

Forthcoming in *Synthese*

Intellectual Autonomy, Epistemic Dependence and Cognitive Enhancement

J. Adam Carter

University of Glasgow

Abstract: Intellectual autonomy has long been identified as an epistemic virtue, one that has been championed influentially by (among others) Kant, Hume and Emerson. Manifesting intellectual autonomy, at least, in a virtuous way, does not require that we form our beliefs in cognitive isolation. Rather, as Roberts and Wood (2007, 259–60) note, intellectually virtuous autonomy involves reliance and outsourcing (e.g., on other individuals, technology, medicine, etc.) to an appropriate extent, while at the same time maintaining intellectual self-direction. In this essay, I want to investigate the ramifications for intellectual autonomy of a particular kind of epistemic dependence: *cognitive enhancement*. Cognitive enhancements (as opposed to therapeutic cognitive improvements) involve the use of technology and medicine to improve cognitive capacities in healthy individuals, through mechanisms ranging from smart drugs to brain-computer interfaces. With reference to case studies in bioethics, as well as the philosophy of mind and cognitive science, it is shown that epistemic dependence, in this extreme form, poses a *prima facie* threat to the retention of intellectual autonomy, specifically, by threatening to undermine our intellectual self-direction. My aim will be to show why certain kinds of cognitive enhancements are subject to this objection from self-direction, while others are not. Once this is established, we'll see that even some extreme kinds of cognitive enhancement might be not merely compatible with, but constitutive of, virtuous intellectual autonomy.

Intellectual Autonomy: Kant, Emerson and Hume

Intellectual autonomy—roughly, a disposition to think (in some to-be-specified sense) *independently*—has received various influential endorsements in the Western intellectual tradition. Immanuel Kant (1784) regards independence of thought as at the very heart of the enlightenment's motto:

Enlightenment is man's leaving his self-caused immaturity. Immaturity is the incapacity to use one's intelligence without the guidance of another. Such immaturity is self-caused if it is not caused by lack of intelligence, but by lack of determination and courage to use one's intelligence without

being guided by another. *Sapere Aude*¹! *Have the courage to use your own intelligence!* is therefore the motto of the enlightenment (1784, §1).

Kant here is effectively praising intellectual *self-determination*, which can be contrasted with what he's calling 'immaturity'², where the latter involves a kind of intellectual cowardice that manifests when one's thinking is wilfully allowed to be guided by something external to oneself.

Like Kant, the American essayist and Transcendentalist Ralph Waldo Emerson, in his famous essay *Self-Reliance*, frames as contrast points what is easy—accepting intellectual guidance—and what is praiseworthy—guiding one's own intellectual life. As Emerson (1841, 55) puts it:

It is easy in the world to live after the world's opinion. It is easy in solitude to live after our own. But the great man is he who in the midst of the crowd keeps with perfect sweetness the independence of solitude.

While both Kant and Emerson extol in different ways the virtues of self-trust, Emerson's aversion to intellectual conformity of any kind is perhaps even stronger than Kant's. Whereas Kant's primary objective in 'What is Enlightenment' is a celebration of the value of self-determination in our intellectual lives, Emerson to a greater extent focuses on the disvalue of (even weak forms of) intellectual servility³. This is apparent in Emerson's overarching argument for self-trust in *Self-Reliance*, where the opinions of others, and their potential influence, are treated as a threat to undermining the value of the kind of self-worth that is grounded in one's individuality⁴.

A more moderate though highly influential defence of self-trust, albeit of a more fine-grained sort, can be found in David Hume's thinking about both induction and testimony. Hume (1772) distinguished, as epistemologically problematic, evidence for induction conceived of as evidence concerning what goes 'beyond the present testimony of the senses, or the records of our memory'⁵. And it's precisely the testimony

¹The translation from Latin is 'Dare to be wise' or alternatively 'Dare to know', a term used originally by the Roman poet Horace in the *Epistles*. In the context of Kant's essay, the phrase is often understood as the imperative: 'Have the courage to use your own reason' (see, e.g., Gardner (1999, 2)).

²Kant's use of the German *Unmündigkeit* has also been translated as 'nonage', or, the condition of not being of age.

³Emerson's disdain for intellectual conformity was already apparent in his 1837 speech 'On The American Scholar' in which Emerson had written that man, '[...]In the degenerate state, when the victim of society, [...] tends to become a mere thinker, or, still worse, the parrot of other men's thinking' (1837, §1).

⁴A more contemporary expression of this point can be found in Charles Taylor's (1991) defence of the value of authenticity.

⁵Hume (1772, sec. 4, pt. 1).

of one's own senses and of memory which are always necessarily required, for Hume, in order to ever justifiably believe the word of someone else⁶. In this respect, Hume gives a epistemic privilege to what is apparent to the self (through perception and memory) which is not afforded to any other kind of evidence, including the kind we acquire by relying on others.

Although Kant, Emerson and Hume are all clear cases where epistemic independence and self-direction are, in different ways, advanced as worthy aims, notice that there is hardly any unified expression here of the disvalue of *epistemic dependence*, which is something any viable account of intellectual autonomy, conceived of as a virtue, should be sensitive to. Whatever is praiseworthy about intellectual autonomy should be sensitive to both what is valuable about self-direction and what is disvaluable about epistemic dependence; an account that misfires in either direction misses the mark.

We can imagine an account of intellectual autonomy which zealously incorporates Hume's epistemological demands with the most radical construal of Kant and Emerson. According to such an account, the virtuously autonomous agent (i) should never uncritically trust others (Hume), (ii) should never allow her intelligence or reason to be guided by another (Kant) and—perhaps most radically of all—(iii) should actively see to it that she is not shaped by the opinions of others (Emerson). While such a person would be intellectually autonomous in the sense of maximally independent and self-directive, she would hardly be virtuously so⁷, and this is because such an individual disvalues epistemic dependence to her own intellectual detriment.

On most any account of epistemic value, the acquisition of true beliefs, knowledge and understanding are epistemic goods⁸, and aspects of character are typically explained as intellectually good to have (as opposed to say morally good to have) in virtue of their connection to such goods⁹. The kind of cognitive isolation that results from disproportionately disvaluing epistemic dependence (as in the extreme form of autonomy sketched above) cuts one off from such cognitive goods.

⁶This is a component of Hume's global reductionist view about testimony. The classic counterreply, advanced most notably by Thomas Reid (1764, 197), is that individuals 'would be unable to find reasons for believing the thousandth part of what is told them' and thus that a non-sceptical approach to testimonial knowledge acquisition should take testimony as itself a basic source. For a contemporary expression of this rejoinder to testimonial reductionism, see Coady (1992, 82).

⁷To draw an analogy to open-mindedness, consider that one is not virtuously open-minded when one's mind is so open that one lacks any intellectual convictions whatsoever.

⁸For some representative work, see for example Haddock, Millar, and Pritchard (2009).

⁹According to virtue responsibilists (e.g., Battaly 2015; Montmarquet 1993; Code 1987), the relevant connection is unpacked in terms of motivation toward epistemic goods. Virtue reliabilists (e.g., Sosa 2009; 2010; 2015; Greco 2010; 2012) by contrast regard traits as virtues provided their manifestation reliably generates epistemic goods such as true belief and the avoidance of error. Cf., Baehr (2011) for an alternative 'personal worth' construal of this connection.

The foregoing considerations can be captured pithily in slogan form: impressive intellectual self-direction and independence of thought without a suitable knowledge base—from others, technology, the internet, etc.—is effectively empty, whereas knowledge acquired in the absence of a suitable capacity for autonomous self-direction is blind¹⁰.

I want to suggest that it's possible to articulate a more reasonable way to incorporate some of what the aforementioned thinkers take to be valuable about independence of thought without committing to an implausibly restrictive stance towards epistemic dependence. Firstly, a concession to Hume is that *blind* trust of any kind of testimony—be it particular mundane items of information, or testimony aimed at guiding inquiry itself—is epistemically criticisable. Though, this concession is compatible with the epistemic appropriateness of certain kinds of trust¹¹. Secondly, a concession to Kant and Emerson: cognitive outsourcing—to what is external to one's own intelligence (Kant) and individuality (Emerson)—is epistemically criticisable in cases where such outsourcing undermines (in some relevant, non-trivial way) one's capacity for intellectual self-direction. And this concession is compatible with the thought that sometimes—perhaps even often—we should make use of available resources (other people, technology, etc.) when forming particular beliefs and also when determining what inquiries to pursue.

Putting this all together—and without committing to any detailed account of intellectual autonomy—it's reasonable to suppose that virtuous intellectual autonomy simply cannot mean, as Roberts and Wood (2007, 259–60) put it, 'that one never relies on the intellectual labor of another'. But nor should a virtuously autonomous agent necessarily *rarely* rely on anything other than one's own endowed faculties. Rather, the crucial idea is that the virtuously autonomous agent actually must rely on others, and outsource cognitive tasks as a means to gaining knowledge and other epistemic goods, *up until the point that doing so would be at the expense of her own capacity for self-direction*. And this makes intellectual autonomy, essentially, a virtue of self-

¹⁰In the former case, we can imagine at one limit, the individual who—from her position in cognitive (near)-isolation—is (while free from the influence of others' opinions) lacking of the essential knowledge base that is necessary to inform intellectually virtuous inquiries. At the other limit, an individual acquires (through deferential and extra-agential means) an impressive knowledge-base, but lacks the crucial capacity to assess, on the basis of this wealth of information, what further inquiries should be pursued without extra-agential assistance.

¹¹This is a point that can be embraced by reductivists as well as anti-reductivists in the epistemology of testimony.

regulation¹² in the acquisition and maintenance of our beliefs¹³.

Over-reliance on the opinion of others is perhaps—as Roberts and Wood (2007, 259) have suggested—the most straightforward threat to an individual’s capacity for intellectual self-direction, but it is hardly the only such threat. The virtuously autonomous individual must also be sensitive to other less obvious ways in which her own agency can become disconnected from the way she acquires and maintains her beliefs.

Increasingly, in order to meet our cognitive goals, we tend to ‘offload’ tasks (traditionally performed through the use of our endowed biological faculties) to technological gadgets with which we regularly and uncritically interact. Moreover, the latest science and medicine has made it possible to improve cognitive functioning along various dimensions using such methods as nootropics or ‘smart drugs’ (e.g., Adderall, Ritalin, Provigil, Oxiracetan), implants (e.g., neuroprosthetics), direct brain-computer interfaces, and (to some extent) genetic engineering¹⁴.

Both high-tech cognitive offloading and reliance on medicine to achieve cognitive goals represent increasingly ubiquitous ways in which an individual’s own intellectual agency can—at least potentially—become disconnected from the way she acquires and maintains her beliefs.

In what follows, I want to examine how the foregoing considerations about the connection between virtuous intellectual autonomy and maintaining self-direction interface with the potential threat posed by various forms of cognitive enhancement. The forms of enhancement explored have in common—and in a way that is relevant to intellectual autonomy—that an individual, in virtue of the enhancement, is such that the contribution of her own biologically endowed cognitive faculties to her cognitive projects is diminished¹⁵.

The remainder of the paper proceeds as follows. §2 sharpens the notion of cognitive enhancement and, in the course of doing so, distinguishes it from the related notion of therapeutic cognitive improvement. §3 presents three kinds of cognitive enhancement cases which appear to undermine, for slightly different reasons, the intellectual autonomy of the cognitive enhanced agent, by undermining (each case, in

¹²As Roberts and Wood (2007, 259) put it, such self-regulation will involve relying on others to the appropriate extent (e.g., by being cautious and trusting in the right circumstances), without devolving into what they call intellectual *heteronymity*—i.e., defined as the opposite of autonomy—which involves being others- rather than self-directed.

¹³That the arena in which autonomy is exercised is the acquisition and maintenance of beliefs is a point that’s been developed by Zagzebski (2013, 259).

¹⁴For some helpful overviews of the state of current cognitive enhancement technologies, see Bostrom and Savulescu (2009) and Sandberg and Bostrom (2006).

¹⁵For a detailed discussion of the relationship between cognitive enhancement and cognitive achievements (in the sense of ‘achievement’ deployed by virtue epistemologists), see Carter and Pritchard (2016).

a different way) her capacity for intellectual self-direction. §4 responds to the three example cases in a way that will clarify why some enhancements are immune to the objection from self-direction while others are not. Once this point is appreciated, it will be suggested how—in the right circumstances—availing ourselves to the latest technology and medicine is not only compatible with, but can fruitfully augment, our intellectual self-direction and autonomy.

Cognitive Enhancement

One very common reason why we rely on technology and/or medicine to assist us in our cognitive tasks is that our endowed biological cognitive faculties sometimes fail us. Take, for example, Alzheimer’s disease, which has symptoms that include short-term memory loss and confusion. Sufferers of Alzheimer’s disease increasingly rely on drugs such as acetylcholinesterase inhibitors (e.g., Donepezil) to slow the cognitive symptoms of the disease¹⁶. Specifically, drugs like Donepezil slow the breakdown of acetylcholine, which is a chemical (in low supply in Alzheimer’s patients) that helps to send messages between nerve cells¹⁷.

Contrast now the use of Donepezil to slow the progression of Alzheimer’s, from a superficially similar case, where a drug is likewise involved to improve cognition, but which (as we’ll see) might generate a different intuitive reaction. Consider the ‘smart drug’ Modafinil (i.e., Provigil), a eugeroic drug prescribed to patients suffering from narcolepsy, though which is widely used ‘off label’ not to correct any pathology or defect, but to gain some kind of cognitive advantage¹⁸. A comprehensive meta-study conducted by Battleday and Brem (2015) has shown Modafinil to be consistently efficacious in enhancing, in non-sleep-deprived healthy individuals, attention, executive functions, and learning, especially in complex cognitive tasks. A potential cost of these gains, as reported in recent studies by Mohamed (2014) and Mohamed and Lewis (2014)—comes in the area of creativity. Such studies indicate that Modafinil, despite its benefits in focus in healthy individuals, has a deleterious effect in convergent and divergent creative thinking tasks, aimed at narrowing possibilities and generating novel ideas, respectively¹⁹.

In both the case of Donepezil as well as Modafinil use, something extra-agential—viz., drugs—is a significant causal difference maker with respect to the nature and quality as well as the direction of one’s cognitive projects. Also, in both cases, the contribution of the agent’s biologically endowed cognitive faculties to her cognitive

¹⁶See, for example, Boada-Rovira et al. (2004).

¹⁷https://www.alzheimers.org.uk/site/scripts/documents_info.php?documentID=147

¹⁸See Peñaloza et al. (2013) for data on recent trends in off-label usage of Modafinil in the United States.

¹⁹Mohamed (2014, 2). Note that these effects are during use.

projects is diminished given the increased role drugs are playing. However, while it's not clear that the former case in any way represents a threat to the cognitively improved agent's intellectual autonomy (if anything, it seems the former case *facilitates* intellectual self-direction in Alzheimer's patients), the second case—viz., regarding Modafinil—is a quite a bit murkier. That is, intuitively, it seems as though allowing a drug like Modafinil to substantially shape the way one manages one's cognitive life (in matters such as focus, learning and creativity) involves a certain sacrifice of self-direction that one doesn't seem to be making in the case of Donepezil.

One might initially assume that this is because 'smart drugs' such as Modafinil generally have a more substantial causal influence on one's cognitive life (relative to not taking the drug) than do drugs like Donepezil, when the latter is taken therapeutically and the former is taken 'off label' by healthy individuals. Though this is hardly the case. A faster rate of neural degeneration would be a much more dramatic shift in one's cognitive life than an absence of above-the-mean focus and attention. On closer consideration, it seems as though some kind of *normative* consideration must be what's underlying the intuition that dependence on Modafinil poses a more credible threat to undermining intellectual autonomy than does dependence on Donepezil²⁰.

Here is an obvious (though ultimately, I'll suggest, misguided) candidate normative consideration: although Donepezil is itself something extra-agential which is causally responsible in an important sense to patients' belief retention and memory, its use is aimed at correcting a pathology or cognitive defect, and as such to bring the agent closer to normal healthy levels of cognitive functioning. Improvements to human functioning which have this kind of goal are termed *therapeutic cognitive improvements*²¹. Modafinil by contrast, when used by healthy individuals to gain a kind of cognitive advantage, constitutes a cognitive *enhancement* and as such aims go beyond mere healthy human cognitive functioning²².

If this normative difference between the two cases is the right explanation, then it would have to be premised upon a more general claim to the effect that cognitive enhancements, *as such*, pose a distinctive kind of challenge to the retention of intellec-

²⁰To sharpen this comparative intuition, we can simply run a pair of variant hypothetical cases, where the causal efficacy of these drugs in each case is dramatically increased. Suppose 'Donepezil-Extra' is a drug, approved by the FDA in the year 2040, which not only slows the breakdown of acetylcholine, but halts it completely, and then through other mechanisms reverses other symptoms of Alzheimer's. Donepezil-Extra, suppose, can have this impressive effect even in cases where cognitive degeneration is severe. 'Modafinil-Extra', by contrast not only improves focus and attention—at the expense of creativity—but does so profoundly, such that users of Modafinil-Extra have substantially different (perhaps even unrecognisably different) maintenance and acquisition tendencies than they did previously, and consequently, very different cognitive character traits.

²¹See, for example, Bostrom and Sandberg (2009, 312).

²²There may potentially be some borderline cases—viz., where the limits of normal cognitive functioning are unclear.

tual autonomy that is not posed by therapeutic cognitive improvements. Even if this more general claim were true, it would just raise yet a further, more difficult question: *why* should the fact that enhancements raise cognitive functioning beyond healthy levels be relevant at all to whether autonomy is at risk of being compromised?

Ultimately, what I want to suggest is that the cognitive norm that is, on closer inspection, most fundamental in explaining why some cases where cognition is partly driven by extra-agential factors really do compromise intellectual autonomy, while others don't, is framed not in terms of enhancement, but rather, in terms of (a lack of) *cognitive integration*, in a sense that will be articulated in more detail in §4.

In order to appreciate why it's not enhancement, *as such*, that's the problem from the perspective of retaining one's intellectual autonomy, it will be helpful to consider three example enhancement cases, which draw from work by Michael Lynch (2016) on neuromedia, Felicitas Kraemer (2011) on pharmacological enhancement and authenticity and Google design ethicists on framing effects and the illusion of choice, respectively. Each case features a strand of cognitive enhancement with a different proximate cause for why the agent's capacity for intellectual self-direction is undermined. What I hope to show is that these three proximate causes all have an underlying or distal cause which can account for why intellectual autonomy is undermined in these enhancement cases. We'll see, further, that such a cause needn't be present in all cases of cognitive enhancement, which is why not all forms of cognitive enhancement are a threat to one's intellectual autonomy. Moreover, the view advanced can also explain why therapeutic cognitive improvements are generally speaking (though not always) compatible with the retention of intellectual autonomy.

Three Objections from Self-Direction

Learned helplessness²³

While drugs can help to help improve cognitive functioning, so can *cognitive scaffolding*—viz., reliance on technologies that complement our endowed cognitive abilities²⁴. Generally, gadgets used for cognitive scaffolding (i.e., iPhones, laptops, Google Glass) are located outside of our brains. However, this might just be temporary. As Michael Lynch (2016) has pointed out, the gadgets we rely on to store, process and acquire information have become, even just over the past decade, significantly smaller and wearable—as he puts it, trending toward seamless and 'invisible'. One example of such 'invisible' scaffolding includes the new Google 'Smart Lens' project, which is bringing to the market 'smart contact eye lenses' with tiny wireless chips inside along

²³For a related discussion of such cases in the context of autonomy and education, see Carter (*forthcoming*).

²⁴See for example Sutton (2010) and Heersmink (2015). Cf. Vygotsky (1980).

with a wireless antenna thinner than a human hair²⁵. Since the launch of this project, Samsung has countered (in February 2016) by patenting its own smart contact lenses which includes an invisible camera, with a display that can ‘project images directly into the human eye’²⁶.

One of the most provocative kinds of seamless cognitive scaffolding however comes in the form of wireless neural implants. Neural implants, increasingly used to assist individuals with prostheses to allow their brain and nerves to control and receive feedback from movements of prostheses, had previously required the use of wires that are connected to a device outside the agent’s body²⁷. A recent (2016) smart chip development can now be paired with the implants to allow for wireless transmission of brain signals²⁸. While this technology is currently being developed exclusively for therapeutic purposes, it doesn’t take much imagination to envision non-therapeutic uses for wireless neural implants.

In his recent book *The Internet of Us*, Michael Lynch (2016) anticipates a not unrealistic future scenario where sophisticated wireless neural implants—what he calls ‘neuromedia’—are the norm in a society. However, Lynch’s story comes with a twist:

NEUROMEDIA: Imagine a society where smartphones are miniaturized and hooked directly into a person’s brain. With a single mental command, those who have this technology—let’s call it neuromedia—can access information on any subject [...] Now imagine that an environmental disaster strikes our invented society after several generations have enjoyed the fruits of neuromedia. The electronic communication grid that allows neuromedia to function is destroyed. Suddenly no one can access the shared cloud of information by thought alone. [...] for the inhabitants of this society, losing neuromedia is an immensely unsettling experience; it’s like a normally sighted person going blind. They have lost a way of accessing information on which they’ve come to rely [...]

The moral Lynch draws, and which he develops in the book, is of course a cautionary one²⁹. Though while the worry expressed by the neuromedia thought experiment

²⁵(Otis and Parvis 2014). Cf., Senior (2014) for discussion.

²⁶<http://www.sciencealert.com/samsung-just-patented-smart-contact-lenses-with-a-built-in-camera>

²⁷<http://www.todaysmedicaldevelopments.com/article/smart-chip-enables-low-powered-wireless-neural-implants/>

²⁸This discovery was pioneered by scientists at Nanyang Technological University in Singapore on 11 February 2016. Here is their press release: <http://media.ntu.edu.sg/NewsReleases/Pages/newsdetail.aspx?news=81889a1d-edc2-479e-8350-dad40a767029>

²⁹Among other larger points, Lynch contrasts the ease by which knowledge is acquired with the more cognitively involved task of gaining understanding, where the latter is at risk of being undermined rather than promoted by certain ways of managing information acquired from the web.

appeals to a futuristic scenario featuring ‘extreme’ cognitive scaffolding³⁰, the crux of the worry can be abstracted away from the details of his case, so as to apply more broadly to some of our currently available cognitive scaffolding, such as smartphones.

Here, the social-psychological notion of “learned helplessness” (e.g., Seligman 1972) will be useful in capturing the lesson. Learned helplessness, generally construed, occurs when one repeatedly experiences a lack of control over his or her environment, then resigns herself to such lack of control³¹. The former president of the Royal Institute of Navigation, Roger McKinlay (2016) offers, in a recent article in *Nature*, a clear everyday example of learned helplessness as it pertains to the maintenance of navigation skills. McKinlay notes that increased reliance on satellite navigation has led drivers to be less vigilant in tracking where they have previously driven compared to those drivers accustomed to relying on paper maps, and so (in simulation tests) are more inclined to drive past the same place twice without noticing. More generally, deteriorating navigation skills has in turn only increased reliance on satellite navigation, through a process whereby individuals are increasingly ‘giving up’ attempting to orient themselves—and are accordingly failing to develop parts of the brain responsible for spatial orientation³².

Putting this all together, a rationale emerges for at least one specific way in which cognitive enhancement can threaten intellectual autonomy—viz., by rendering individuals increasingly intellectually helpless³³. To the extent that one is helpless, one is unable even if one tries, to direct one’s cognitive affairs, in the absence of the enhancement in question. Let’s now consider two other potential rationales, which are motivated on the basis of different kinds of considerations.

Scaffolding, framing effects and the illusion of preference-based choice

An well-known variety of cognitive bias called the framing effect occurs when the presentation of a choice influences how individuals react to it (e.g., Tversky and Kahneman 1985). Framing effects reveal that our perceived control over certain kinds

³⁰By ‘extreme’, I mean that in Lynch’s case, the neural implants aren’t merely supplementing endowed cognitive faculties, but rather, effectively replacing such faculties. At the very least, the implants are relegating biological cognitive faculties to a auxiliary role in both belief maintenance and formation.

³¹Cf., Brehm (1966) for a different approach to human behaviour in the face of lack of control, called psychological reactance theory. Psychological reactance theory however is predicated upon the individual feeling that something is taken away from them. Such feelings however are less likely to accompany cases where loss of control is more subtle, as when loss of control is a byproduct of achieving other gains (e.g., afforded by new technologies). For an overview, see Baumeister and Vohs (2007, 723–5).

³²McKinlay (2016). For discussion on this point, see Maguire, Woollett, and Spiers (2006) (cited in McKinlay). More generally, see Rupert (2004) for an expression of worries about biological cognitive atrophy, in response to Rowlands (1999).

³³Of course, not all such individuals will have been intellectually helpless to begin with—the idea is that individuals trend toward becoming more helpless, by degree.

of choices is often illusory, and the shape that our online inquiries take is especially susceptible to such an illusion of control.

A concrete example of such a framing effect involves online searching. We are often led to believe that our own preferences are what's primarily responsible for determining what inquiries we in fact pursue online, in a way that overlooks, as Google design ethicist Tristan Harris (2016) puts it, the way 'choices are manipulated upstream by menus we didn't choose in the first place.' As Harris puts it:

When people are given a menu of choices, they rarely ask: "what's not on the menu?" "why am I being given these options and not others?" "do I know the menu provider's goals?" "is this menu empowering for my original need, or are the choices actually a distraction?"

Suppose, for example, you make an online choice (e.g., between five options which turn up in a Yelp search³⁴) about what activity to pursue when visiting a new town. In such a circumstance, you might believe the choice you make, from the options on the Yelp search menu best represents your own preferences. You then search again, to learn some more information about the specific activity you've tentatively chosen, and are in the process nudged by Google-auto-complete, when generates another choice from a menu³⁵. At the end of such a series, your curiosity is satisfied and your inquiry is complete. It is an interesting philosophical question to what extent you have just directed the particular chain of inquiries which have culminated in the new set of beliefs you've settled upon. This much seems plausible: the shape that online inquiry chains take is significantly influenced by undetected framing effects that are themselves the product of upstream technological design decisions³⁶.

The more general line of argument that emerges is the following: enhancement via intelligence augmentation, as when we outsourcing cognitive tasks to smartphones and other gadgets, subjects us to constant framing effects which often go unnoticed. While such gadgets obviously aid us in acquiring knowledge quickly and seamlessly, they—as this line of argument contends—undermine our intellectual self-direction by (and in a manner that typically goes undetected) diminishing the contribution that our own biological cognitive faculties make towards the shape our inquiries take³⁷.

³⁴This is an example used by Harris (2016), when noting we are often led to believe that the only salient choices, relative to the initial inquiry (e.g., nearby activities), are the ones which websites like Yelp provide.

³⁵See also Heersmink (2016) for additional discussion of some of the epistemic implications of auto-complete.

³⁶For further discussion, see Edelman (2014).

³⁷It is worth noting that our inquiries are often subject to framing effects that are non-technological in nature. This is a point Sunstein (2014) notes when discussing the inevitability of choice architecture—

Authenticity and self-direction

Some virtue epistemologists such as Christopher Hookway (2003) and Catherine Elgin (1996) have argued, in different ways, for the epistemological significance of certain kinds of emotions³⁸. Emotions can be influenced pharmacologically, sometimes therapeutically, though sometimes with the aim of enhancing emotional well-being in healthy individuals, as in what Peter Kramer (1994) calls ‘cosmetic pharmacology’³⁹.

As Felicitas Kraemer (2011, 52) has asked, when drugs are relied on to enhance our emotional well-being ‘Can emotional authenticity or inauthenticity be inferred from the naturalness or artificiality [...] of such drugs? How this question is answered is relevant to intellectual self-direction. For if those such as Hookway and Elgin are right that emotions can be epistemically significant, then to the extent such emotions are inauthentic, the self-directedness of inquiries influenced by such emotions appears *prima facie* called into doubt.

Kraemer’s own line, drawing from work by Kramer (1994) and Elliott (2004), is that we should take seriously that (for example) someone using Prozac might feel ‘like themselves’ for the first time ever and is not in doing so mistaken⁴⁰. On accounts of emotional authenticity which make the naturalness or artificiality of the enhancing agent relevant to emotional authenticity, such cases are difficult to explain.

Kraemer is led to the conclusion that emotional authenticity is to be regarded as a phenomenally felt quality. She writes:

The notion ‘emotional authenticity’ thus means the phenomenally felt quality that a person perceives with respect to his or her inner emotional state, no matter by which means (natural or artificial) it has been brought about [...] and whenever ‘the individuals experiencing it recognize their own feelings really as their own and identify with them’ (2011, 57).

viz., it is inevitable that many kinds of choices will be framed in some particular way rather than another. ‘Opting out’ of framing effects, at least in a wide class of decision points, is not a straightforward option. One might accordingly wonder to what extent technological design decision-making represents a case of particular interest (*vis-à-vis* autonomy) beyond the general interest of non-technological framing effects? I submit two points in response, firstly, cognitive offloading is becoming increasingly widespread as a strategy for gaining information (e.g., Lynch 2016), a consequence of which is that we are to a much greater extent appropriating new kinds of framing effects into our belief-forming processes. Secondly, the kinds of framing effects are, in the case of technology design in particular, a product of *other people’s* interests and purposes, where these purposes include deliberate manipulation (e.g., Sunstein and Thaler 2008). Thanks to Glen Pettigrove and Jennifer Corns for discussion on this point.

³⁸For an overview of recent work on the relevance to emotion in epistemology, see Brun and Kuenzle (2008). Cf., Brady (2013) for a notable recent contribution to debates about epistemic significance of emotional experience.

³⁹See for example, Elliott (2004).

⁴⁰She notes, in particular, Peter Kramer’s case of ‘Tess’ (2003, 1–21; 278), cf., Kraemer (2011, 52–3).

Kraemer's phenomenological account of emotional authenticity has a number of advantages over other accounts which will (for example) be faced with the problem of accounting for why some non-natural or artificial enhancing agents detract from authenticity while more common cases (as she writes, listening to Mahler, drinking wine) do not.

However, Kraemer's own account invites a further epistemological worry, one that is highlighted in recent experimental work by Newman, Bloom, and Knobe (2014) and Newman, Freitas, and Knobe (2015). In short, what these experimental studies found was that 'people's true self attributions appear to be influenced in a complex way by their moral judgments'⁴¹ through what Knobe calls 'positivity bias'. In particular, as Newman, Bloom, and Knobe (2014) put it:

people have a general tendency to conclude that the true self is fundamentally good—that is, that deep inside every individual, there is something motivating him or her to behave in ways that are virtuous' (2014, 203).

Such studies cast doubt on the thought that (as per Kraemer) the phenomenally felt quality that a person perceives with respect to his or her inner emotional state is such that, when one experiences it, one can reliably *recognize* their own feelings really as their own and identify with them. At least, from these findings, it is reasonable to conclude that, given the prevalence of positivity bias, one will be *ceteris paribus* more inclined to regard some emotions as aspects of one's true self if those emotions are judged positively than otherwise.

Putting this all together, an epistemological worry materialises: the more positively one regards a given emotional experience, the more likely one is to *mistake* that experience as characteristic of one's true self even when it is not. Pharmacological enhancement of emotions tends to generate in one the kinds of emotional experiences that will be regarded positively, from the subject's perspective. Enhanced emotions, as such, are thus likely to be mistaken as characteristic of one's authentic self *even when they are not*. This includes enhanced epistemic emotions (e.g., attentiveness, curiosity, pride), the sort which (as per Hookway, Elgin and others) can be relevant to virtuous inquiry. To the extent that the kind of intellectual self-direction that is crucial to intellectual autonomy involves one's authentic self, the foregoing looks problematic; enhancement of epistemically relevant emotions threatens to make one more likely to regard her inquiries as self-directed, in the sense of authentically self-directed, when they are in fact not.

⁴¹Newman, Freitas, and Knobe (2015, 4)

Cognitive Integration and High-Tech Autonomy

We're now in a position to put several pieces together. In §1, it was shown why overly self-reliant conceptions of intellectual autonomy are self-undermining. The virtuously autonomous agent, it was argued, *must* rely on others, and outsource cognitive tasks as a means to gaining knowledge and other epistemic goods, *up until the point that doing so would be at the expense of her own capacity for self-direction*. This will happen when a subject's intellectual agency is (in some relevant sense) caused to be disconnected from the way she acquires and maintains her beliefs. Call this kind of disconnect *agential disconnect*.

In §3, we saw three kinds of cases where some form of cognitive enhancement led to agential disconnect. In §3.1, cognitive enhancement via Lynch-style neuromedia appeared to cause agential disconnect by diminishing through habituation (viz., as in cases of learned helplessness) the contribution of one's biological faculties in her cognitive projects. In §3.2, cognitive enhancement via intelligence augmentation/scaffolding appeared to cause agential disconnect via subjecting individuals to constant framing effects which often go unnoticed and which generate an illusion of choice in the course of cognitive projects. In §3.3, pharmacological enhancement of epistemic emotions appeared to cause agential disconnect by making the subject more likely to regard her inquiries as authentically self-directed, when they are in fact not.

In what follows, I want to argue that these three proximate causes of agential disconnect canvassed in §§3.1-3.3 share a common distal cause⁴². Once this point is appreciated, it will be shown that cognitive enhancement, as such, is not doing any interesting work in accounting for the threat to autonomy posed in the three enhancement cases surveyed in §3. Rather, what's important is just how the cognitive artifacts in question are being incorporated into the subject's cognitive character.

Cognitive character and 'extended' agency

There is a well-developed framework in contemporary virtue epistemology which offers a way to model the comparative contribution of a subject's own agency versus other non-agential factors in one's intellectual endeavours. A key reference point in the history of epistemology where this framework was especially important came in

⁴²Of course, it might be tempting to conclude that cognitive enhancement, as such, is ultimately what's responsible for the agential disconnect in each of these cases. According to such a line, the fact that in each of these cases (of neuromedia, scaffolding and pharmacological enhancement, respectively), cognition is improved in healthy individuals—rather than merely therapeutically to correct some pathology—is itself a kind of difference maker in each case. A diagnosis along these lines would be implied by one who regards the threat to autonomy posed by Modafinil as opposed to Donepezil, as surveyed in §2, to be explained by the fact that the former is an enhancement, rather than an improvement. The argument that is advanced in this section shows why this kind of diagnosis is unworkable.

the wake of meta-incoherence objections⁴³ to standard process reliabilist accounts of knowledge. According to process reliabilist (e.g., Goldman 1976) approaches to knowledge, knowledge is true belief issued by a reliable belief forming process.

Problematically for this view, some processes which lead to belief generation are themselves disintegrated from the agent's own cognitive psychology, as was the case in Keith Lehrer's (1990) famous case of Mr. Truetemp. In that case, an individual—Mr. Truetemp—has (unbeknownst to him) a thermometer planted in his head, and which causes him to reliably form true beliefs about the ambient temperature. Even though Mr. Truetemp's beliefs are formed by a reliable process, they plausibly (contra the process reliabilist) do not qualify as knowledge⁴⁴.

The virtue reliabilist's explanation for why such beliefs do not qualify as knowledge is that (in Lehrer's case) the thermometer is not appropriately integrated within the agent's—to use Greco's terminology—cognitive character⁴⁵. In support of this point, consider that when Truetemp forms a true belief, it is hardly something for which we can credit *him*.

But this raises a further question: under what conditions does an external artifact, such as a thermometer, count as being appropriately integrated within the psychology of the individual such that it *is* part of one's character—viz., so that the product of the inquiry is something the agent can take credit for?

A promising answer to this question—one which offers possibilities for diagnosing the cases considered in §§3.1-3.3, can be found in recent literature at the intersection of virtue epistemology and active externalist approaches in the philosophy of mind and cognitive science⁴⁶. Consider, for example, the hypothesis of *extended cognition* (e.g., Clark and Chalmers 1998; Clark 2008), according to which cognitive processes can cross-cross the boundaries of the skin and skull and includes extra-organismic parts of the world, such as notebooks, iPhones, laptops, etc. Proponents of extended cognition insist that the matter of whether to include something as part of a cognitive process should be made on the basis of the functional role that thing plays, rather than on the basis of its location or material constitution. As Clark and Chalmers (1998, 8) put it, such judgments should be guided by a kind of parity principle:

‘if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is part of the cognitive process.’

⁴³For an early formulation of this kind of objection, see Lehrer and Cohen (1983).

⁴⁴See Goldman (2011) for discussion.

⁴⁵See for example, Greco (2010; 2003).

⁴⁶This literature has emerged around 2010 and has gained traction since. For an overview, see Carter et al. (2014).

What this principle implies is that in cases where, for example, one habituates oneself to outsourcing information storage and retrieval to a device (e.g., an online diary) as opposed to biological memory, then—given that we're prepared to count biological memory as part of a cognitive process—so likewise we should include one's iPhone, which is playing the same functional role as biological memory vis-a-vis information storage and retrieval, despite being located outside the head⁴⁷.

Proponents of extended cognition however are careful not to include just any kind of external artefact we interact with as part of a cognitive process. It would be problematic for example to suggest that certain devices we consult only occasionally qualify. And so it is a theoretical project for proponents of extended cognition to explain which extra-organismic artifacts we rely on are to be ruled-in as parts of our cognitive processes, and which should be excluded⁴⁸.

As Duncan Pritchard (2010) has pointed out, the question the virtue-epistemologist must answer—viz., under what conditions does an external artifact, such as a thermometer, count as being appropriately integrated within the psychology of the individual such that it *is* part of one's character—bears similarities to the question the proponent of extended cognition must answer in order to avoid ruling in too much as part of a cognitive process. Pritchard (2010) has argued, in the earliest paper to explicitly connect extended cognition and virtue epistemology, that these related questions might have a similar answer. In particular, the line he has pressed is that for some extra-agential item to be included as part of an extended cognitive process, it must *also* be cognitively integrated in such a way as to qualify as a part of the agent's own cognitive character.

One obvious such integration condition, one that has been advanced also by Clark (2008), involves *accessibility*; the external artifact must be accessible in a way that is broadly analogous to our innate faculties. In the case of cognitive scaffolding, this means that the iPhone or gadget must be (literally) available to one much like one's own biological memory is. Cases of individuals who only occasionally carry a phone thus do not qualify; likewise, analogously, in the pharmacological case, a smart drug, regardless of its efficacy, which is taken irregularly does not qualify. Another such condition involves *automatic endorsement*; the representational deliverances of the extra-organismic artifact should be automatically endorsed and should not normally be subject to critical scrutiny. Thus, in the scaffolding case, a gadget—say, a tactile visual substitution system⁴⁹ (TVSS)—is not cognitively integrated if the agent does not trust in a default way the deliverances of the TVSS, much like, in normal circum-

⁴⁷ For some notable critiques of this line of thinking, see Adams and Aizawa (2001) and Rupert (2004).

⁴⁸ See, for example, Palermos (2014) and Heersmink (2015).

⁴⁹ These convert the images recorded by a camera to tactile stimulation on the tongue. See Bach-y-Rita and Kercel (2003) for an overview. Cf., Palermos (2011).

stances, we trust uncritically the deliverances of our biological eyes; likewise, in the pharmacological case.

A third and important condition, one which Pritchard puts forward, is a kind of *cognitive ownership* condition—viz., whereby the individual in question endorses in some epistemically respectable way the extra-organismic entity which is playing a role in her belief formation and to appreciate it as reliable. Consider the following case: suppose some ethically compromised cognitive scientists were to surreptitiously install a tactile visual substitution system in an individual with non-working biological eyes—and without that individual having any awareness that such a system had ever been installed. The epistemic pedigree of whatever spatial orientation beliefs such an individual forms via the tactile inputs generated by the system is in this case undermined⁵⁰. Furthermore, notice that such an individual is in the relevantly same situation as Lehrer's Mr. Truetemp: neither is aware of the source of the reliability of the relevant beliefs, having never originally endorsed the extra-organismic source as a reliable one. By contrast, Pritchard suggests that if, for example, Mr. Truetemp *were* to come to find out about the thermometer and to appreciate it as a reliable source of his beliefs about the ambient temperature, then his cognitive success could be creditable to him in a way it is not in the original version of the case, even though in both cases, the temperature beliefs are caused by the thermometer. *Mutatis mutandis*, in the case of the individual with the tactile visual substitution system.

To the extent that this thinking is on the right track, we have a precedent, as well as some tools, for thinking about the cognitive enhancement cases canvassed in §§3.1–3.3 featuring agential disconnect, and just how in these cases contribution of the subject's own agency compares to the contribution from what is external to it.

Cognitive enhancement and cognitive integration

Cognitive enhancement via Lynch-style neuromedia appeared to cause agential disconnect, and thus to undermine the kind of intellectual self-direction that is crucial to intellectual autonomy, by generating through habituation a kind of 'learned helplessness' (in the sense articulated in §3.1). I want to now suggest that there is an even more fundamental cause of the agential disconnect we find in such cases. The more fundamental cause can be expressed in the language of cognitive integration.

Consider again Pritchard's diagnosis of the original presentation of the Truetemp case. Mr. Truetemp fails the cognitive ownership condition on cognitive integration, specifically because Mr. Truetemp lacks any conception of the reliability of the source of his own beliefs—which happen to be issued correctly by the implanted thermome-

⁵⁰Pritchard himself makes this point with reference to Clark and Chalmers' case of 'Otto'. See Pritchard (2010, 145).

ter. Lynch's neuromedia cases are presumably not like Truetemp in the respect that, in Lynch's thought experiment, those with neuromedia are plausibly aware that they are relying on wireless neural implants to form their beliefs. However, there is another respect in which Lynch's case lines up closely with the Truetemp case. In neither case does the subject have an epistemically respectable conception of the source of the reliability of the beliefs. Truetemp lacks such a conception because he fails to know about the thermometer. In Lynch's case, the neglect of epistemic hygiene has led to a profound kind of cognitive atrophy. As Lynch puts it,

Just as overreliance on one sense can weaken the others, so overdependence on neuromedia might atrophy the ability to access information in other ways, ways that are less easy and require more creative effort (2016, 22).

A consequence of such atrophy is that the individuals in question lack any conception, *in the absence of relying on the neuromedia itself*, of the neuromedia's reliability. Just as one cannot take cognitive ownership of the legality of the laws enshrined in a legal text by reading in that very text *that* its laws are legal, so likewise Lynch's neuromedia proponents are not in a position to satisfy the cognitive ownership condition by bootstrapping—viz., by simply relying on what their neuromedia tells them about the epistemic pedigree of neuromedia⁵¹, and endorsing the reliability of neuromedia on this basis. In sum, a lack of cognitive integration would suffice to account for the agential disconnect, that is, for why intellectual agency is caused to be disconnected from the way she acquires and maintains her beliefs.

Consider now the second kind of scaffolding case canvassed—where cognitive enhancement via cognitive scaffolding appeared to cause agential disconnect via subjecting individuals to constant framing effects which often go unnoticed and which generate illusions of choice. The paradigmatic example is one who comes to form a set of beliefs through some form of intelligence augmentation (e.g., Google Glass) as a result of an series of menu choices. Again, in the case described, the cognitive integration conditions are not suitably met; this is in particular because the cognitive ownership condition on cognitive integration fails, though for a different reason than in Lynch's neuromedia case. In the neuromedia thought experiment, the individual will plausibly concede her helplessness in the absence of the neuromedia (and in the presence of neuromedia, can not satisfy the condition in an epistemically respectable way). In the cases discussed in §3.2, however, the individual is likely to claim cognitive ownership of her own inquiries. The problem is just that the individual significantly mistaken about what she is entitled to claim.

⁵¹For a related argument to do with extended cognition and epistemic circularity, see Carter and Kallestrup (2017).

Given that, as Harris (2014) points out, design decisions often significantly influence upstream menu choices, the salient explanation for why—in the kind of case described, at least—certain inquiries take the shapes they do, and culminate in the beliefs that they do (correct or otherwise), is technological design rather than the individual's own preferences. For those who outsource certain kinds of inquiries entirely to menu-driven search apps (e.g., Yelp reviews, Spotify music suggestions, etc.) the beliefs which serve as the termination of these inquiries might be very different if the menu choices were different, if one searched through different apps, or by means with fewer automated suggestions.

This is not to say that individual preferences play no explanatory role in such inquiries at all. Preference plays a guiding role even in inquiries that are almost entirely and uncritically influenced by menu-choice design. Rather, the point is that in some of these cases, the subject is led to claim a mistaken level of ownership, oblivious to the framing effect and its influence in her belief formation. This failure of cognitive integration between the individual and the technology whose influence she lacks a conception of, suffices to account for the agential disconnect which undermines intellectual self-direction.

Finally, in the case of pharmacological enhancement considered in §3.3, things are more complex. In this case, the problem was one of authenticity: it was shown that pharmacological enhancement of epistemically efficacious emotions threatens to make one more likely to regard her inquiries as self-directed, in the sense of *authentically* self-directed, when they are in fact not. And this was due to the fact that enhanced emotions, as such, are for reasons outlined in §3.3, likely to be mistaken as characteristic of one's authentic self *even when they are not*.

Here I think it will be helpful to draw a diagnostic parallel between (i) the failure of cognitive integration in intelligence augmentation cases where technological design generates an illusion of choice in the direction of one's inquiries, and (ii) what can be a failure of cognitive integration in some (but not all) cases of pharmacological enhancement. In the former case, the subject's lack of a conception of how the technology relied upon influenced her cognising resulted in the subject's failing the cognitive ownership condition on cognitive integration. An analogous problem surfaces in the latter case, when the subject lacks a conception of how the pharmacological enhancement relied upon influences her cognising. To the extent that the former case constitutes a failure of cognitive integration, by parity of reasoning, so should the latter case.

High-tech autonomy

We've arrived at the conclusion that the three kinds of enhancement cases (e.g., as surveyed in §§3.1-3.3) which seemed to pose a threat to the retention of intellectual

autonomy, share a common feature: in each case, plausible conditions on cognitive integration are not satisfied⁵², and in a way that accounts for the kind of agential disconnect that underlies the agents' (respective) defect in intellectual self-direction.

I want to now go further to suggest that, in so far as these cases are ones where intellectual autonomy is undermined, the fact that they are cases of cognitive enhancement, as opposed to therapeutic cognitive improvement, is just an accidental feature. In order to appreciate why, we can briefly run variations on all three cases. In each variation, we'll hold fixed the cognitive enhancement element of the case but shift the cognitive integration present, so that the cognitive ownership condition on integration is satisfied. In each case, what results is that the agential disconnect characteristic of compromised intellectual self-direction is not present.

Consider first a spin on our 'learned helplessness' case. Suppose that the individuals in Lynch's thought experiment rely on neuromedia in a more epistemically hygienic fashion than he has us suppose they do—viz., by maintaining ways, without relying on neuromedia itself, to monitor and calibrate against their environment the neuromedia they're relying on, and so without allowing their other ways of forming and maintaining beliefs to atrophy. In such a circumstance, these individuals will be such that they would not surrender the capacities to be sensitive to potential faults in their neuromedia, specifically by maintaining other reliable methods for assessing its reliability. In such a circumstance, even when the neural implants are in perfectly working order, the individuals in this revised version of the story are plausibly a position to take a kind of cognitive ownership over the truths they acquire on the basis of their implants. They are, in this respect, in an analogous position with an 'enlightened Mr. Truetemp' who is able to monitor the deliverances of the thermometer with other faculties. In neither the enlightened Truetemp nor the enlightened neuromedia cases—where the cognitive ownership condition is satisfied—is it at all clear that we have agential disconnect of the sort that undermines intellectual self-direction⁵³.

⁵²One difference worth noting between intelligence augmentation devices and pharmaceuticals is that the former often lead directly to beliefs/information, whereas the relationship between pharmaceuticals and belief-forming is comparatively indirect. Despite this difference, both are integrated only if a kind of cognitive ownership condition is satisfied (even if the content of what one must endorse differs to some extent across these cases in light of the direct/indirect distinction).

⁵³Of course, one kind of rejoinder will be to embrace a strong form of what Kallestrup and Pritchard (2012) term *epistemic individualism*. According to the most general version of this thesis, positive epistemic status supervenes exclusively on biological properties of the subject. If epistemic individualism is true, then the notion of extended agency is hard to make sense of, which means the kind of cognitive integration we'd find in a revised version of the neuromedia case is ruled out *ex ante*. Epistemic individualism, widely embraced by epistemic internalists, is also tacitly embraced by epistemic externalists, such as Goldman (1979), who has remarked that epistemic justification is a matter of reliable processes, where the process themselves are seated in the agent. As Goldman (1979) puts it, 'A justified belief is, roughly speaking, one that results from cognitive operations that are, generally speaking,

Similar remarks can be made with respect to ‘enlightened’ variants of the enhancement cases noted in §§3.2-3.3. In the case of intelligence augmentation and framing effects, the subject is oblivious to the framing effect and its powerful influence in her belief formation. This is why the individual is not in a position to take cognitive ownership of the results of the inquiry, as self-directed. However, the situation is different if the individual were to become aware of the framing effects particular to the gadgets she is relying on, and to attain some conception of how such affects are inclined to nudge inquiries in particular directions, and thus how to avoid such effects when they conflict with her other preferences. One such mechanisms through which this could be accomplished is by actively undertaking ‘framing-effect’ debiasing, which a study by Almashat et al. (2008) has shown has been effective in medical contexts. In a similar vein, an enlightened variation of the pharmacological enhancement case from §3.3, where the relevant ‘debiasing’ would pertain to positivity bias (as reported by Knobe and co-authors) as opposed to framing-effect biases.

In each of the three enhancement cases, the more information one acquires about the mechanisms by which the extra-organismic artefact (be it drugs, technology) culminate in beliefs the agent holds, and the conditions that are required for these mechanisms to function reliably, the better positioned the individual is to appreciate when such mechanisms are not reliable, and thus to take cognitive ownership even when (as in the enlightened Truetemp case) the extra-organismic element seems to be doing most of the relevant work⁵⁴.

If this is right, then two final points can be gleaned. The first point is that, when cases of cognitive enhancement genuinely threaten to undermine our intellectual autonomy, it is not because they are cases of enhancement, as opposed to therapeutic improvement. Rather, the cases feature a lack of suitable cognitive integration. The

good or successful. But ‘cognitive’ operations are most plausibly construed as operations of the cognitive faculties, i.e., information-processing equipment internal to the organism’ (1979, 13). See also, for discussion, Carter and Kallestrup (2016, sec. 3.2). One reason to embrace epistemic individualism is that one might be opposed to the very possibility of extended cognition. However, this is a false choice; as Kallestrup and Pritchard (2012) have argued, epistemic individualism actually has a hard time making sense of mundane cases of testimonial knowledge dependence. See, along with Kallestrup and Pritchard (2012), also Kallestrup and Pritchard (2013b) and Kallestrup and Pritchard (2013a) for arguments from arguments against epistemic individualism.

⁵⁴Note that the kind of cognitive ownership that is necessary for cognitive integration might in some cases require simply appreciating of an external resource that it is reliable, and possessing some rough conception of what the mechanisms is reliable at doing (without additional cognitive command of the details). At least, this—as opposed to more robust requirements (e.g., that one understands at a greater level of sophistication how the mechanisms works) would seem to be sufficient for integration of the sort that’s apposite to *intellectual autonomy*. There are, of course, other varieties of autonomy: moral, political, etc. Perhaps different articulations of the kind of cognitive ownership condition on cognitive integration might be germane to these different aspects of autonomy. Thanks to Fiona Macpherson for discussion on this point.

corollary to this point is that cases of cognitive enhancement which feature suitable cognitive integration pose no obvious threat to intellectual autonomy, despite the fact that such cases might involve heavy epistemic dependence on extra-organismic elements of the world.

This verdict seems to conflict with our original diagnosis of the Donepezil versus Modafinil case from §2. That verdict was, recall, that cognitive enhancement in healthy agents, such as by relying on Modafinil, seems to pose a *prima facie* threat to intellectual autonomy that is not posed in equal measure in cases where an individual relies on drugs such as Donepezil for purely therapeutic purposes, viz., to slow the progression of Alzheimer's. This reading suggested that in the case of Modafinil, it was the *fact* of enhancement (which is not present in the Donepezil case) that seemed to make a difference with respect to intellectual autonomy.

The second key point to make is that cases like this, no less than the cases from §3, can be diagnosed in the language of cognitive integration, and so involve no essential appeal to enhancement. There are two elements to establishing this point. First, consider again Donepezil as used for Alzheimer's patients, which does not intuitively threaten intellectual autonomy. That it is used therapeutically is not an essential part of the explanation for why. Consider that, in the case of pharmacological therapeutic cognitive improvements such as Donepezil, the drug is administered with the aim of *preventing* change (e.g., by slowing change) in the agent's cognitive psychology, rather than causing it. However, other drugs with the same kind of therapeutic purpose can aid in achieving this cognitively ameliorative aim while dramatically inducing new changes in the agent's cognitive psychology in other ways. A notable example here are benzodiazepines (e.g., Xanax), which can be used to treat anxiety problems. When used therapeutically, these drugs can help to improve poor concentration for anxiety sufferers. However, drugs like Xanax, while helping to quell anxiety-driven distraction and lack of focus, can have cognitively detrimental side effects, including memory loss⁵⁵ and in some extreme cases anterograde amnesia⁵⁶. Given the prevalence of memory-loss denial⁵⁷ for those suffering memory loss, the use of Xanax therapeutically can engender cognitive disintegration in the subject despite the ameliorative cognitive affects it brings about by fulfilling its therapeutic function. Such individuals can accordingly have their intellectual autonomy compromised even when therapeutic drugs are fulfilling their intended function⁵⁸.

⁵⁵For example, as reported in a study by Chouinard (2004).

⁵⁶See Mejo (1992).

⁵⁷In the UK, memory loss denial as noted as among the factors which make it difficult to convince individuals suffering memory loss to seek proper medical care. <http://www.nhs.uk/conditions/memory-loss/Pages/Introduction.aspx>

⁵⁸Therapeutic use of Xanax can be brief, and while autonomy can be undermined in the long term, some threats to autonomy will be more short-term.

To the extent that therapeutic cognitive improvements, on the whole, generally do not typically pose a threat to autonomy, this is only because the introduction of new changes to system is an aberrant and accidental property of drugs administered in these circumstances. In normal circumstances—as in the example of successful administering of Donepezil—the subject is prevented by the drug from further cognitive deterioration (or at least, caused to have such deterioration forestalled), without the introduction of comparatively more dramatic new changes to the way she maintains and forms beliefs.

As for the case of Modafinil—to the extent that our comparison in §2 elicited the response that Modafinil is a greater threat to intellectual autonomy than Donepezil as used therapeutically, that is because Modafinil when functioning normally as a cognitive enhancement (as when taken by healthy individuals to gain a cognitive advantage) causes new changes which must be in some epistemically respectable way appreciated by the subject in order for her to take cognitive ownership. An individual taking Modafinil for the first time is likely to be unaware of the specific effect the drug is having and how it is contributing to the individual's cognitive success. Such an individual begins to trend closer to the unenlightened (and cognitively disintegrated) Mr. Truetemp who trusts the deliverances of the thermometer but fails to appreciate it as a reliable source of his beliefs. However—and this is a point that has been expressed by Pritchard (2010, sec. 4), time is a factor which can make possible such appreciation. As Pritchard notes, provided the individual (who undergoes a change to her cognitive architecture) in question is suitably epistemically vigilant, she will acquire track record evidence about the way she is forming beliefs when utilising the drug. Over time, one who is vigilant in this way can plausibly take a kind of cognitive ownership of the beliefs formed through the drug which is not possible, say, the first time the drug is taken.

Putting these points together, given that enhancements when functioning normally induce new changes in the individual's cognitive architecture (whereas therapeutic improvements when functioning normally aim to correct or slow the progression of some pathology or defect) additional demands are made for suitable cognitive integration in all enhancement cases which are more relaxed in the cases where the drug (e.g., Donepezil) when functioning normally does *not* cause substantial new changes but rather functions so as to *prevent* or forestall new changes.

These observations about therapeutic cognitive improvements and cognitive enhancements generally speaking suffice to explain initial reactions to the comparison in §2. Modafinil, in short, is the sort of thing which requires more by way of cognitive integration than Donepezil. And so for any given case of Modafinil use, the likelihood of cognitive integration is lower than in the case of Donepezil for which

the standards are more relaxed⁵⁹. And this is so even though *some* cases of therapeutic improvements (e.g., consider the memory-loss side effects of Xanax) bring about significant new changes to the agents' cognitive architecture and thus require more by way cognitive integration, and *some* cases of cognitive enhancements (e.g., Modafinil, as used epistemically vigilant subject) can become cognitively integrated over time. What goes for pharmacological enhancements goes for other forms of cognitive enhancement, such as cognitive scaffolding⁶⁰.

Concluding Remarks

Cognitive enhancement is profoundly controversial. Bioconservatives and other critics of what they perceive as 'techno-progressivism' and 'post-humanism' have offered a range of anti-enhancement arguments, many of which are based on ethical considerations for why cognitive enhancement is dangerous or immoral. These include arguments to the effect that enhancement will undermine human dignity and preclude the possibility of meaningful achievements, by artificially removing obstacles the overcoming of which gives meaning to our lives⁶¹. Ethically driven arguments against cognitive enhancement are not only registered by bioconservatives or for that matter by ethical deontologists; on utilitarian grounds, Julian Savulescu and Ingmar Persson (2012)—ardent proponents of embracing moral bioenhancement—have influentially maintained that, given the ease by which available technologies have made possible the destruction of the human race (e.g., bioweapons, nuclear weapons, etc.) cognitive enhancement is currently too dangerous to pursue, at least until we can morally improve ourselves.

These ethical concerns about cognitive enhancement might be valid⁶². They are however orthogonal to the specifically *epistemic* question of whether availing ourselves to the latest science and medicine in order to improve ourselves cognitively (beyond health levels of functioning) threatens to undermine our capacity for intellectual autonomy. I've shown how, at least initially, it looks like cognitive enhancement (in three different kinds of cases) poses a direct threat to undermining intellectual autonomy by undermining our capacity for intellectual self-direction. Furthermore, it

⁵⁹The standards for cognitive integration should be understood as commensurate with the degree of change in the individual's cognitive architecture. In the Donepezil case specifically, the standards will be lower because the changes will be merely accidental ones, rather than (as in the case of enhancement) the *intended* effect.

⁶⁰A classic example of cognitive scaffolding as used for therapeutic purposes is Clark and Chalmers' classic case of Otto, who slowly replaces his failing biological memory with a notebook for the purposes of information storage and retrieval.

⁶¹For expressions of such arguments, see for example Kass (2004), Sandel (2009) and Harris (2011). For recent criticism to this general line of argument, see Carter and Pritchard (2016) and Bostrom (2005).

⁶²See, however, Carter and Gordon (2015) for a recent critique of Savulescu and Persson's argument.

appeared that epistemic dependence on technology and drugs for the purpose of therapeutic cognitive improvement (with a purely restorative aim) was immune to this charge. I have attempted to provide a different diagnosis. Drawing from recent work on virtue epistemology and extended cognition, I hope to have shown that the notion of enhancement as such is theoretically unimportant for accounting for why certain kinds of high-tech epistemic dependence genuinely threaten to undermine intellectual autonomy and others such kinds of dependence don't. If my diagnosis is correct, then just as some therapeutic uses of technology and medicine can undermine autonomy, so likewise, 'high-tech' intellectual autonomy is not an oxymoron, but a very natural result of combining epistemic dependence with epistemic vigilance. In short, whether embracing new cognitive enhancement technologies is a ultimately threat to maintaining virtuous intellectual autonomy is not a matter of what we're depending on (e.g., material constitution, location), or why we're depending on it (to correct a pathology or gain an advantage), but rather is a matter of *how* we're depending on it—which, and contrary to some bioconservative jeremiads to the contrary, remains largely in our own hands⁶³.

References

- Adams, Fred, and Ken Aizawa. 2001. "The Bounds of Cognition." *Philosophical Psychology* 14 (1): 43–64.
- Almashat, Sammy, Brian Ayotte, Barry Edelstein, and Jennifer Margrett. 2008. "Framing Effect Debiasing in Medical Decision Making." *Patient Education and Counseling* 71 (1): 102–7.
- Bach-y-Rita, Paul, and Stephen W Kercel. 2003. "Sensory Substitution and the Human-machine Interface." *Trends in Cognitive Sciences* 7 (12). Elsevier: 541–46.
- Baehr, Jason. 2011. *The Inquiring Mind*. Oxford: Oxford University Press.
- Battaly, Heather. 2015. *Virtue*. John Wiley; Sons.
- Battleday, RM, and A-K Brem. 2015. "Modafinil for Cognitive Neuroenhancement in Healthy Non-Sleep-Deprived Subjects: A Systematic Review." *European Neuropsychopharmacology* 25 (11): 1865–81.

⁶³I am grateful to Mark Alfano, Jennifer Corns, Emma C. Gordon, Orestis Palermos, Richard Heersmink, Fiona Macpherson, Glen Pettigrove, Duncan Pritchard and Jesús Vega Encabo for helpful discussion. I'm also grateful to audiences at the 2017 Pacific APA, the University of Edinburgh and the University of Glasgow Postgraduate reading party. Finally, I'd like to thank two anonymous referees at *Synthese* as well as Jesús Vega Encabo and Fernando Broncano-Berrocal for their work in putting together this special issue.

- Baumeister, Roy F, and Kathleen D Vohs. 2007. *Encyclopedia of Social Psychology*. Sage Publications.
- Boada-Rovira, Merce, Henry Brodaty, Patrick Cras, Stavros Baloyannis, Murat Emre, Richard Zhang, and Ranbir Bahra. 2004. "Efficacy and Safety of Donepezil in Patients with Alzheimer's Disease." *Drugs & Aging* 21 (1): 43–53.
- Bostrom, Nick. 2005. "In Defense of Posthuman Dignity." *Bioethics* 19 (3): 202–14.
- Bostrom, Nick, and Anders Sandberg. 2009. "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." *Science and Engineering Ethics* 15 (3): 311–41.
- 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: Oxford University Press.
- Brady, Michael S. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. OUP Oxford.
- Brehm, Jack W. 1966. *A Theory of Psychological Reactance*. Oxford: Academic Press.
- Brun, Georg, and Dominique Kuenzle. 2008. "A New Role for Emotions in Epistemology." In *Epistemology and Emotions*, edited by Georg Brun, Ulvi Dogluoglu, and Dominique Kuenzle, 1–31. Ashgate Publishing Company.
- Carter, J Adam, and Emma C Gordon. 2015. "On Cognitive and Moral Enhancement: A Reply to Savulescu and Persson." *Bioethics* 29 (3): 153–61.
- Carter, J. Adam, and Jesper Kallestrup. 2016. "Extended Cognition and Propositional Memory." *Philosophy and Phenomenological Research* 92 (3): 691–714.
- . 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.
- . Forthcoming. "Autonomy, Cognitive Offloading and Education." *Educational Theory*, special issue on *Cheating Education* (eds.) D. Aldridge and J. Tillson.
- Carter, J. Adam, and Duncan Pritchard. 2016. "The Epistemology of Cognitive Enhancement." *Unpublished Manuscript*, 1–25.
- Carter, J. Adam, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard. 2014. "Varieties of Externalism." *Philosophical Issues* 24 (1): 63–109.
- Chouinard, Guy. 2004. "Issues in the Clinical Use of Benzodiazepines: Potency,

- Withdrawal, and Rebound.” *Journal of Clinical Psychiatry* 65: 7–12.
- Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
- Clark, Andy, and David Chalmers. 1998. “The Extended Mind.” *Analysis* 58 (1): 7–19.
- Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.
- Code, Lorraine. 1987. *Epistemic Responsibility*. Hanover, NH: University Press of New England.
- Edelman, Joe. 2014. “Choice Making and Interface.” <http://nxhx.org/Choicemaking/>.
- Elgin, Catherine. 1996. *Considered Judgment*. Princeton, New Jersey: Princeton University Press.
- Elliott, Carl. 2004. *Better Than Well: American Medicine Meets the American Dream*. WW Norton & Company.
- Emerson, Ralph Waldo. 1841. “Self-Reliance.” In *Essays: First Series*.
- Gardner, Sebastian. 1999. *Kant and the Critique of Pure Reason*. Routledge.
- Goldman, Alvin. 1976. “Discrimination and Perceptual Knowledge.” *Journal of Philosophy* 73 (20): 771–91.
- . 2011. “Reliabilism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2011. <http://plato.stanford.edu/archives/spr2011/entries/reliabilism/>.
- Goldman, Alvin I. 1979. “What Is Justified Belief?” In *Justification and Knowledge*, 1–23. Springer.
- Greco, John. 2003. “Knowledge as Credit for True Belief.” In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, edited by Michael DePaul and Linda Zagzebski. Oxford: Oxford University Press.
- . 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- . 2012. “A Different Virtue Epistemology.” *Philosophy and Phenomenological Research* 85 (1): 1–26.
- Haddock, Adrian, Alan Millar, and Duncan Pritchard, eds. 2009. *Epistemic Value*.

- Oxford: Oxford University Press.
- Harris, John. 2011. "Moral Enhancement and Freedom." *Bioethics* 25 (2): 102–11.
- Harris, Tristan. 2016. "How Technology Hijacks People's Minds—from a Magician and Google's Design Ethicist." *Medium Magazine*, May.
- Heersmink, Richard. 2015. "Extended Mind and Cognitive Enhancement: Moral Aspects of Cognitive Artifacts." *Phenomenology and the Cognitive Sciences*, 1–16. doi:10.1007/s11097-015-9448-5.
- . 2015. "Dimensions of Integration in Embedded and Extended Cognitive Systems." *Phenomenology and the Cognitive Sciences*, 14(3), 577–598.
- . 2016. "The Internet, Cognitive Enhancement, and the Values of Cognition." *Minds and Machines*, 26(4): 389–407
- Hookway, Christopher. 2003. "Affective States and Epistemic Immediacy." *Metaphilosophy* 34 (1-2): 78–96.
- Hume, David. 1772. *An Enquiry Concerning Human Understanding*. Hackett Publishing Company.
- Kallestrup, Jesper, and Duncan Pritchard. 2012. "Robust Virtue Epistemology and Epistemic Anti-Individualism." *Pacific Philosophical Quarterly* 93 (1): 84–103.
- . 2013a. "Robust Virtue Epistemology and Epistemic Dependence." In *Knowledge, Virtue, and Action: Essays on Putting Epistemic Virtues to Work*, edited by Tim Henning and David P Schweikard. London: Routledge.
- . 2013b. "The Power, and Limitations, of Virtue Epistemology." In *Powers and Capacities in Philosophy: The New Aristotelianism*, edited by John Greco and Ruth Groff, 248–69. London: Routledge.
- Kant, Immanuel. 1784. "Beantwortung Der Frage: Was Ist Aufklärung?" In *Berlinische Monatsschrift*, edited by Friedrich Gedike and Johann Erich Biester.
- Kass, Leon R. 2004. *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*. Encounter books.
- Kraemer, Felicitas. 2011. "Authenticity Anyone? The Enhancement of Emotions via Neuro-Psychopharmacology." *Neuroethics* 4 (1): 51–64.
- Kramer, Peter D. 1994. *Listening to Prozac*. Viking Press.
- Lehrer, Keith. 1990. *Theory of Knowledge*. Routledge.

- Lehrer, Keith, and Stewart Cohen. 1983. "Justification, Truth, and Coherence." *Synthese* 55 (2): 191–207.
- Lynch, Michael P. 2016. *The Internet of Us: Knowing More and Understanding Less in the Age of Big Data*. W.W. Norton.
- Maguire, Eleanor A, Katherine Woollett, and Hugo J Spiers. 2006. "London Taxi Drivers and Bus Drivers: A Structural Mri and Neuropsychological Analysis." *Hippocampus* 16 (12): 1091–1101.
- McKinlay, Roger. 2016. "Technology: Use or Lose Our Navigation Skills." *Nature* 531 (7596): 573–75.
- Mejo, Susan L. 1992. "Anterograde Amnesia Linked to Benzodiazepines." *The Nurse Practitioner* 17 (10): 44–50.
- Mohamed, Ahmed Dahir. 2014. "The Effects of Modafinil on Convergent and Divergent Thinking of Creativity: A Randomized Controlled Trial." *The Journal of Creative Behavior* 50(4): 252–267.
- Mohamed, Ahmed Dahir, and Chris Roberts Lewis. 2014. "Modafinil Increases the Latency of Response in the Hayling Sentence Completion Test in Healthy Volunteers: A Randomised Controlled Trial." *PloS One* 9 (11): 1-9.
- Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*. Rowman & Littlefield Publishers.
- Newman, George E., Paul Bloom, and Joshua Knobe. 2014. "Value Judgments and the True Self." *Personality & Social Psychology Bulletin* 40 (2): 203-216..
- Newman, George E., Julian De Freitas, and Joshua Knobe. 2015. "Beliefs About the True Self Explain Asymmetries Based on Moral Judgment." *Cognitive Science* 39 (1): 96–125.
- Otis, Brian, and Babak Parvis. 2014. "Introducing Our Smart Contact Lens Project." *Google Blog*, January. <http://googleblog.blogspot.co.uk/2014/01/introducing-our-smart-contact-lens.html>.
- Palermos, Spyridon Orestis. 2011. "Belief-Forming Processes, Extended." *Review of Philosophy and Psychology* 2 (4). Springer: 741–65.
- . 2014. "Loops, Constitution, and Cognitive Extension." *Cognitive Systems Research*, 27, 25-41.
- Peñaloza, Renée A, Urmimala Sarkar, David M Claman, and Theodore A Omachi.

2013. "Trends in on-Label and Off-Label Modafinil Use in a Nationally Representative Sample." *JAMA Internal Medicine* 173 (8). American Medical Association: 704–6.
- Persson, Ingmar, and Julian Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. OUP Oxford.
- Pritchard, Duncan. 2010. "Cognitive Ability and the Extended Cognition Thesis." *Synthese* 175 (1). Springer: 133–51.
- Reid, Thomas. 1764. "An Inquiry into the Mind on the Principles of Common Sense." In *The Works of Thomas Reid*, edited by W.H. Bart. Maclachlan & Stewart.
- Roberts, Robert C., and W. Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. OUP Oxford.
- Rowlands, Mark. 1999. *The Body in Mind: Understanding Cognitive Processes*. Cambridge University Press.
- Rupert, Robert D. 2004. "Challenges to the Hypothesis of Extended Cognition." *The Journal of Philosophy* 101 (8): 389–428.
- Sandberg, Anders, and Nick Bostrom. 2006. "Cognitive Enhancement: A Review of Technology." <http://diyhl.us/bryan/papers2/neuro/implants/Anders>.
- Sandel, Michael J. 2009. *The Case Against Perfection*. Harvard University Press.
- Seligman, Martin EP. 1972. "Learned Helplessness." *Annual Review of Medicine* 23 (1): 407–12.
- Senior, Melanie. 2014. "Novartis Signs up for Google Smart Lens." *Nature Biotechnology* 32 (9): 856–56.
- Sosa, Ernest. 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford University Press.
- . 2010. *Knowing Full Well*. Princeton: Princeton University Press.
- . 2015. *Judgment and Agency*. Oxford: Oxford University Press.
- Sutton, Jonathan. 2010. "Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process." In *The Extended Mind*, edited by Richard Menary. Cambridge: MIT Press.
- Taylor, Charles. 1991. *The Ethics of Authenticity*. Cambridge: Cambridge University Press.

- Tversky, Amos, and Daniel Kahneman. 1985. "The Framing of Decisions and the Psychology of Choice." In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, 107–29.
- Vygotsky, Lev Semenovich. 1980. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard university press.
- Zagzebski, Linda Trinkaus. 2013. "Intellectual Autonomy." *Philosophical Issues* 23: 244–61.