

# Cognitive Enhancement, Hyperagency, and Responsibility Explosion

Emma C. Gordon

*University of Glasgow*

[emma.gordon@glasgow.ac.uk](mailto:emma.gordon@glasgow.ac.uk)

*Abstract.* Hyperagency objections (e.g., Sandel 2009; Owens 2007; Nagel 2010; Hauskeller 2013) warn us, broadly, that cognitive enhancement has the potential to negatively impact our well-being by giving us *too much* control. In this paper, I aim to carefully formulate and engage with a prominent version of the hyperagency objection due to Sandel (2009)—viz., that cognitive enhancement may create an ‘explosion’ of responsibilities that will inevitably generate an influx of unfulfilled desires. I will first outline why this kind of worry might look *prima facie* persuasive, and then I’ll show that we can satisfactorily defend cognitive enhancement against it. To that end, I will canvass and evaluate three distinct strategies for responding to this style of hyperagency objection: (i) denying the capacitarian claim that enhanced capacities generate correspondingly increased responsibilities, (ii) questioning whether new responsibilities reliably generate additional desires to fulfil these responsibilities, and (iii) showing why experiencing an influx of unfulfilled desires is less likely for the hyperagent than we might first assume. This third strategy, I will argue, is the most effective. Further, the same strategy leads us to see that the hyperagency argument is ultimately a double-edged sword, in that its underlying reasoning problematically generalises in such a way that it prescribes the *reduction* of cognitive capacities in unenhanced agents. At the end of the day, then, if we are to resist cognitive enhancement, it should not be on the basis of a Sandel-style hyperagency argument.

## 1. Introduction

The human enhancement debate explores the ethics of using new and emerging drugs and other biotechnologies to improve aspects of ourselves and our lives<sup>1</sup>, from our cognitive capacities to our moral disposition<sup>2</sup> and relationships<sup>3</sup>. For present purposes, we will be interested in one

---

<sup>1</sup> See Persson and Savulescu (2008) for the origins of the debate on whether we have an “urgent imperative” to pursue moral enhancement. For some of the many responses to Persson and Savulescu, see e.g., Fenton (2010), Harris (2011), de Melo-Martín (2018), Jotterand and Levin (2017) and Hardcastle (2018). See also Earp (2018) for a contemporary proposal that aims to overcome some of the previously highlighted issues in the moral enhancement debate.

<sup>2</sup> This thought experiment is of course implausible with today’s technology. Of relevance to arguments related to hyperagency are, however, both actual and merely possible cases. Both kinds will be discussed here. For some caution though on overreliance on overstating what possibilities are currently available, see de Melo-Martín and Salles (2015).

<sup>3</sup> See e.g., Liao (2011) for work on enhancing the parent-child relationship and Earp et al. (2012) for arguments in favour of using “love drugs” to enhance romantic relationships when the well-being of children is at stake. More generally, see Clarke et al. (2016) and Juengst and Moseley (2016) for concise summaries of some of the main considerations that bear on whether—or when—we ought to enhance ourselves.

particular subset of enhancement—*cognitive enhancement*, which aims to augment our cognitive capacities past the point of correcting pathology (Juengst and Moseley, 2015). More specifically, we will consider one way in which the improvements associated with more radical forms of cognitive enhancement are – perhaps paradoxically – claimed to have a negative impact on our well-being<sup>4</sup>.

Imagine for a moment that you could effortlessly ‘level yourself up’ in real life, in much the same way as you might level up a character in a video game with a few clicks of a button.<sup>5</sup> Perhaps you have a harmless brain computer interface (BCI) or neural implant<sup>6</sup> implanted through a simple, non-invasive procedure, or perhaps you swallow a perfectly safe pill (i.e., one with no dangerous side effects) that immediately doubles your intelligence, your reasoning abilities, your information processing abilities and your focus. In short, imagine that through the use of cognitive enhancement biotechnology you are now cognitively ‘superhuman’. It is easy to see how, with this superhuman boost, your life could go better in a number of ways. For one thing, with all of these boosted capacities, you can more easily achieve your goals<sup>7</sup>. If you are a lawyer, you can now more easily win cases. If you’re a chess player, you can see dozens or even hundreds of moves ahead. And if you’re an oncologist, you can now diagnose cancer more quickly and more effectively design treatment protocols.

However, much more interesting than how your life could go better in such a circumstance is how your life could go *worse*. Along with acquiring all of this new power – including, the power to continue upgrading your powers, altering what options are available – you at the same time acquire a lot more *responsibility* than you had before<sup>8</sup>. Whereas before, you were less responsible for how things go for yourself, your loved ones, and your wider community—given the limitations on what you were *able* to anticipate and do about these things in light of the capacities you were originally gifted<sup>9</sup>—these limitations (or at least some of them) have now been lifted.

---

<sup>4</sup> For related discussion, see Earp et al. (2014).

<sup>6</sup> For an introduction to new developments in BCI and neural implant technology, see e.g., He et al. (2020).

<sup>7</sup> At least, you could more easily achieve the present goals that you have. As we will see shortly, Sandel maintains that the situation becomes very different once we factor in additional goals that you have on account of possessing the enhancements themselves, and the responsibilities possessing them generates.

<sup>8</sup> For a detailed discussion on this issue, see Maslen et al. (2012).

<sup>9</sup> An additional negative consequence of hyperagency, according to Sandel, is that it diminishes our capacity to appreciate something he takes to be valuable, which is an appreciation of ‘the gifted character of human powers and achievements’ (2012: 78). However, the negative import of this consequence would presumably be contested by an opponent of Sandel’s argument; that is, the arbitrary gifts of our original capacities are *worthy* of appreciating only if there is already something antecedently valuable about the kind of ‘chance’ Sandel notes we are all subject to.

The worry here—which we will call the *argument from hyperagency*—is primarily discussed in work by Michael Sandel (2009). Sandel’s thinking is that, perhaps counterintuitively, life for a superpowered person would inevitably be an unpleasant one, one that is marred by an ‘explosion’ of responsibilities which in turn give rise to an explosion of unfulfilled desires. As Danaher (2014) puts it, when we think about what it means to have control over this many constitutive aspects of our agency, we see it is not welcome. And to the extent that cognitive enhancement gives us this control, it in this respect a burden disguised as a blessing.

The following is a charitable reconstruction of Sandel’s hyperagency objection to enhancement, which takes the connection between hyperagency and an allegedly problematic ‘explosion of responsibility’ (2007: 103-4) as the driving idea, where this explosion of responsibility is taken to generate a ‘moral burden’ (2007: 104), presumably (we may assume) one we are not equipped to meet, and on the relatively weak presumption that we would desire to do so. The reasoning (which makes implicit premises explicit, to make the reasoning valid) is as follows:

*Argument from Hyperagency*

(P1) The enhancement of capacities makes probable a corresponding increase of responsibilities for the results of one’s choices proportionate to the degree of the enhancement.

(P2) To the extent that the enhancement of capacities makes probable additional responsibilities for the results of one’s choices, it generates additional desires to fulfil these additional responsibilities.

(P3) If these additional desires are not fulfilled, well-being will be impeded.

(P4) It is more probable that, through the enhancement of capacities, an increase of responsibilities and desires to fulfil them is not accompanied by fulfilled desires than that it is.

(C1) Therefore, it is more probable that enhanced capacities will impede well-being than not.

(P5) If it is more probable that doing something,  $\phi$ , will impede one’s well-being than not, then *ceteris paribus* one should not  $\phi$ .

(C2) Therefore, *ceteris paribus* one should not enhance one’s capacities.

---

However, the proponent of bioenhancement will find scope to potentially press back here, either by denying there is something inherently valuable about having the capacities we have by chance, or by granting this but maintaining that the value of enhancing our capacities overrides whatever value there is *inherent* in possessing the capacities we have by chance.

Sandel's hyperagency argument, part of his wider project in *The Case Against Perfection*, effectively subverts what we might initially be inclined to think about the relationship between our capabilities and our well-being. In this paper, I will suggest that we ought to reject this kind of template version of the hyperagency argument (as well as another version of the argument I'll consider, which doesn't rely on (P2) and (P3)), and that, importantly, there are potentially several different routes to doing so, all of which press back against different aspects of the above reasoning. In section 2, I'll briefly note why (P3) and (P5) should both be granted—at least if we interpret them appropriately charitably. Thereafter, I'll turn to the three potentially contested premises—(P1), (P2), and (P4). In section 3, I'll evaluate the *prima facie* plausible (P1), consider some further support for it, while also registering some potential concerns. Section 4 then offers cause to doubt (P2) but will ultimately suggest that all Sandel needs to do to accommodate my concerns is to *amend* his second premise in a way that does not involve any substantial concession. In Section 5, I show why – perhaps contrary to initial appearances – the weak spot in the argument is actually (P4) – namely, the claim that it is more probable that, through the enhancement of capacities, an increase of responsibilities and desires to fulfil them is not accompanied by fulfilled desires than that it is. I will show that this claim faces intractable objections on the basis of which we should reject it, and further, that the underlying reasoning behind (P4) also generalises in a problematic way. On the basis of these problems, the argument is unsound.

Section 6 concludes by considering the prospects of running a version of the hyperagency argument that doesn't rely on P2 (or by extension P3), and which instead opts for a more direct connection between responsibility failure and well-being depletion. We'll see that on such a revision of the argument, there remains a problem with a (variation of) P4, and that the revised version of the argument accordingly faces structurally similar problems.

## **2. Why we should accept P3 and P5**

Premise (3) of the argument is arguably the most plausible; an increase in unfulfilled desires – including desires to fulfil additional responsibilities, or to realise what we take to be exceptional potential – will very likely contribute to a depletion of well-being. We can see this result if we apply any one of a range of theories of well-being<sup>10</sup>. For example, if we look at hedonistic

---

<sup>10</sup> See e.g., Fletcher (2016) for a comprehensive overview of many different forms of well-being.

theories of well-being<sup>11</sup> – which tell us, roughly, that the more pleasure we experience the better off we are – we see that unfulfilled desires are very likely to induce *dis*pleasure, and thereby reduce well-being. And if we apply desire-fulfilment theories of well-being—which Heathwood (2017) notes are “nowadays undoubtedly one of the leading theories of well-being” (p.135)—we have an even more straightforward explanation of how hyperagency depletes well-being<sup>12</sup>. On such views, well-being is fundamentally a function of desire-fulfilment—and the unfulfilled desires of the hyperagent (e.g., to fulfil responsibilities, to realise seemingly unlimited potential, etc.) thereby immediately make a negative contribution to well-being.

For present purposes, I also think we can grant Sandel (P5)—i.e., the implicit premise that if it is more probable that doing something,  $\phi$ , will impede one’s well-being than not, then *ceteris paribus* one should not  $\phi$ . It’s important though to clarify what this premise is not saying. By virtue of the *ceteris paribus* clause, the premise is not committed to the implausible suggestion that the all-things-considered one should never do what will impede their well-being. (Such a thesis would rule out, contentiously, any normative requirements for sacrifice under any circumstances). Secondly, the clarification ‘*one’s*’ well being implies the premise is not making any claim about how the permissibility of our actions are restricted in the light of how they relate to *all* persons’s well-being. The idea captured here, and which I think can be simply conceded (in order to focus on more contentious aspects of the hyperagency argument) is that *all things equal*, one should refrain from doing what would probably impede one’s well-being.<sup>13</sup>

As a more general point – regarding P5’s place in the wider hyperagency argument—note that it is common to directly appeal at least in part to an individual’s own well-being when making evaluations of whether we ought to endorse new and emerging biotechnology—indeed, the popular welfarist approach to enhancement tells us that whether something even counts as an enhancement depends on whether it “increases [a person’s] chances of leading a good life in the relevant set of circumstances” (Savulescu, Sandberg and Kahane, 2011). There is a precedent, then, for arguing for or against certain forms of potential enhancement with explicit reference to how well-being is impacted—just as (P5) does. In sum then, I will assume from here on out that if we are to reject the *Hyperagency Argument*, then, we will need to look beyond (P3) and (P5).

---

<sup>11</sup> For contemporary defences of hedonism about well-being, see e.g., Feldman (2004), Crisp (2006), and Bramble (2016).

<sup>12</sup> See also Dorsey (2013) and Heathwood (2005).

<sup>13</sup> Note that the use of ‘impede’ I’m employing here is meant to be neutral with respect to substantive moral theses with respect to, e.g., consequentialism or deontology.

### 3. Evaluating (P1): Do enhanced capacities generate a corresponding increase of responsibilities?

Intuitively, Premise (1), which tells us that enhanced capacities makes probable a corresponding increase in responsibilities, might appear just as plausible as either of (P3) or (P5). We give a child fewer responsibilities than an adult, plausibly on account of children having less developed capacities—and we generally take them to be more responsible about how their life goes as their capacities develop. Likewise, those with dementia are thought to be less than fully responsible for what they do (and thus obligated to a lesser extent) because they lack the right kind and/or degree of mental capacity for the relevant kind of reasoning.<sup>14</sup>

There is theoretical precedent for taking the connection between capacities and responsibility to be a tight one. The view that responsibility tracks mental capacity, *capacitarianism*, is a central underlying assumption of much of our reasoning about responsibility (Vincent 2011; 2013).<sup>15</sup>

In order to appreciate the thrust of this capacitarian idea— which underlies Sandel’s P1 (on which new capacities generate new responsibilities) – it will be helpful to step away from the enhancement debate specifically and consider the motivation we find for the more fundamental idea that, even when we hold all else fixed, capacity increase is plausibly a difference maker when it comes to what responsibilities one has.

With this in mind, consider the following case, due to Vanessa Carbonell (2013), who thinks of capacity acquisition as capable of ‘triggering’ responsibilities one otherwise would have lacked:

...[S]uppose a man collapses on the railway platform and is dying while waiting for the paramedics. As the sole bystander I would be obligated to save his life but I do not know how. (Fortunately, the paramedics arrive just in time.) Coincidentally, a CPR course is offered at my workplace that day, and I take it. On my return commute, shockingly, another man collapses on the railway platform. No one else on the platform has the relevant knowledge, but now I do.

---

<sup>14</sup> For discussion on this point, see, e.g., Peel (2013).

<sup>15</sup> See, e.g., Fischer and Ravizza (1998), Fischer (2000) and Mele (2001) for discussion.

As Carbonell sees it, in this case, an obligation has been ‘triggered’ (2013, 247) by the capacity you have to *do* CPR in the second version of the CPR case but not in the first. Whereas your lacking the capacity to perform CPR if you try (in the first version, by lacking the relevant know-how) insulates you from any obligation to make the attempt at CPR, your possessing it (in the second version of the case, and holding fixed everything else) seems sufficient to trigger the obligation. That is, you are plausibly *responsible* for refraining to act in the second version of the case in a way you would not be responsible for refraining to act in a scenario (like the first version of the case) where your attempts at CPR would be no more reliable than chance.

Notice now how Carbonell’s triggering point seems to hold if we run a twist on the set up of the case and shift the details so that rather than to *take the CPR* course, you instead purchase the CPR Tempo app, which offers both audio and visual cues that aid the timing of chest compressions during the process of cardiopulmonary resuscitation (CPR).<sup>16</sup> As Clark et al. (2017) have suggested, if a capacity-generated obligation is generated by your gaining the capacity to do CPR via a class, it would also be generated by a capacity acquired via intelligence augmentation via the CPR Tempo app. It would at any rate, as Clark et al. suggest, seem unprincipled to diagnose the two versions asymmetrically simply on the grounds the latter capacity is technology derived and the former is not.

And what goes for capacity generated obligations to others plausibly goes likewise for the kind of responsibility we have over *our* own lives.<sup>17</sup> If our capacities to govern our own lives are limited, then likewise is the extent to which we are responsible for how our own lives go.

Accordingly, if the core capacitarian thesis is right, then it seems that as a person’s capacities are enhanced beyond the “normal” range, the person in possession of these capacities might in some sense become, as Vincent (2013) puts it, “hyperresponsible.”<sup>18</sup> More will be rightly expected of us, both in terms of (i) the set of things we are expected to do – and expect of ourselves that we do, (ii) the level of proficiency with which we are expected to do them.

At this point, Sandel’s P1 is looking to be on safe ground. Let’s consider now a few challenges, one of which comes from experimental data. Maslen et al.’s (2015) preliminary empirical work

---

<sup>16</sup> This example is due to Clark et al. (2017).

<sup>17</sup> Granted, a strict libertarian may disagree with this presumption; though some libertarians may only reject obligations imposed by the state while accepting that there may be moral obligations one has to oneself.

<sup>18</sup> de Sio et al. (2014) defend something similar in course of arguing that people in certain high-responsibility jobs—such as surgeons and pilots—may have a duty to use certain enhancement technologies.

indicates that lay people seem not to see an obvious positive link between use of cognitive enhancement substances and an increase on responsibility or accountability. This is so *even though* there is general agreement that responsibility tracks capacity in cases like those of the child and dementia patient above. This kind of result puts some pressure on Sandel's (P1), then—as Maslen et al. point out, “great divergence between lay and philosophical theories of responsibility puts an extra burden on the philosophers to explain why their theory is justified” (p. 125).

One way to defend (P1) in response to Maslen et al.'s (2015) empirical results might be to embrace a kind of ‘error theory’—an explanation of *why* the participants surveyed tended to deny that capacitarianism extends into the enhanced range. Here, we might appeal to some of the objections to this claim—the sort of objections that those who rejected that capacitarianism extends to cover enhanced capacities might make—and then argue that these objections are not fatal. This would give us a potential diagnosis of why the view was rejected, at the same time as further defending the view.

One obvious candidate for such an error theory is the main worry that Maslen et al. (2015) themselves raise about extending capacitarianism to cases of cognitive enhancement. In a nutshell, this worry is that the cognitive capacities that result from enhancement drugs and technologies are not capacities that *genuinely belong* to the enhanced person. We can see a similar sentiment reflected elsewhere in the literature, such as in the debate about whether one's enhanced achievements are really creditable to one, a question that bioconservative ethicists such as, e.g., Leon Kass (2004) answer in the negative.

However, Maslen et al. suggest that the foregoing idea is not an intractable problem for capacitarianism—or, for our purposes, for (P1) of the hyperagency argument. At most, they argue, the objection implies that capacities that derive from “*unnatural, external* means of becoming enhanced” (p. 126, italics mine) are not truly our own, and therefore, don't generate extra responsibilities. Here, we can assume that – by external – they mean advanced brain computer interfaces (BCIs), for example, which leaves other “modest, internal” forms of cognitive enhancement—most notably, pharmacological cognitive enhancements (PCEs)—untouched. The hyperagency argument would then, at the very least, still potentially raise a serious concern about highly effective cognitive enhancement drugs.



However, regardless, there are also other defences available to the capacitarian, and thereby to Sandel in defence of (P1). The idea is that – at least as recent work in the philosophy of cognitive science (e.g., Clark and Chalmers 1998; Palermos 2011; 2014; 2015) suggests – *even when* capacities do in fact derive from “unnatural, external means of becoming enhanced,” it may very well still be that the capacities are *ours* in the sense that is relevant to credit attributions and reactive attitudes such as praise and blame that go hand and hand with responsibility. To take one notable example often used to illustrate this idea, consider the use of a smartphone rather than on-board biomemory for the purposes of storing and retrieving information. According to proponents of extended cognition, what matters for the purposes of whether the information you’ve stored in the smartphone is part of *your* memory is not simply settled by asking whether the smartphone itself is something ‘unnatural’ or ‘external’ to your biological cognition. Rather, the relevant question is whether the smartphone is suitably *integrated* into your cognitive architecture. For Clark (2010), at least, such integration – which isn’t going to be satisfied in just any use of a smartphone – will require that one’s smartphone is functionally on a par with biomemory along a number of key dimensions; that is, it must be as easily accessible, the information retrieved must be default trusted, it must be easily available as needed, and the information (as with biomemory) must have been previously endorsed. In short, as Clark sees it, if you are *using* your smartphone in a way that is functionally analogous to how you use biomemory for the task of information storage and retrieval, there is no in-principle barrier to counting the phone’s storage as part of your (extended memory), such that – when you then retrieve information from external memory -- we may then attribute that memory based belief to *you*, and credit you in a way that is analogous to how we might credit those relying on biomemory. And what goes for smartphones, will – for proponents of extended cognition – go for other kinds of enhancement via intelligence augmentation, including, e.g., tactile-visual substitution systems, BCI brain chips, etc.

Of course, the extended cognition approach in the philosophy of cognitive science, though increasingly popular, remains contested (see, e.g., Adams and Aizawa 2008 for some notable criticisms). I mention it here merely to sketch a potentially even stronger line one might take in insisting that the kind of reasoning one might appeal to in denying P1 by denying that capacitarianism extends to enhanced capacities doesn’t hold water. That is, one might maintain that in principle even the most radical kinds of enhancements to our capacities could generate an influx of responsibilities, so long as the external artifacts that feature in the enhancements are

suitably integrated into one's cognitive architecture by the functionalist criteria of extended cognition.

In sum, then, while there is at least some cause to doubt the capacitarian commitment underlying (P1), so too are there available defences—and, at worst, merely a concession that the target of the hyperagency argument is narrower in scope than we might first think, applying only to pharmacological cognitive enhancements.<sup>19</sup> With that said, we can now turn to consider two other potential weak spots in the hyperagency argument—Premise (2) and Premise (4).

#### **4. Against P2: Do such new responsibilities generate desires to fulfil them?**

What about (P2), which tells us that the new responsibilities of the enhanced generate new desires to fulfil these responsibilities? For one thing, this claim is not necessarily or analytically true; it is, after all, of course *possible* that one could have additional responsibilities while at the same time—e.g., due to some kind of moral defect—simply lack the corresponding desires to fulfil them. Imagine, for example, a disillusioned CEO of a company whose shareholders have just given her more responsibilities than before. It is very possible, perhaps even realistic, that at least some of these responsibilities will not correspond with desires to fulfil them; one might be entirely *nonplussed* by the expansion of her responsibilities. Accordingly, as the worry goes, as well as not being analytically true, (P2) is also most likely not true simpliciter—there are probably individuals like the imagined CEO whose have absent desires to fulfil additional responsibilities incurred.

However, these kinds of worries do not really force any major revision to Sandel's argument, and so don't really give us substantive resources to defend cognitive enhancement against the hyperagency objection. This is for at least two important reasons. First, a weaker version of (P2)—i.e., one that makes the relationship between (i) additional responsibility and (ii) additional desires to fulfil these additional responsibilities a mere 'propensity' relation weaker than an entailment relation—would actually suffice with the other premises to generate the conclusion. In a bit more detail, just consider (P3), which says "If these additional desires are not fulfilled, well-being will be impeded." Suppose we were to tweak (P2) so that it makes explicit that the

---

<sup>19</sup> One might wonder whether there is a limited scope to resist P1 on the basis of the possibility of potential enhancements with a very specific function – which is to cause us to better fulfil our responsibilities. One might think that the possibility of such enhancements is in tension with the idea that increasing capacities thereby increases responsibilities as per P1. However, even if there were fine-grained moral enhancements whose central function was to cause us to better meet responsibilities, this concession is compatible with the premise that *additional* responsibilities would emerge on the basis of the individual's gaining these responsibility-fulfilling capacities. As such, the kind of case envisioned remains compatible with P1.

relevant relation is a propensity rather than an entailment relation; notice P3's connection between unfulfilled desires and a depletion of well-being doesn't rely on the relevant unfulfilled desires being generated each time a responsibility is generated by an enhanced capacity. The claim (which we've already seen is plausible) just maintains that, *ceteris paribus*, unfulfilled desires impede well-being; thus, if (by a weaker version of P2) additional responsibilities (in light of enhanced capacities) generate *some* unfulfilled desires (*viz.*, to fulfil some of those additional responsibilities), then *to this extent* (via (P3)) well-being is going to be thereby impeded.

A second point of reply to the above worry, however, takes us a bit deeper into the initial framing of the argument. *Why* would failing to live up to an explosion of new responsibilities impede well-being? There are various potential fine-grained stories here. It might be that failing to live up to one's responsibilities diminishes one's well-being by diminishing one's sense of self-worth, which is displeasurable or otherwise debilitating for one.<sup>20</sup> Or perhaps, the sense that one is unable to meet her new-found 'hyperresponsibility' might give one a kind of 'decision paralysis'<sup>21</sup>, where one's own efficacy as an agent who can act on her own choices is (somewhat paradoxically) thwarted by a sense that she is in some respect not equal to the many new responsibilities she now has (for a discussion of this kind of line, see Owens 2007). Or perhaps one is simply psychologically affected in negative ways by what Sandel describes as the additional 'moral burden' one has to meet the explosion of responsibilities that enhancement might bring about.

What is important to note is that that the simple and relatively neutral idea that we have desires to fulfil our responsibilities that if unmet would impede wellbeing is broadly compatible with various of these more fine-grained explanations. Moreover, making explicit that unfulfilled desires to fulfil responsibilities impedes well-being allows us (in what is perhaps the least contentious way) to charitably construct Sandel's hyperagency argument to get the relevant *normative* conclusion that we should (on the basis of a claimed explosion of responsibility via enhancement) not enhance our capacities; We can't, after all, derive this conclusion from reasoning that does not include any specification of why it is that failing to fulfil responsibilities (including to oneself) negatively impacts one in a way that would be normatively strong enough to override the positive value associated with possessing new capacities via enhancement.

---

<sup>20</sup> A related idea here is that hyperagency, and the corresponding level of responsibility that it engenders, might in some way negatively affect one's sense of meaning. On this point, however, see Danaher (2014) for a useful rebuttal that aims to subvert this kind of worry by articulating some ways in which enhancement can promote rather than hinder meaning in one's life.

<sup>21</sup> See, e.g., Nagel (2010) for this kind of a concern; see Danaher (2014) for critical discussion.

I mention the above to contextualise the place of (P2) in the argument in a way that allows us to better appreciate certain kinds of criticisms against it. For example, let's return to our CEO who lacks desires to fulfil their additional responsibilities, and simply imagine a more extreme version of this character – a wholesale nihilist, subsisting in a state where (except for basic bodily sensations like hunger, etc.) lacks any commitment to any values, including to the value of fulfilling any would-be responsibilities whatsoever. Should the possible or even actual existence of such a character lead the proponent of Sandel's hyperagency argument to revise (P2)? No; this is because (as we've seen above) *so long as* there is at least some general propensity individuals have to desire to meet the responsibilities they have, making this fact in P2 explicit would offer us a suitably uncontentious way to capture a consideration about 'exploded responsibilities' that would have an important *normative* bearing, and in the way that is more than sufficient for (in addition with P1, P3, and P4) generating the argument's prescriptive conclusion. In sum, then, (P2) turns out to be like much like (P1); seemingly vulnerable to objections, but such that it can be defended in light of those objections at no important cost to Sandel's wider argumentative strategy.

That said, we will return to the place of (P2) in the argument in §6. There I will consider and evaluate what I take to be the most plausible way to run a variation on the Hyperagency Argument that does not rely at all on any premises about desire or desire fulfilment. That argument, we'll see, fares no better.

### **5. Against P4: Are unfulfilled desires really so likely for the hyperagent?**

The real weak spot is neither (P1) nor (P2) but instead (P4). Consider that if, *ex hypothesi*, our capacities were held fixed at present levels, and then additional responsibilities were incurred along with desires to fulfil these additional responsibilities, it is indeed very likely that this would result in unfulfilled desires to fulfil these responsibilities (and the ensuing lack of well-being that is implicated by such unfulfilled desires). But the situation is in fact very different when we appreciate that we are holding fixed not our present capacities but our *enhanced* capacities when considering the likelihood of desire fulfilment/unfulfillment. Even if enhanced capacities generate additional responsibilities and additional desires to fulfil these responsibilities, they also, at the same time, *make it proportionately easier* to fulfil whatever additional desires and responsibilities having one's abilities enhanced would plausibly generate.

Compare by way of analogy: right now, you plausibly lack any *responsibility* to disarm a complicated bomb, even if it is near you and its going off would cause significant destruction<sup>22</sup>. However, if you were an off-duty bomb dismantling expert and just happened to be walking by, the situation is different; such a person would plausibly have a kind of capacity-generated responsibility<sup>23</sup> to dismantle the bomb that is broadly analogous to the kind of responsibility a medical doctor, but not an ordinary passenger, has on a plane to attend to an unwell passenger. These points I take it, are in line with Carbonell's CPR example discussed in in support of the core capacitarian insight.

Now, consider this: the bomb dismantler and the doctor have additional capacity-generated responsibilities in these cases, but do these capacities at the same time generate additional unfulfilled desires they would not otherwise have had? They certainly would *if* these individuals desired to dismantle the bomb and help the passenger, respectively, but were unable to fulfil these desires. But they *are* able to fulfil them! That is, after all, precisely why we think that they have the relevant responsibilities that *they* have (but not untrained/unskilled individuals) in the first place.

The central problem with (P4), thus, is that overlooks the sense in which there is more plausibly a kind of *symmetry* than there is an asymmetry between (i) the extent of responsibilities the acquisition of enhanced capacities would generate and the (ii) extent to which subsequent desires to fulfil these responsibilities are themselves apt to be fulfilled.

There is, however, another problem with the premise, which is that the more general reasoning of which it is an instance would seem to support, counterintuitively, the *diminishment* of existing levels of capacities. After all, if there is a positive correlation between capacities and unfulfilled desires, then it seems that, to the extent that unfulfilled desires are a detriment to well-being such that we should avoid enhancing ourselves to avoid such unfulfilled desires, we should (by parity of reasoning) militate against unfulfilled desires by actively undercutting our current capacities! However, almost no one would think there is any moral imperative to undercut our current levels of functioning. Plausibly, Sandel himself would not want to explicitly sign on to this commitment. But the challenge then becomes explaining why—in a way that is not morally

---

<sup>22</sup> For the present purposes, we can remain agnostic on the matter of whether in such a circumstance an untrained individual would have a responsibility to *attempt* to dismantle the bomb. The point here is just that it's not the case that such an individual has a responsibility to in fact dismantle it.

<sup>23</sup> For a discussion of capacity-generated responsibilities in the special case of epistemic capacities, see, e.g., Carbonell (2013) and Carter, Clark, and Palermos (2018).

arbitrary—the supposed correlation between capacities and well-being-relevant lack of desire fulfilment justifies curtailing *enhancement* of our cognitive capacities but not curtailing *present* capacities.

In sum, there are two independent reasons to reject (P4): there is a reason to do with a kind of *proportional symmetry* that can be expected between capacities generating new responsibilities and one's capacity to meet them; and second, there is an overgeneralisation worry, such that a defence of P4 would seem to imply unpalatable results about curtailing our present unenhanced abilities.

I want to close by considering and replying to potential objections to both of these reasons to reject P4. First, let's zero in on the point about symmetry. Perhaps we could envision a proponent of P4 insisting that we are overlooking something important about how enhanced abilities generate new responsibilities – in a way that might be modelled not as linear growth but as an *exponential growth*, such that each additional increase in abilities correspond with exponentially increasing levels of responsibility. On such a picture, for example, a minor enhancement might generate only some seemingly manageable new level of responsibilities, whereas a *significant* enhancement might generate responsibilities that, modelled as an exponential function, approach infinity – viz., such that we should plausibly think that with any such significant enhancement our capacity to meet these exponentially generated responsibilities would be impossible or at least highly impractical.

There is, I think, a relatively straightforward way to respond the above point, which is to cast doubt on the idea that *even if* enhancing abilities generated *exponentially increasing* responsibilities (a claim that is itself dubious) that there is any interpretation of this idea that plausibly implies that a consequence will be *unfulfilled* responsibilities. First, a note about why the exponential growth interpretation is manifestly dubious. It is dubious because, in *unenanced cases*, the correspondence between increases in abilities and corresponding responsibilities seems best described as linear, and in a sense that allows us to make sense of a kind of symmetrical correspondence between the two. For example, a child has less responsibility to mow the lawn than an adult. How much less? The answer seems to be: approximately as much less as the adult has a greater capacity to do it. That is, the extent to which the adult seems to have more responsibilities than the child seems to be the very same extent to which the adult has greater capacities than the child. Extrapolating from this idea: it seems that the presumption should be in favour of thinking that

a similar kind of linear growth will characterise an increase in enhanced abilities and responsibilities rather than otherwise. This is the reason that the exponential interpretation is *prima facie* dubious. However, crucially, *even if* the relevant growth were exponential, this would seem plausible in itself only in so far as we would accept also that whatever radical enhancements generate near-infinite responsibilities also at the same time raise one increasingly toward the kind of omnipotence that would permit them to meet those abilities. Put another way, even if the proponent of P4 were granted the dubious claim that the relevant growth here is exponential, it would still be incumbent upon them to explain why we should accept this claim *and* the further claim that we should *not* expect whatever high levels of enhancements would trigger near-infinite responsibilities to themselves be enhancements to abilities that give us near-infinite powers to meet them. In sum, then, the envisioned ‘exponential growth’ style reply to the symmetry based objection to P4 doesn’t ultimately hold up.

Even so, we might envision a proponent of P4 taking issue with the second (independent) reason given for rejecting the premise, which is that it overgeneralises. For example, we can imagine someone might attempt to submit that there is some special feature of enhancement that makes it such that, when our abilities are enhanced, whatever corresponding responsibilities that are generated by their enhancement become more difficult to fulfil. The wider idea here would be that such a point is simply inapplicable to our natural abilities, and that for this reason, the kind of overgeneralisation argument against P4 (that it implies we should curtail existing abilities) doesn’t hold up.

The worry here, though, is that attempts to resist the overgeneralisation strategy in this kind of way are going to be arbitrary. At least, in the absence of a good reason to think that enhancement *as such* has an effect on abilities whereby when abilities are increased *in that way* (rather than by traditional training, education, etc.), then the responsibilities generated by them are harder to fulfil, the default presumption should be that the strategy in P4 is going to overgeneralise. The burden of argument, is, accordingly on the proponent of P4 to explain why the reasoning here would not overgeneralise, and it’s unclear to me how such an argument would not ultimately be an arbitrary one.

## **6. The Hyperagency Argument, Redux:**

By this point, we’ve seen that even if all other premises of the argument are granted to Sandel, P4 represents a crucial hole in the argument, one that reveals a seemingly intractable problem

with an attempt to reason – as Sandel does – from the capacitarian insight that responsibilities track capacities to the idea that an enhancing our capacities risks as he puts it ‘burdening’ us with too much responsibility, and in a way that will make our lives ultimately worse. The underlying problem, we’ve seen, is that – to the extent that we can expect enhanced capacities to generate additional responsibilities – those very same enhancements would more plausibly than not furnish us with the resources to meet whatever responsibilities the acquisition of those capacities brings about. And so the scenario on which enhanced capacities are accompanied with ‘too many’ responsibilities is ultimately not such that it represents a risk with reference to which (all else equal) we should forbear from enhancing capacities, provided we have other good reasons to do so.

That said, I want to conclude by circling back once more to (P2). One might wonder whether there is scope to save the Hyperagency Argument by (i) noting that (P4) – the crucial premise challenged – makes reference to *desires* about fulfilling responsibilities; and (ii) then insisting that a version of the Hyperagency Argument could conceivably be run without any commitment at all to premises that reference desires.

Let’s consider what such an argument might look like.

*Argument from Hyperagency (Version 2)*

(V2-P1) The enhancement of capacities makes probable a corresponding increase of responsibilities for the results of one’s choices proportionate to the degree of the enhancement.

(V2-P2) If these additional responsibilities are not met, then *ceteris paribus*, well-being will be impeded.

(V2-P3) It is more probable that, through the enhancement of capacities, the increased responsibilities would be unmet than met.

(V2-C1) Therefore, it is more probable that enhanced capacities will, *ceteris paribus*, impede well-being.

(V2-P4) If it is more probable that doing something,  $\phi$ , will impede one’s well-being than not, then *ceteris paribus* it is not the case that one should  $\phi$ .

(V2-C2) Therefore, *ceteris paribus* it is not the case that one should enhance their capacities.

There are two key points to make about the above alternative version of the Hyperagency Argument. First, note that (V2-P2) requires more defence than (P2) of the original version of the



argument. It leaves it incumbent upon the proponent of the Hyperagency Argument to explain *why* unmet responsibilities would all else equal impede well-being. One very weak claim here that would do the trick is the claim that we desire to meet our responsibilities, in conjunction with the observation that unfulfilled desires impede well-being. This weak claim is exactly what the original version of the argument made explicit. In this respect, the original version is meant to be a more charitable construction of the reasoning than would be this version.

However, more crucially, note that this alternative version of the argument that omits desire-based premises does not in any way eliminate the *kind* of problem that faced P4 in the original argument. Omitting desire-premises from the argument simply kicks the can down the road, in the sense that the argument relies on a premise *equally* as problematic as P4, which is premise (V2-P3), according to which it is more probable that, through the enhancement of capacities, the increased responsibilities would be unmet than met. The same kinds of reasons offered in §5 that count against P4 apply, *mutatis mutandis*, to V2-P3. Thus, the alternative version of the argument not only does not do better than the original version, but it seems ultimately to do worse, in that it inherits the kind of problem that faced P4 of the original argument, while also relying on a premise -- (V2-P2) -- which the original version characterised in a more charitable way.

## 7. Concluding remarks

In short, I hope to have shown here, firstly, that Sandel draws attention to an initially arresting idea—that new powers gained via cognitive enhancement bring with them new responsibilities of a sort that is often overlooked in discussions of how enhanced capacities bear on well-being. However, as we've seen, it is not straightforward to translate this idea into a compelling argument against cognitive enhancement. In the foregoing, I have attempted to focus in on some of the subtleties that feature in a charitable construction of a Sandel-style argument that transitions from enhancement and alleged hyperresponsibility to a diminishment of well-being that would justify forgoing enhancements despite their benefits. We've seen that while there are a few premises of the argument that might initially seem contentious, the real problem with this strategy of argument lies in a mistaken idea that we should *not* be expected to meet whatever additional responsibilities we have that would be implied by enhancing our abilities. Moreover, we've seen, that alternative versions of the argument seem no more promising than the original version. Of course, it might be that we *should* forego cognitive enhancements for reasons utterly divorced from considerations to do with hyperagency; equally, it might be that a very different

kind of hyperagency-style argument might have more promise than one that focuses on alleged problems with responsibility explosion.<sup>24</sup> However, I hope to have clarified just why it is that hyperagency arguments against enhancement that (like Sandel's) do trade on responsibility explosion are not ultimately promising.

## References

- Bramble, Ben 2016. 'A New Defense of Hedonism about Well-Being'. In *Ergo* 3.
- Carbonell, Vanessa. 2013. 'What We Know and What We Owe'. *Oxford Studies in Normative Ethics* 3.
- Carter, J. Adam. and Pritchard, Duncan. 2017. 'The Epistemology of Cognitive Enhancement'. In *Journal of Medicine and Philosophy* 44(2): 220-242
- Clark, Andy. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. OUP USA, 2008.
- Clark, Andy. 'Memento's Revenge: The Extended Mind, Extended'. *The Extended Mind*, 2010, 43–66.
- Clark, Andy, and David Chalmers. 'The Extended Mind'. *Analysis* 58, no. 1 (1998): 7–19.
- Andy Clark, Carter, J. Adam, Palermos, S. Orestis. 2018. "New Humans? Ethics, Trust, and the Extended Mind" in (eds.) Carter, J.A., Clark, A., Kallestrup, J., Palermos, S.O. and Pritchard, D.H. *Extended Epistemology*. Oxford: Oxford University Press, pp. 331-351.
- Crisp, Roger. 2006. *Reasons and the Good*. Clarendon Press.
- Danaher, John. 2014. 'Hyperagency and the Good Life – Does Extreme Enhancement Threaten Meaning?'. In *Neuroethics* 7: 227-242.
- de Melo-Martín, I. 2018. 'The Trouble With Moral Enhancement'. In *Royal Institute of Philosophy Supplement* 83: 19-33.
- de Melo-Martín, I., & Salles, A. 2015. 'Moral Bioenhancement: Much Ado about Nothing?', *Bioethics* 29(4): 223-232.
- de Sio, Filippo Santoni, Nadira Faulmüller and Nicole A. Vincent. 2014. 'How cognitive enhancement can change our duties.' In *Frontiers in System Neuroscience* 8.
- Dorsey, D. 2013. 'Desire-satisfaction and welfare as temporal'. In *Ethical Theory and Moral Practice* 16: 151-171.
- Earp, Brian. 2018. 'Psychedelic Moral Enhancement'. In *Royal Institute of Philosophy Supplement* 83, 415-439.
- Earp, Brian, Anders Sandberg and Julian Savulescu. 2012. 'Natural Selection, Childrearing, and the Ethics of Marriage (and Divorce): Building a Case for the Neuroenhancement of Human Relationships'. In *Philosophy and Technology* 25(4): 561-587.
- Earp, Brian, D., Sandberg, A., Kahane, G., & Savulescu, J. 2014. 'When is Diminishment a Form of Enhancement? Rethinking the Enhancement Debate in Biomedical ethics. *Frontiers in Systems Neuroscience*, 8, 12.
- Feldman, Fred. 2004. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Clarendon Press.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge university press.
- Fletcher, Guy. 2016. *The Philosophy of Well-Being: An Introduction* (Routledge: New York).
- Giubilini, Alberto and Minerva, Francesca. 2019. 'Enhancing Equality'. In *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 44(3): Volume 44, 335–54.

---

<sup>24</sup> A very different style of hyperagency objection, one that I've not taken up in this paper, is due to Michael Hauskeller (2011) – and which focuses moreso on the pursuit of hyperagency than its effects. See Danaher (2014) for helpful critical discussion of Hauskeller's proposal.

- Hardcastle, Valeria 2018. 'Lone Wolf Terrorists and the Impotence of Moral Enhancement'. In *Royal Institute of Philosophy Supplement* 83: 271-291.
- Harris, John. 2011. 'Moral enhancement and freedom'. In *Bioethics* 25(2):102–11.
- Hauskeller, Michael. 2011. Human enhancement and the giftedness of life. *Philosophical Papers* 40(1): 55–79.
- . 2013. *Better humans? Understanding the enhancement project*. London: Acumen.
- He, B., H. Yuan, J. Meng and S. Gao. 2020. 'Brain–Computer Interfaces. In *Neural Engineering*, edited by B. He. Springer: Cham.
- Heathwood, Chris. 2005. 'The problem of defective desires'. In *Australasian Journal of Philosophy* 83(4): 487–504
- . 2017. 'Desire-Fulfilment Theory'. In *The Routledge Handbook of the Philosophy of Wellbeing*, edited by Guy Fletcher. London: Routledge.
- Jotterand, F. and Levin, S. 2017. 'Moral Deficits, Moral Motivation and the Feasibility of Moral Bioenhancement'. In *Topoi*.
- Juengst, Eric. and Daniel Moseley. 2016. 'Human Enhancement' in *The Stanford Encyclopedia of Philosophy*, edited by E. Zalta.
- Kass, Leon. 2003. 'Beyond Therapy: Biotechnology and the Pursuit of Human Improvement,' *President's Council on Bioethics*, Washington, DC, 16.
- Liao, S. Matthew. 2011. 'Parental Love Pills: Some Ethical Considerations'. In *Bioethics* 25(9): 489-494.
- Maslen, Hannah, Filippo Santoni de Sio, and Nadira Faber. 2015. 'With Cognitive Enhancement Comes Great Responsibility?'. In *Responsible Innovation* 2: 121-138.
- Mele, Alfred R. 2001. *Autonomous agents: From self-control to autonomy*. Oxford University Press on Demand.
- Nagel, Saskia. 2010. 'Too much of a good thing? enhancement and the burden of self-determination.' in *Neuroethics* 3: 109–119.
- Owens, David. 2007. 'Disenchantment'. In *Philosophers without Gods*, edited by Louise Antony. Oxford: OUP.
- Palermos, S. Orestis. 2011. 'Loops, Constitution, and Cognitive Extension'. *Cognitive Systems Research* 27 (2014): 25–41.
- . 2014. 'Knowledge and Cognitive Integration'. *Synthese* 191, no. 8 (2014): 1931–51.
- Palermos, S. Orestis, and Duncan Pritchard. 'Extended Knowledge and Social Epistemology'. *Social Epistemology Review and Reply Collective* 2, no. 8 (2013): 105–20.
- Peel, Elizabeth (2013) 'Panic' and 'Blame' in Contemporary Dementia Discourse: an Analysis and Critique. In: 8th Biennial Conference, International Society for Critical Health Psychology, 22-24 July, University of Bradford.
- Persson, Ingmar and Julian Savulescu. 2008. 'The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity'. In *Journal of Applied Philosophy* 25(3): 162-177.
- Pritchard, Duncan. 'Cognitive Ability and the Extended Cognition Thesis'. *Synthese* 175, no. 1 (2010): 133–51. <https://doi.org/10.1007/s11229-010-9738-y>.
- Sandel, Michael J. 2009. *The Case against Perfection*. Harvard University Press.
- . 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.
- Savulescu, Julian, Sandberg, Anders and Kahane, Guy. 2011. 'Well-being and enhancement.' In *Enhancing Human Capacities*, edited by Julian Savulescu et al. Oxford: Wiley-Blackwell.
- Vincent, Nicole A. 2011. 'Capacitarianism, responsibility and restored mental capacities'. In *Technologies on the stand: Legal and ethical questions in neuroscience and robotics*, edited by B. van den Berg and L. Klaming. Netherlands: Wolf Legal Publishers.

———. 2013. 'Enhancing Responsibility.' In *Neuroscience and legal responsibility*, edited by Nicola A. Vincent. Oxford: Oxford University Press.