

Virtue Epistemology, Enhancement and Control

J. Adam Carter

University of Glasgow

Abstract: An interesting aspect of Ernest Sosa's (2017) recent thinking is that *enhanced* performances (e.g., the performance of an athlete under the influence of a performance-enhancing drug) fall short of aptness, and this is because such enhanced performances do not issue from genuine competences on the part of the agent. In this paper, I explore in some detail the implications of such thinking in Sosa's wider virtue epistemology, with a focus on cases of *cognitive enhancement*. A certain puzzle is then highlighted, and the solution proposed draws from both the recent moral responsibility literature on *guidance control* (e.g., Fischer and Ravizza 1998; Fischer 2012) as well as from work on cognitive integration in the philosophy of mind and cognitive science (e.g., Clark and Chalmers 1998; Clark 2008; Pritchard 2010; Palermos 2014; Carter 2017).

1. According to Ernest Sosa's virtue epistemology (e.g., Sosa 2009, 2010b, 2010a, 2015, *passim*), the normative structure of epistemic performances can be fruitfully modelled within a wider framework for assessing performances more generally. The basic features of the wider framework are as follows:

- A. Any performance with an aim can be evaluated along three dimensions: (i) whether it is *accurate*, (ii) whether it is *adroit*, and (iii) thirdly, whether it is accurate *because* adroit.
- B. A performance in some domain of endeavour *D* is accurate because adroit when its success issues from a (complete) *D* competence; such performance are *apt*.
- C. A competence, in a given domain of endeavour, is a disposition to perform well in that domain of endeavour.
- D. Competences have a 'triple-S' constitution—*seat*, *shape* and *situation*—with reference to which three kinds of dispositions can be distinguished: the innermost competence (*seat*), the inner competence (*seat + shape*), and the complete competence (*seat + shape + situation*).

The theoretical core of (A-D) are Sosa's two 'triples': (i) the 'AAA' structure of the normativity of performances (accurate/adroit/apt); and the 'SSS' structure of the constitution of competences (seat/shape/situation).

To appreciate how the 'triple S' (i.e., SSS) structure of the constitution of a competence works, just consider the illustrative example Sosa offers concerning one's competence to drive a car:

With regard to one's competence in driving, for example, we can distinguish between (a) the innermost driving competence that is seated in one's brain, nervous system, and body, which one retains even while asleep or drunk; (b) a fuller inner competence, which requires also that one be in proper shape, that is, awake, sober, alert, and so on; and (c) complete competence or ability to drive well and safely (on a given road or in a certain area), which requires also that one be well situated, with appropriate road conditions pertaining to the surface, the lighting, etc. The complete competence is thus anSSS (or an SeShSi) competence (2017, 191-92)

We test for an innermost competence to drive a car by asking: Would the driver drive reliably enough¹ *if in proper shape and properly situated*? (We don't test for a driving competence by asking: would the driver perform reliably enough if deprived of oxygen and placed on a slick road.) Likewise, the possession of a visual-perceptual competence requires just that one's perceptually formed beliefs are reliably enough correct when one is in proper shape (i.e., awake, alert) and properly situated (not in the dark, in thick fog, etc.).

Crucially, for Sosa, a performance may be apt *even if* it issues (non-deviantly) from a complete competence (hereafter, competence) in circumstances under which the obtaining of certain elements of the competence is *unsafe*—viz., when you are in the proper shape and properly situated, but very easily might not have been. (Imagine, for example, your hitting a bullseye manifests your competence to shoot reliably enough under conditions where you might easily have been in improper shape because you might easily have been struck by a bolt of lightning, but luckily were not). That said, a performance is not apt *unless* its success manifests such a competence.

This all matters for epistemology because beliefs are themselves a kind of epistemic performance, one that aims at truth. Apt belief, on Sosa's view—viz., belief whose correctness manifests an epistemic competence—is *knowledge*. More specifically, apt belief is *animal* knowledge. Animal knowledge (like any other kind of apt

¹What counts as 'reliable enough' for Sosa is determined by the norms governing particular performance domains. See Sosa (2010) for discussion.

performance) must manifest a complete competence, but doesn't require that all of the individual components of the complete competence *safely* hold. Thus, one might attain animal knowledge even in circumstances where (for instance) the shape or situational components of the first-order competence could easily, and unbeknownst to the thinker, have *not* been in place². What Sosa calls *reflective knowledge* demands more, that one's apt belief is aptly noted—viz., that the belief's aptness *itself* manifests a second-order competence of the thinker, one that requires taking a reliable perspective on one's first-order competence and environmental conditions such that not too easily would one have formed one's first-order belief inaptly.

2. We may, for the present purposes, set the complexities of reflective knowledge aside³. The focus will be more basic—apt belief (animal knowledge) and, even more fundamentally, the SSS components of a complete competence.

Here is a question rarely asked: *Does it matter how one comes to acquire the relevant seat/shape?* In the default case, where one forms a visual-perceptual belief by looking at an object while sober/awake and in normal lighting conditions, the *acquisition conditions* of the seat/shape don't seem particularly interesting. (Consider, after all, that on Sosa's view, one can believe aptly *even when* one very easily could not have been in proper shape, but is, and could likewise easily have been improperly situated, but was not. One might be inclined to reason thusly: (i) Sosa says explicitly that it doesn't matter, *vis-à-vis* an apt performance, whether any of the SSS elements of a competence might easily have not obtained, so long as they do obtain. (ii) Therefore, by parity of reasoning, it shouldn't matter, *vis-à-vis* an apt performance, whether the seat and shape elements of a subject's competence have been acquired by the subject through abnormal means, so long as they *have* been acquired.

But on this point, Sosa has—very recently (2017)—taken a different tack. Here are some interesting remarks from his latest book:

For simplicity, I leave aside restrictions on how you acquire the relevant elements of competence, such as the seat and the shape, restrictions that have come to the fore with the cyclist Lance Armstrong (regarding drug-induced shape) and with the baseball player Alex Rodriguez (regarding drug-derived seat). Each of these athletes enhanced his performances by

²For instance, one might have animal knowledge when one uses one's visual-perceptual conditions to identify a red wall when there is no 'funny business' with the lighting conditions, but even if there *easily could have been* funny business with the lighting—viz., a jokester waiting in the wings, but who just so happened to not tinker with the lights, bringing about a lighting condition under which you'd believe the wall was red when it was not.

³For some criticism of this distinction, see Kornblith (2004; 2009; 2012). Cf., however, Carter and McKenna (2017) for a defence of this distinction in light of Kornblith's criticisms.

enhancing his complete SSS dispositions to succeed, but these dispositions did not remain competences once drug-enhanced (Sosa 2017, 195, fn. 2.).

The above remarks indicate a kind of *acquisition restriction principle* in Sosa's thinking: if an SeShSi condition of a complete competence is *drug enhanced*, then the associated SSS disposition to succeed is not a *competence*. And given that successes that do not derive from genuine competences are not apt, it follows from such an acquisition restriction principle that no successes (epistemic or otherwise) that issue from drug-enhanced SSS dispositions to succeed are *apt*.

Of course, 'drug-enhanced' does not simply mean: 'affected' by a drug. Caffeine and vitamin supplements might temporarily sharpen one's senses before a performance; moreover, some drugs (e.g., anti-depressants) might help one to perform better by simply aiding in restoring one to normal healthy levels of functioning. In Sosa's examples, he is referring to individuals who are using medicine to enhance themselves *beyond* normal healthy levels of functioning in order to gain an advantage. Sosa is accordingly using 'enhanced' in a way that comports with the distinction in the bioethics literature between mere *therapeutic improvements* (i.e., which restore one to normal healthy levels of functioning) and *enhancements*, which aim to take healthy individuals beyond such normal levels⁴, generally by availing to the latest science and medicine⁵.

3. Why does Sosa take such a line? He doesn't explicitly say. But others have (in various ways) defended similar ideas, albeit, to different ends.

In *The Case Against Perfection*, political philosopher Michael Sandel (2012) writes:

[...] as the role of the enhancement increases, our admiration for the achievement fades. Or rather, our admiration for the achievement shifts from the player to his pharmacist [...] This suggests that our moral response to enhancement is a response to the diminished agency of the person whose achievement is enhanced. The more the athlete relies on drugs or genetic fixes, the less his performance represents his achievement (2012, 25–26).

⁴Enhancement comes by degree. Plausibly, Sosa has in mind something like 'significantly' (drug enhanced) relative to non-enhanced levels of performance, where significance will partly be fixed by domain of performance.

⁵For further discussion on this distinction, see for example, Bostrom and Savulescu (2009), Bostrom (2009) and Bostrom and Roache (2007).

In a similar vein, bioethicist John Harris (2011) has suggested that performance enhancements take away a valuable kind of freedom we have, a freedom to *fail*, by almost guaranteeing our success in a way that renders our own agency in the course of our endeavours otiose. Drawing inspiration from Milton's *Paradise Lost*, Harris writes 'Milton's insight is the crucial role of personal liberty and autonomy: that sufficiency to stand is worthless, literally morally bankrupt, without freedom to fall', and he remarks elsewhere that 'our freedom to fall is precious' (Harris 2011, 110). One point in common between Sandel's and Harris's assessments is that both see enhancements as undermining responsibility for the performance in a manner that has a deleterious effect on performance's value, even if it may successfully attain its end⁶.

4. What Sosa says about athletic performance enhancement applies, *mutatis mutandis*, for cognitive performance enhancement. Consider, for example, Modafinil, a nootropic 'smart drug' prescribed to patients suffering from narcolepsy, but which is often used off-label as a kind of eugeroic cognitive enhancement agent. Modafinil is an effective way to boost cognitive performance. Comprehensive meta-studies show that Modafinil consistently boosts (in non-sleep-deprived healthy individuals) attention, executive functions, and learning, especially in complex cognitive tasks (Battle-day and Brem 2015).

Here is a dilemma: if we embrace Sosa's acquisition restriction principle, then—absent some compelling disanalogy between athletic and cognitive performance enhancement that would justify a difference in how we normatively assess the two from within Sosa's framework—Modafinil-enhanced cognitive successes are not *knowledge*, animal or otherwise. (After all: knowledge minimally requires *apt* belief, true belief that manifests a competence; but it follows from Sosa's acquisition restriction principle that Modafinil-driven cognitive successes don't derive from competence and so *a fortiori* fall short of knowledge). This is an uncomfortable result.

But—and here is the dilemma—we can't very well avoid the above result simply by jettisoning wholesale Sosa's acquisition restriction principle. Suppose, for *reductio*, that we did, while attempting to retain the broader performance normativity framework in claims (A-D). It would follow that one's performance may qualify as apt even in circumstances in which the enhancement seems to be doing *all* the work⁷.

⁶For a critical discussion of Sandel and Harris on these issues, see Carter and Pritchard (*forthcoming*). See also Bradford (2013) for related remarks on the value of achievements in connection with the overcoming of obstacles.

⁷Consider, for example, the plight of Jeremy Renner's character Aaron Cross in the 2012 film *The Bourne Legacy*. In the film, Cross uses pills called 'chems' which enhance him both physically and mentally, permitting him to function at a superhuman capacity as a black ops agent. *Without* the pills, Cross is—and this is crucial to the plot—of below-average intelligence and capable of none of the things that make him otherwise highly effective in field operations.

My inclination is to think there should be a principled way to split the horns of this dilemma, one that is in principle compatible with Sosa's performance normativity framework. I'll turn now to developing how I think such a rationale might proceed.

5. Take as a starting point a simple idea connecting apt belief and epistemic responsibility: Your believing correctly manifests *your* competence (rather than, say, someone else's) only if *your* epistemic agency (e.g., cognitive traits, capacities that are in some way owned by you) is in some significant way responsible for your cognitive success⁸. Note that this kind of epistemic responsibility doesn't presuppose that you have direct control over your beliefs. (You don't⁹). The idea, rather, is that the kind of responsibility an apt believer has for the correctness of her belief at least requires that the mechanisms giving rise to the belief are suitably her own mechanisms, whatever this may involve.

Suppose now that a benevolent demon *causes* you to believe *p* (making it impossible for you to believe otherwise), and *p* is true. In such a case, that you believe correctly seems to have nothing to do with your own epistemic agency; your own cognitive traits play no role in accounting for why you believe truly rather than falsely. We can vary facts about your own epistemic agency anyway we like, but the outcome remains the same—you believe (correctly) that *p*.

Such a case is suggestive of something like the following initial idea: a subject is epistemically responsible for some cognitive success, *X*, only if it's not the case that no matter how we alter facts about the subject's epistemic agency, the result *X* remains the same. Put another way: the agent isn't responsible for the cognitive success if the subject *couldn't have avoided* that success. This idea comports with a more general idea: that responsibility of any kind requires alternative possibilities.

Frankfurt-style cases¹⁰, however, famously cast doubt on this general claim. And likewise, *epistemic* Frankfurt cases¹¹—viz., where a benevolent demon waits in the wings and will intervene only if the subject doesn't come to form the target belief correctly, but since the agent forms the target belief correctly, the demon accordingly doesn't intervene—cast doubt on the supposition that alternative possibilities are necessary for epistemic responsibility. There is, however, plausibly a freedom-relevant necessary condition on epistemic responsibility that doesn't trade on alternative possibilities, one that has to do with what Fischer and Ravizza (2000) call *guidance control*.

In what follows, I will (i) briefly explain guidance control as a freedom-relevant feature of moral responsibility and show how (a tweaked version of) a guidance con-

⁸This is, in effect, what Duncan Pritchard has termed the 'ability intuition' on propositional knowledge. See, for example, Pritchard (2012).

⁹See, for example, Williams (1970). Cf., McCormick (2014).

¹⁰See Frankfurt (1969).

¹¹For a recent discussion of such cases, see Kelp (2016).

trol condition is a plausible necessary condition on epistemic responsibility of the sort that apt belief plausibly demands; (ii) I'll then attempt to show why *some* kinds of pharmacological cognitive enhancements may fail such a condition while others satisfy it. The upshot, or so I'll suggest, is that there is room to manoeuvre in a principled way between the two horns of the dilemma: the virtue epistemologist who embraces a Sosa-style performance normativity model of epistemic assessment can allow that some beliefs formed via the aid of cognitive enhancing drugs are apt (and so knowledge) while others are not.

6. Fischer and Ravizza (2000) define *guidance control* as the freedom-relevant aspect of moral responsibility. The pithy statement of the view is as follows: an agent has guidance control over an action if and only if that action is caused by a reasons-responsive mechanism M , and S owns M ¹². As Fischer and Ravizza themselves note, and as is suggested by the above definition, there are two key dimensions of guidance control: *reasons-responsiveness* and *mechanism ownership*. On the Fischer-Ravizza view, a mechanism (in short: a way of doing something) is reasons-responsive provided that, *holding fixed that mechanism*, 'the agent would presumably choose and act differently in a range of scenarios in which he is presented with good reasons to do so' (Fischer 2012, 187). The mechanism that is normal practical reasoning functioning in normal conditions is reasons-responsive; even if this mechanism issues act A in the actual world in light of the presence of the actual reasons on which A is based, if we hold fixed this mechanism and vary the presence of reasons so that there is, say, overwhelming reasons for B rather than A , normal practical reasoning will go for B on the basis of these other reasons. The mechanism of *coerced* practical reasoning is not likewise reasons-responsive (after all, even if we adjust the reasons, that mechanism will continue, holding fixed the coercion as a feature of the mechanism, to issue the same action).

Given that one could be manipulated or brainwashed so as to acquire a reasons-responsive mechanism (in the sense just specified) but would not intuitively be responsible in such a circumstance for the deliverances of such a mechanism, Fischer and Ravizza specify further that the mechanism that issues in the relevant behaviour must (in an appropriate sense) *be the agent's own mechanism*. They articulate this mechanism ownership condition as follows: S owns a process/mechanism of type M iff S reasonably takes herself to be the agential source of the outcomes of M and S takes herself to be a fair target of reactive attitudes regarding the outcomes of M ¹³.

¹²See Kruse (2017) for a helpful and concise presentation of Fischer and Ravizza's key definitions. Cf., Mele (2000, 2006) and Todd and Tognazzini (2008) for some critiques of guidance control.

¹³See Fischer (2012, 190) and Fischer and Ravizza (2000, 210–13). I am taking this formalisation from Kruse (2017, 2814).

We may initially define *epistemic guidance control*¹⁴ as an analogous condition on *epistemic* responsibility, one that likewise admits of dimensions that will be epistemic variants on reasons-responsiveness and mechanism ownership. For our purposes, an epistemic variant on reasons-responsiveness (which Fischer and Ravizza are of course thinking about in the context of moral, rather than epistemic, responsibility) will involve—on a first pass—the following core idea: a *belief-forming mechanism* is (epistemically) reasons-responsive provided that, *holding fixed that mechanism*, ‘the agent would presumably form beliefs differently in a range of scenarios in which she is presented with good reasons to do so’. On such a view, the mechanism of normal induction is epistemically reasons-responsive—viz., the process responds differentially in nearby worlds where we vary the inputs—while the mechanism of hypnotized believing is not. These are both intuitive results¹⁵.

7. I submit that belief-forming processes boosted by pharmacological cognitive enhancements are not going to have any obvious problems satisfying *epistemic reasons-responsiveness*, at least, not in ordinary circumstances. The idea, in sum, is this: Let M be a cognitive mechanism and let $E(M)$ be M operating at an enhanced capacity. If M is suitably reasons-responsive, then $E(M)$ will be too, provided $E(M)$ is really an *enhancement* of M , as opposed to a diminution of M . Suppose that M *ex hypothesi* involves attention and executive functions of the sort enhanced via Modafinil in complex cognitive tasks. If M is a suitably reasons-responsive, then when we hold fixed the mechanism (i.e., involving attention and executive functions, etc.) and vary the inputs, M responds differentially in light of the varied inputs. But then $E(M)$, an enhanced variation of this very mechanism, will if anything respond differentially *even*

¹⁴McHugh (2013) is the first I’m aware of to use the term ‘epistemic guidance control’. Though, McHugh is using the term to capture, in the man, the reasons-responsiveness element of (doxastic) guidance control. As will be come clear, the interest here will primarily be to do with the mechanism ownership component. We thus are using the term in different ways.

¹⁵It might be thought that the above-sketched way of modelling the ‘epistemic reasons-responsiveness’ component of what would be epistemic guidance control allies with a distinctively *epistemic internalist* gloss of epistemic responsibility but not so obviously with an externalist gloss of this notion. And Sosa’s epistemology is externalist, a form of virtue reliabilism. Thus, as the line of thought would go, reasons-responsiveness as a component on guidance control is orthogonal to the kind of guidance control that would be germane to epistemic responsibility on an externalist programme like Sosa’s. This is however not a deep worry. For one thing, as Sylvan and Sosa (2014) have pointed out, reasons possession itself may be explicable with reference to the more fundamental notion of competence possession. For another—and more importantly—we may easily retain the core idea of reasons-responsiveness (as a component of responsibility-relevant guidance control) by simply maintaining this: a mechanism (i.e., belief forming process) is reasons-responsive in the way that matters for guidance control provided the process-type is suitably modally robust—viz., that the process responds differentially in nearby worlds where we vary the inputs.

better in light of the varied inputs. And indeed, that's what the empirical evidence would suggest¹⁶.

The situation then seems to be as follows: if cognitively enhancing our ways of forming beliefs has as a consequence that the epistemic responsibility we have for our cognitive successes reached through such enhanced processes is undermined, then this isn't going to be for the reason that such enhancements undermine the guidance control condition on epistemic responsibility by way of the reasons-responsiveness dimension to guidance control. It must be, if anything, that such enhancements stand in tension with *mechanism ownership*.

8. According to Fischer and Ravizza, one *owns* a mechanism just in case one reasonably takes herself to be the agential source of the outcomes, Ω , of that mechanism, and reasonably takes herself to be a fair target of the reactive attitudes (e.g., approval, praise, disapprobation, blame, etc) regarding Ω . We can—following Andrea Kruse (2017, 2814)—define *epistemic mechanism ownership* as the following restricted idea: *S* owns an epistemic process/mechanism of type *M* iff *S* reasonably takes herself to be the agential source of the outcomes of *M* and *S* takes herself to be a fair target of reactive attitudes regarding the doxastic outcomes of *M* (i.e., the beliefs that *M* issues). Let's set aside what 'reasonably' involves for the moment—though we'll return to this.

While beliefs formed via smart-drugs will satisfy the reasons-responsiveness condition on epistemic guidance control, there's a worry that they'll *never* satisfy the epistemic mechanism ownership condition, and that this may be so no matter how we unpack 'reasonably'.

To appreciate why, consider the following case in the moral responsibility literature, due to Neal Judisch (2005), which Judisch intends as an objection to the Fischer-Ravizza account of moral guidance control. What's interesting in particular about the case is the kind of move Fischer thinks must be made in order to deal with it, and what this suggests for epistemic mechanism ownership and enhancement.

BYPASS. [S]uppose that one night, while Chum is soundly asleep, he spontaneously develops a debilitating brain lesion. The lesion is situated in his neural network in such a way that his capacity for practical reasoning is severely impaired—the relevant mechanism no longer even approximates the standards of moderate reasons-responsiveness. Imagine now that a benevolent neuroscientist, Dr. White, is somehow made aware of Chum's plight. Unfortunately, there is no way he can remove the lesion without causing irreparable damage to Chum's brain, but White has

¹⁶See Battleday and Brem (2015).

a few handy electronic devices that enable him (literally) to get around that problem. The first, placed just “upstream” of the lesion, takes as inputs the messages sent through the neural pathways headed right for the spot where the lesion is located, and it transmits the incoming data via radio signals to the other device located just “downstream” of the lesion, which device, in turn, relays the appropriate impulses to the neural pathways just downstream of it. The result is that the lesion is both successfully isolated and bypassed, its potentially deleterious effects completely cut off from the rest of Chum’s brain; indeed, Chum’s post-surgery cognitive architecture is functionally equivalent to his pre-surgery brain... When Chum awakens, he is of course completely unaware of the evening’s events; as far as he is concerned, it is business as usual (2005, 115–30).¹⁷

Judisch’s diagnosis is as follows: (i) Chum is morally responsible for his subsequent behaviour; (ii) but, contra the Fischer-Ravizza view according to which moral responsibility requires guidance control (as they specify it), Chum cannot qualify as morally responsible. Judisch’s rationale for (ii) is that you can’t *take responsibility* in the sense required for mechanism ownership of a mechanism that is manipulated in a clandestine fashion (after all, it seems that you are utterly mistaken in such a circumstance about what the mechanism actually is).

I suggested that Fischer’s reply is interesting and bears relevance to how to think about enhancement cases *vis-à-vis* mechanism ownership. Fischer’s reply, in short, involves an appeal to commonsense functionalism. He writes:

[...] if neuroscientists secretly installed a physically different part of my brain that functioned equivalently to the biological part, this would surely not create a different kind of mechanism. As far as I’m concerned, this would be like replacing one’s carburetor with a functionally identical carburetor but made of slightly more durable material; it would still be the same kind of engine (for most conceivable purposes). Merely changing the physical realization of the processing that goes on (without changing the inputs or the processing of those inputs) does not, it seems to change the mechanism-kind (2012, 202).

A corollary of Fischer’s diagnosis is that if the neuroscientists installed something that was *more than the functional analogue of a sturdier carburetor*—viz., one that didn’t ‘function equivalently’ but functioned *better*, then the situation would be different. And he says as much explicitly in commenting on a twist on Judisch’s case:

¹⁷Cited also in Fischer (2012, 201).

[I]f the way Chum tends to process inputs is fundamentally altered, this changes Chum's mechanism. For example, if Chum is (say) highly risk-averse prior to the manipulation and very adventuresome after, or if he is egoistic before and altruistic after, the intervention has changed Chum's mechanism—his signature way of weighing reasons (2012, 202).

The foregoing reasoning by Fischer suggests the following. Call the mechanism that is [*reasoning* + Modafinil] ' $M(R)$ '. If Modafinil alters a thinker's mechanism in a manner broadly analogous to the moral case described above—such that $M(R)$ stands to normal reasoning, R , as the alteration to Chum's mechanism in the passage above stands to Chum's normal mechanism—then the use of Modafinil has fundamentally altered the mechanism. But then, if the mechanism is fundamentally altered, there's a problem: it's not clear how the subject can *reasonably* take responsibility in the way required for mechanism ownership. After all, what would such evidence *be* in virtue of which such a taking of responsibility is reasonable? It's at best unclear¹⁸.

9. At this point, though, we may anticipate an optimistic rejoinder: in the typical kinds of cases of cognitive enhancement under consideration, you are freely *choosing* to take Modafinil! This feature makes such cases very different from the BYPASS case where you were manipulated against your will. Consideration of this difference may lead one to reason as follows: your responsibility for your choice to take Modafinil 'carries over' to the known consequence of this choice (i.e., your cognitive behaviour while on the drug) in such a way that you could be responsible for the cognitive success while on the drug *even if* you fail the mechanism ownership condition on epistemic guidance control at the time at which you are forming Modafinil-fueled cognitive outputs.

Such an appeal, is, effect, an appeal to what Fischer and Ravizza call a *tracing condition*. And indeed, Fischer and Ravizza help themselves to just such a condition in order to handle cases where it seems that one is responsible at a time despite not having a reasons-responsive mechanism at that time—as in drunk driving. According to Fischer and Ravizza (2000), 'When one acts from a reasons-responsive mechanism at time T_1 , and one can reasonably be expected to know that so acting will (or may) lead to acting from an unresponsive mechanism at some later time T_2 , one can be held responsible for so acting at T_2 ' (2000, 50).

¹⁸Consider that such evidence would at least have to include evidence that one is the agential source of the outcomes of the enhanced mechanism, *as well as* additional evidence that S takes herself to be a fair target of reactive attitudes regarding the doxastic outcomes of M . If one knows the Modafinil has been efficacious, it's not clear initially how either of these would be satisfied in normal cases of Modafinil use. We will, however, return to this point in §10.

Question: Can we say then that a cognitively enhanced individual may satisfy a plausible guidance control requirement on epistemic responsibility simply because—even though the agent seems to fail the mechanism ownership condition on guidance control at a time T_2 —the agent nonetheless acted from an owned reasons-responsive mechanism at T_1 (i.e., through an owned mechanism used to take the pill in the first place)?

It's not clear that we can. I want to suggest now that a tracing condition, though it seems plausible in the case of *moral responsibility* as Fischer and Ravizza are advertent to it, looks quite a bit less plausible when we shift from moral to *epistemic* responsibility. Furthermore, a tracing condition—if plausible at all *vis-à-vis* epistemic responsibility—seems comparably more plausible in cases of epistemic blame as opposed to epistemic praise (e.g., of the sort befitting an epistemic achievement).

To see this point, compare the plausibility of the tracing condition in a drunk driving case (a case with a negative moral outcome—where the free choice at a previous time to knowingly put oneself in a position whereby one drove dangerously seems intuitively to warrant moral blame) with the following case of epistemic responsibility for a positive epistemic outcome.

Imagine here an extreme case: suppose there is a futuristic pill *Euler-Pro* which does the following: firstly, the pill incepts in one's memory detailed information about topology, graph theory and analytic number theory; the pill further dramatically enhances one's raw mathematical skill. Aaron, an ordinary maths student with average grades, take the *Euler-Pro* pill and, within five minutes, works out the connection between the Riemann zeta function and the prime numbers. He proceeds to write up a proof for what is effectively the Euler product formula for the Riemann zeta function (and his teacher gives him an A).

If an 'epistemic tracing condition' were a plausible supplement to epistemic guidance control as a condition on epistemic responsibility, we should expect Aaron to be responsible, epistemically speaking, for his proof (regardless of what is going on while he is proving the Euler product formula for the Riemann zeta function) given that he had suitable guidance control over his previous action of taking a pill the taking of which put him in the state whereby this formula became provable for him.

But this doesn't seem right at all. To see why, consider a tracing-relevant variation on the famous case of Norman the Clairvoyant (Bonjour 1980) in the classic reliabilist literature.

NORMAN-PILL: Norman wants to know where the President is at all times, but the Secret Service frequently hide the President's whereabouts for security reasons, constantly frustrating Norman's desire to know where the President is. Fortunately for Norman, an angel decides to help him out. The angel gives him a pill with the following explanation. Taking

the pill will cause two things: first, it will cause Norman to form a true belief about where the President is any time Norman thinks about this. Secondly, the pill will cause Norman to forget his encounter with the angel, and thus, will cause him to lack any appreciation for why he is believing what he does. Norman freely chooses to take the pill. The next day, Norman forms the (true) belief ‘The President is in Boston’ and has no idea why he thinks this.

In *NORMAN-PILL*, Norman’s belief is correct: the President really is in Boston. But, upon recognising that he has no idea why he believes this, Norman acquires a defeater for his belief—what is called in the defeasibility literature a ‘no-reason defeater’¹⁹. He’s no longer justified in continuing to believe the President is in Boston *even though* this belief issues from an infallible process. Continuing to believe the President is in Boston despite the presence of such a defeater is epistemically irresponsible.

Such epistemic irresponsibility isn’t ‘cancelled’ by pointing out the further fact that a tracing condition is met in *NORMAN-PILL*. After all, Norman *did* have full command of his faculties and made a free and informed decision to take the pill at an earlier time, a pill that he knew would bring it about both that he would know where the President was any time he thought about it *and* that he would forget at the later time the explanation for why he believed this.

And what goes for Norman in *NORMAN-PILL* goes for Aaron who takes the *Euler-Pro* pill. What these cases indicate is an interesting way in which moral and epistemic responsibility come apart: regardless of whether we can supplement an account of guidance control with a tracing condition to handle problem cases with respect to moral responsibility (e.g., such as drunk driving cases), the same doesn’t seem to be the case with epistemic responsibility. In particular, looking to past free, fully informed decisions isn’t always going to be helpful in assessing epistemic responsibility at a later time. And what *this* means is that attempting to explain how we may be suitably epistemically responsible for the correctness of beliefs formed via enhancement, but which seem to fail a mechanism ownership condition, simply by pointing to how such non-owned mechanisms are known-consequences of owned reasons-responsive mechanisms exercised at a previous time is not going to be compelling. Accounting for how we are suitably responsible for the correctness of such beliefs in the face of the mechanism ownership problem requires, thus, looking elsewhere.

¹⁹A no-reason defeater is a type of psychological or mental state defeater (see Pollock 1986) that is a reason for thinking it’s no longer reasonable to believe a proposition, given that one lacks any reason for believing it, and yet, it’s the sort of belief that would be reasonable to hold only if one did have some reason for it (Bergmann 1997, 102–103). See also Sudduth (2014).

10. Let's take stock. We were initially faced with a dilemma: if we embrace Sosa's acquisition restriction principle, then it looks as though *Modafinil*-enhanced cognitive successes will inevitably fall short of *knowledge*, animal or otherwise—a result that seemed problematic (for surely we can know at least *some* of the things we come to believe through cognitive enhancing drugs?) But, if we tried to avoid this result by jettisoning wholesale Sosa's acquisition restriction principle, it would follow that one's performance may qualify as apt, and accordingly that one's belief may qualify as animal knowledge, even in circumstances in which the enhancement seems to be doing *all* the work. The objective, then, was to seek a way to navigate through the horns of this dilemma.

Doing so, it was suggested, would seem to require some principled rationale for a difference in treatment in extreme enhancement cases as opposed to moderate ones—where *Modafinil* use would seem *prima facie* to fall in the latter category. A natural starting point was to consider how such a difference might be a difference in the level of epistemic responsibility the agent may have for the successful outcome. Here, a promising place to look was to the moral responsibility literature, which is replete with rich discussion of how disconnections between agency and action can be responsibility undermining. In particular, one popular way of thinking of such a connection is as follows: (i) moral responsibility requires guidance control; and (ii) guidance control requires the satisfaction of two conditions: a reasons-responsiveness condition and a mechanism ownership condition. Taking this line of thinking as a guide, we've seen that an analogous requirement on epistemic responsibility—viz., that epistemic responsibility for a cognitive output (e.g., such that the success of the cognitive output is suitably a product of the agent's *own* cognitive agency) requires exhibiting *epistemic guidance control*, will—in the kind of cognitive enhancement cases of interest to us—ultimately turn on the mechanism ownership condition on epistemic guidance control.

The problem, which we saw in the previous section can't be fixed by adding a tracing condition of the sort that Fischer and Ravizza endorse in the case of moral responsibility, is that it's not clear how *any* cognitive enhancements, moderate or otherwise, are going to satisfy such a mechanism ownership condition. The requirement, recall, is as follows: *S* owns an epistemic process/mechanism of type *M* iff *S* reasonably takes herself to be the agential source of the outcomes of *M* and *S* takes herself to be a fair target of reactive attitudes regarding the doxastic outcomes of *M*²⁰.

At this point, it will be fruitful to consider more carefully what 'reasonably' should be understood to involve in the mechanism ownership condition. Plausibly, Fischer and Ravizza intended such a condition in what is on the whole a subjective account of mechanism ownership, to rule-out cases where one's taking herself to be the source

²⁰This formulation is due to Kruse (2017, 2814).

of the outcomes of the mechanism and a fair target of the reactive attitudes regarding the doxastic outcomes of that mechanism are unfounded, delusional, or otherwise disconnected from reality.

But what kinds of considerations might one base one's assessments of one's self, mechanism and outputs on in order for them to qualify as reasonable? Even more specifically, what kinds of evidence would actually serve to *favour* a given self-assessment, on the part of a subject, *S*, that *S* is the source of the outcomes of a given mechanism, as opposed to a competing self-assessment according to which *S* is *not* the source of the outcome of such a mechanism?

Fortunately, there is a burgeoning field at the intersection of epistemology and the philosophy of mind and cognitive science which has attempted to answer such questions with reference to considerations to do with *cognitive integration*²¹. For example, according to S. Orestis Palermos (2014a; 2014b; 2015) the relevant kind of evidence would consist in the presence of *feedback loops*. The idea, which is grounded in a dynamical systems theory (DST)²² approach to theorising about cognitive processes, maintains that when an external element (e.g., a drug, a piece of technology, etc.) is *non-linearly* related to the agent's biological cognitive system, it can count as a constitutive part of an overall cognitive system that extends to include all the contributing parts. A key motivation for this is that, as DST maintains, such non-linear relations arising out of continuously and reciprocally interacting parts on the basis of feedback loops from one part to the other give rise to an overall non-decomposable system that consists of all of them. In short, when there are ongoing feedback loops or 'continuous reciprocal causation' (CRC) between internal and external parts (see Palermos 2014b), the result is an *extended cognitive system* that consists of all of them.

Notice that, in (extreme) enhancement cases such as NORMAN-PILL, no such feedback loops are present. After all, the pill seems to be disintegrated from the rest of Norman's cognitive psychology, such that the pill is having distinctive cognitive effects on Norman (causing him to believe certain things), even though the mechanisms of the pill are in no way *affected* by anything Norman does or thinks²³. The pill is certainly a causal antecedent of a process that is Norman's own, but if DST is any guide to what should be included in a cognitive process, the pill is not a *part* of Norman's process. Norman thus lacks at least one kind of evidence available to him (the would-be presence of feedback loops) for thinking that he would be the source of the outcomes of that mechanism (i.e., the pill-induced beliefs about the President).

²¹For some representative discussions, see Menary 2006; Pritchard 2010; 2018; Carter et al. (eds.) (2014; 2018).

²²See, for example, Chemero (2009), Froese et al. (2013) and Palermos (forthcoming).

²³Indeed, we may suppose even that if Norman were to try to believe otherwise than that the President is where the pill causes him to believe the President is, then he would be unable to do so.

The same does not obviously hold in the case of Modafinil. Whereas nothing about one's own epistemic agency is going to have much of an effect with Norman in *NORMAN-PILL*, one can potentially monitor one's use of Modafinil. This might happen in various ways. For one thing, there are known side effects of Modafinil, including (in a minority of cases) dizziness, hallucinations, unusual thoughts and behaviour²⁴, which may themselves affect the reliability of Modafinil use. One may monitor for such symptoms and refrain from trusting the deliverances of one's enhanced mechanism if one detects such symptoms. In this respect, the causal relations between Modafinil and the user are not so obviously linear one-way relations as they are the case of *NORMAN-PILL*. Thus, one might have with respect to DST some defeasible evidence for reasonably taking it that one is the source of the outcomes of that mechanism, and likewise, a fair target of the reactive attitudes.

However, it may be that additional evidence, beyond merely evidence for the presence of feedback loops (evidence that may be somewhat elusive) is needed in order for one to be reasonably take oneself to be the source of the outcomes of a cognitively enhanced mechanism.

At this point it will be helpful to consider Duncan Pritchard's (e.g., 2010; 2018) recent thinking about cognitive integration. Pritchard (2010) sets out to reconcile virtue epistemology and extended cognition, where—according to the latter—cognitive processes such as memory can supervene partly on instruments and technology located outside of the agent herself, and which are not biologically endowed features of the agent. As one might be initially inclined to read the situation, the deliverances of 'extended processes' (i.e., those partly involving work done by gadgets located outside our heads) are not going to be due to *ability* or virtue to a sufficient extent in order to qualify as knowledge by virtue epistemology's lights. Of course, the proponent of extended cognition (Clark and Chalmers 1998) wants to insist that the matter of whether something is located outside one's head is irrelevant to whether that thing is part of a genuine cognitive process; what matters is just the functional role that the thing is playing.

Pritchard's (2010) novel move to reconcile these two programs takes as a starting point the following 'cognitive integration' condition on knowledge:

(COGA_{WEAK}) If *S* knows that *p*, then *S*'s true belief that *p* is the product of a reliable belief-forming process which is appropriately integrated within *S*'s cognitive character such that her cognitive success is to a significant degree creditable to her cognitive agency (2010, 136-7).

Pritchard's formulation is importantly neutral with respect to whether or not the process that gives rise to the belief must be entirely comprised of one's biologically

²⁴https://www.emedicinehealth.com/drug-modafinil/article_em.htm

endowed faculties, or whether it may in some cases incorporate features that have been ‘added on’ to the original process. What matters is that the overall (reliable) process responsible for the true belief is suitably integrated.

A brief consideration of this view suggests how it may fruitfully be applied to the vexed case of smart drugs, and to the present conundrum. To this end, let’s run a quick case comparison. One well-known case in the reliabilist literature is Alvin Plantinga’s (1993) case of someone (call him ‘Al’) who develops a brain lesion which (unbeknownst to Al) causes him produce reliable beliefs on a given subject (say, for instance, mathematics). What is the epistemic status of the ensuing reliable beliefs?

The prevailing assessment is that they are not known, despite being reliably produced, and Pritchard’s explanation is that such beliefs fail COGA_{WEAK}. But crucially, such beliefs don’t fail COGA_{WEAK} *in principle*, simply because the tumour (an addition to the original cognitive process) was what was responsible. It’s because the tumour in the initial case isn’t suitably integrated. A corollary of this thinking is that if the tumour *were* suitably integrated, there would be no barrier to gaining knowledge through a process that relies on it. And this, Pritchard supposes, might be something that could happen over time. The following remarks Pritchard offers are illuminating on this score:

For now suppose that Alvin becomes aware that there exists brain lesions of this sort, and gains additional good grounds for supposing that he possesses just such a brain lesion, such that he now knows that he has one of these brain lesions. Perhaps, for example, he comes across an article about these brain lesions in a reliable newspaper and researches the matter in reliable medical journals and on this basis comes to know that he is the victim of the brain lesion in question. Intuitively Alvin’s beliefs in the target mathematical propositions now qualify as knowledge. But notice that it is also true that Alvin has in this way integrated this belief-forming process within his cognitive character to a sufficient degree that his cognitive success is now primarily creditable to his cognitive agency, rather than being creditable to something external to his cognitive agency (albeit a factor which was under his skin, and hence in this sense internal) (2010, 138).

We might then say something similar for beliefs produced through the assistance of smart drugs. Suppose that, rather than developing a brain lesion which aids him in his mathematical beliefs, ‘Al-Drug’ begins using Modafinil or some other cognitive performance-enhancing drug, but with *no conception* whatsoever of what the drug does or how it effects his formation of mathematical beliefs. With no such conception (and no appreciation of how the drug may go wrong, when it does—viz., in the case of side

effects), ‘Al-Drug’ is in a relevant respect in a similar position as Al is in in Plantinga’s original case *prior* to any appreciation of the brain lesion or how it is affecting him.

Things seem very different, however, if we suppose that ‘Al-Drug’ (like Al from Pritchard’s variation on the case) comes over time, and through calibration and feedback, to appreciate the source of the reliability of his beliefs, along with a sensitivity to when things go wrong (e.g., tell-tale signs of any epistemically problematic side effects), an appreciation that may be developed and refined over time and through further feedback. In such a circumstance, there is, on Pritchard’s rationale, no principled reason to treat ‘Al-Drug’ any differently from ‘Al-lesion’; both *may* potentially come to gain knowledge through the deliverances of their cognitive processes, respectively.

A pleasing result of the foregoing is that the very kinds of evidence one would plausibly require in order to cognitively integrate a drug into one’s belief-forming process in such a way as to satisfy Pritchard’s COGA_{WEAK} condition would at the same time be evidence with respect to which one may (by taking a view of herself on the basis of such evidence) plausibly satisfy the epistemic mechanism ownership condition for the guidance-control requirement on epistemic responsibility—viz., evidence on the basis of which one might *reasonably* take oneself to be the agential source of the outcomes of the relevant belief forming mechanism, and a fair target of the reactive attitudes.

And this is a good news in so far as we want to split the horns of Sosa’s dilemma. Sosa’s restriction principle indicated that no cognitively enhanced performances aspire to aptness, a consequence of which is that no beliefs formed via cognitive enhancement aspire to knowledge. I hope to have shown that there is scope to resist this strong conclusion *without* tolerating cognitive enhancement wholesale in our epistemic assessments. Apt beliefs (i.e., knowledge) must be beliefs we are suitably responsible for, and this requires (among other things) that we *own* the relevant mechanism giving rise to our beliefs. We can’t own them without being in a position to reasonably take ourselves to be the agential source of the outcomes of the such mechanisms. Fortunately, the kind of evidence plausibly needed for such self-assessments to be reasonable can be found by looking to recent work at the intersection of epistemology and the philosophy of mind and cognitive science—viz, evidence of suitable cognitive integration²⁵.

REFERENCES

Battleday, Ruairidh M., and A.K. Brem. 2015. ‘Modafinil for Cognitive Neuroen-

²⁵Thanks to Emma C. Gordon, Michel Croce, an anonymous referee and to an audience at the University of Cologne for helpful feedback.

- hancement in Healthy Non-Sleep-Deprived Subjects: A Systematic Review'. *European Neuropsychopharmacology* 25 (11):1865–1881.
- Bergmann, Michael. 1997. 'Internalism, Externalism, and Epistemic Defeat'. (PhD Dissertation: University of Notre Dame).
- BonJour, Laurence. 1980. 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy* 5, 53-73.
- Bostrom, Nick. 2009. 'Cognitive Enhancement: Methods, Ethics, Regulatory Challenges'. *Science and Engineering Ethics* 15 (3):311–41.
- Bostrom, Nick, and Rebecca Roache. 2007. 'Human Enhancement: Ethical Issues in Human Enhancement'. In *New Waves in Applied Ethics*, edited by Jesper Ryberg, Thomas S. Petersen, and Clark Wolf. Palgrave Macmillan.
- Bostrom, Nick, and Julian Savulescu. 2009. 'Human Enhancement Ethics: The State of the Debate'. In *Human Enhancement*, edited by Nick Bostrom and Julian Savulescu, 1–22. Oxford University Press.
- Bradford, Gwen. 2013. 'The Value of Achievements'. *Pacific Philosophical Quarterly* 94 (2):204–224.
- Carter, J. Adam. 2017. 'Intellectual Autonomy, Epistemic Dependence and Cognitive Enhancement'. *Synthese*, 1–25. <https://doi.org/10.1007/s11229-017-1549-y>
- Carter, J. Adam, and Robin McKenna. 'Sosa versus Kornblith on Grades of Knowledge'. (Unpublished manuscript).
- Carter, J. Adam, and Duncan Pritchard. Forthcoming. 'The Epistemology of Cognitive Enhancement'. *Journal of Medicine and Philosophy*, 1–25.
- Carter, J. Adam, Andy Clark, Jesper Kallestrup, Orestis Palermos and Duncan Pritchard (eds.) 2018. *Extended Epistemology*, Oxford: OUP.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT press.
- Fischer, John Martin. 2012. *Deep Control: Essays on Free Will and Value*. Oxford University Press, USA.
- Fischer, John Martin, and Mark Ravizza. 2000. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge university press.
- Frankfurt, Harry G. 1969. 'Alternate Possibilities and Moral Responsibility'. *The Journal of Philosophy* 66 (23):829–839.

- Froese, T., Gershenson, C., and Rosenblueth, D., A. (2013). 'The Dynamically Extended Mind'. <http://arxiv.org/abs/1305.1958>
- Harris, John. 2011. 'Moral Enhancement and Freedom'. *Bioethics* 25 (2):102–11.
- Judisch, Neal. 2005. 'Responsibility, Manipulation and Ownership: Reflections on the Fischer/Ravizza Program'. *Philosophical Explorations* 8 (2):115–130.
- Kelp, Christoph. 2016. 'Epistemic Frankfurt Cases Revisited'. *American Philosophical Quarterly* 53 (1):27–37.
- Kornblith, Hilary. 2004. 'Sosa on Human and Animal Knowledge'. In *Ernest Sosa and His Critics*, edited by John Greco, 126–34. Wiley-Blackwell.
- . 2009. 'Sosa in Perspective'. *Philosophical Studies* 144 (1): 127–36.
- . 2012. *On Reflection*. Oxford University Press.
- Kruse, Andrea. 2017. 'Why Doxastic Responsibility Is Not Based on Direct Doxastic Control'. *Synthese* 194 (8):2811–2842.
- McCormick, Miriam Schleifer. 2014. *Believing Against the Evidence: Agency and the Ethics of Belief*. Routledge.
- McHugh, Conor. 2013. 'Epistemic Responsibility and Doxastic Agency'. *Philosophical Issues* 23 (1):132–157.
- Mele, Alfred R. 2000. *Reactive Attitudes, Reactivity, and Omissions*. JSTOR.
- . 2006. 'Fischer and Ravizza on Moral Responsibility'. *The Journal of Ethics* 10 (3):283–294.
- Menary, Richard. 2006. *Cognitive Integration: Mind and Cognition Unbound*. Basingstoke: Palgrave MacMillan.
- Palermos, S. Orestis. 2014a. 'Knowledge and Cognitive Integration', *Synthese* 191 (8): 1931–1951.
- . 2014b. 'Loops, Constitution, and Cognitive Extension', *Cognitive Systems Research* 27: 25–41.
- . (2015). 'Active Externalism, Virtue Reliabilism and Scientific Knowledge', *Synthese*, 192(9), 2955–2986.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*, Oxford University Press, Oxford.

- Pollock, John. 1986. *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.
- Pritchard, D. 2012. 'Anti-Luck Virtue Epistemology'. *Journal of Philosophy* 109:247–79.
- . (2010). 'Cognitive Ability and the Extended Cognition Thesis', *Synthese* 175: 133–151.
- . (2018). 'Extended Knowledge'. In *Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos and Duncan Pritchard. Oxford: OUP.
- Sandel, Michael J. 2012. 'The Case against Perfection: What's Wrong with Designer Children, Bionic Athletes, and Genetic Engineering'. In *Arguing About Bioethics*, edited by Stephen Holland, 93. Routledge.
- Sosa, Ernest. 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford University Press.
- . 2010a. *Knowing Full Well*. Princeton: Princeton University Press.
- . 2010b. 'How Competence Matters in Epistemology'. *Philosophical Perspectives* 24 (1):465–75. <https://doi.org/10.1111/j.1520-8583.2010.00200.x>.
- . 2015. *Judgment and Agency*. Oxford: Oxford University Press.
- . 2017. *Epistemology*. Princeton, N.J.: Princeton University Press.
- Sylvan, Kurt, and Ernest Sosa. 2014. 'The Place of Reasons in Epistemology'. In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star. Oxford University Press. <https://eprints.soton.ac.uk/362902/>.
- Sudduth, Matthew. 2013. 'Defeaters in Epistemology'. *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/ep-defea/>
- Todd, Patrick, and Neal A. Tognazzini. 2008. 'A Problem for Guidance Control'. *The Philosophical Quarterly* 58 (233):685–692.
- Williams, Bernard. 1970. 'Deciding to Believe'. In *Language, Belief, and Metaphysics*, edited by Howard E. Kiefer and Milton K. Munitz, 95–111. Albany, NY: SUNY Press.