nous.12288>.
- Carter, J. Adam, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard. 2014. "Varieties of Externalism." *Philosophical Issues* 24 (1): 63–109.
- Carter, J. Adam, and S. Orestis Palermos. 2015a. "Active Externalism and Epistemic Internalism." *Erkenntnis* 80 (4): 753–72.
- . 2015b. "Active Externalism and Epistemology." *Oxford Bibliographies*. <https://doi.org/10.1093/OB0/9780195396577-0285>.
- Carter, J. Adam, and Duncan Pritchard. 2020. "Extended Entitlement." In *New Essays on Epistemic Entitlement*, edited by Peter J Graham and Nikolaj Pederson.
- Christman, John. 1991. "Autonomy and Personal History." *Canadian Journal of Philosophy* 21 (1): 1–24.
- . 2007. "Autonomy, History, and the Subject of Justice." *Social Theory and Practice* 33 (1): 1–26.
- Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of*

- Human Intelligence*. Oxford University Press.
- . 2015. “What ‘Extended Me’ Knows.” *Synthese* 192 (11): 3757–75.
- Cohen, Stewart. 1984. “Justification and Truth.” *Philosophical Studies* 46 (3): 279–95.
- Cuyppers, Stefaan E. 2006. “The Trouble with Externalist Compatibilist Autonomy.” *Philosophical Studies* 129 (2): 171–96.
- Cyr, Taylor W. 2019. “Why Compatibilists Must Be Internalists.” *The Journal of Ethics* 23 (4): 473–84.
- Dworkin, Gerald. 1981. “The Concept of Autonomy.” *Grazer Philosophische Studien* 12: 203–13.
- Fischer, John Martin. 2011. *Deep Control: Essays on Free Will and Value*. Oxford University Press.
- Fischer, John Martin, and Mark Ravizza. 2000. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.
- Frankfurt, Harry G. 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.
- Goldman, Alvin. 2016. “Reply to Nagel.” In *Alvin Goldman and His Critics*, edited by Hilary Kornblith and Brian McLaughlin, 253–56. Oxford: Blackwell.
- Goldman, Alvin I. 1979. “What Is Justified Belief?” In *Justification and Knowledge*, 1–23. Springer.
- . 1999. “Knowledge in a Social World.”
- Greco, John. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- Kaplan, Mark. 2018. *Austin’s Way with Skepticism: An Essay on Philosophical Method*. Oxford University Press.
- Kass, Leon R. 2004. *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*. Encounter books.
- Korcz, Keith Allen. 2019. “The Epistemic Basing Relation.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2019. <https://plato.stanford.edu/archives/fall2019/entries/basing-epistemic/>; Metaphysics Research Lab, Stanford University.

- Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. Penguin.
- Lackey, Jennifer. 2007. "Why We Don't Deserve Credit for Everything We Know." *Synthese* 158 (3): 345–61.
- Lehrer, Keith. 1990. *Theory of Knowledge*. Routledge.
- Lehrer, Keith, and Stewart Cohen. 1983. "Justification, Truth, and Coherence." *Synthese* 55 (2): 191–207.
- Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press UK.
- Littlejohn, Clayton. 2009. "The New Evil Demon Problem." *Internet Encyclopedia of Philosophy*.
- Mele, Alfred R. 2001. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press on Demand.
- Mele, Alfred R. 2003. "Agents' Abilities." *Noûs* 37 (3): 447–70.
- Menary, Richard. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Springer.
- Nagel, Jennifer. 2016. "Knowledge and Reliability." In *Alvin Goldman and His Critics*, edited by Hilary Kornblith and Brian McLaughlin, 237–56. Oxford: Blackwell.
- Palermos, S. Orestis. 2011. "Belief-Forming Processes, Extended." *Review of Philosophy and Psychology* 2 (4): 741–65.
- . 2014. "Knowledge and Cognitive Integration." *Synthese* 191 (8): 1931–51.
- Pritchard, Duncan. 2010. "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175 (1): 133–51.
- . 2012. "Anti-Luck Virtue Epistemology." *Journal of Philosophy* 109 (3): 247–79.
- Sandel, Michael J. 2009. *The Case Against Perfection*. Harvard University Press.
- Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge University Press.

- . 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*. Oxford University Press.
- . 2010. “How Competence Matters in Epistemology.” *Philosophical Perspectives* 24 (1): 465–75.
- . 2015. *Judgment and Agency*. Oxford: Oxford University Press.
- Weimer, Steven. 2009. “Externalist Autonomy and Availability of Alternatives.” *Social Theory and Practice* 35 (2): 169–200.
- Yli-Vakkuri, Juhani, and John Hawthorne. 2018. “The Necessity of Mathematics.” *Noûs*.
- Young, Robert. 1980. “Autonomy and Socialization.” *Mind* 89 (356): 565–76.