



This is a repository copy of *Enhancement and the conservative bias*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/206845/>

Version: Accepted Version

Article:

Davies, B. orcid.org/0000-0003-4612-7894 (2017) Enhancement and the conservative bias. *Philosophy & Technology*, 30 (3). pp. 339-356. ISSN 2210-5433

<https://doi.org/10.1007/s13347-016-0245-z>

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://doi.org/10.1007/s13347-016-0245-z>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Enhancement and the Conservative Bias

Ben Davies – This is a pre-publication draft. Please cite the published version.

The possibility of enhancing human beings has the propensity to exhilarate and appal in equal measure. A particular concern is that in pursuing the goods that an enhanced life promises, we will lose touch with the value of what we already have. A stance of *evaluative conservatism* suggests that even if enhancement promises significant goods, we should have some preference for the values inherent in our current mode of being. This paper considers one such argument from Nicholas Agar (2010; 2013). Agar raises a number of worries about what he calls ‘radical’¹ enhancement. This is an attempt to draw a threshold between absolute distrust of enhancement *per se*, and unbridled enthusiasm for the enhancement project in all its forms. Agar argues that at some level of enhancement – in particular, radical cognitive enhancement and radical anti-ageing enhancement – our values will change quite dramatically. He further insists that, however good our lives may be according to the values that we will come to have *post*-enhancement, we have reason from our current perspective to resist such a change because we risk ceasing to care about particular things that we currently value.² This conservative stance applies even if we will not care about that loss once we have undergone enhancement (since our values will change sufficiently that we no longer see our failing to care about previously valued objects as a bad thing).

After drawing some parallels with a similar stance outlined by G.A. Cohen, I outline Agar’s concerns as they relate to radical cognitive enhancement, and suggest that although some of his empirical predictions are overly pessimistic, his concern may have some force in particular cases. I then outline the related concern regarding radical anti-ageing enhancement. In this case, I suggest that as well as predictive problems, Agar’s argument suffers a conceptual defect that Cohen’s discussion helps to highlight.

1. The conservative bias

Before outlining Agar’s worries in detail, I will briefly note some similarities with a view outlined by G.A. Cohen (2011; 2012). Cohen insists more explicitly on a ‘conservative bias’ towards some sources of value: even if we could realise greater total value by destroying a current source of value (e.g. destroying a beautiful statue to make an even better one) a conservative bias gives us reason not to do

¹ He has defended more moderate enhancements elsewhere (2004).

² A related criticism is found in Hauskeller (2011; 2013, 163-181), who suggests that the relentless pursuit of enhancement involves a failure to be properly grateful for what we have already. He cites Voltaire’s (borrowed) aphorism that “the better is the enemy of the good” (2013, 175).

so. We should cherish sources of value not just for the values they realise, but for their own distinct nature, which includes respecting the historical trajectories by which they came to be the way they are.

As Cohen acknowledges, not everyone will find convincing all cases in which he suggests that we should hold a conservative bias; and nor will people necessarily agree about the threshold of added value that is sufficient to overturn a conservative bias in any particular case (since Cohen does not hold that the conservative bias should present an absolute prohibition on destroying existing sources of value). In my discussion, I will focus on just one particular source of value, for which it seems to me most people hold some degree of conservative bias more resiliently than they do regarding many other sources of value. This source is interpersonal relationships. Generally, in close relationships, each individual is understood to matter in large part by virtue of their distinctive relational qualities to the valuing agent, not just for intrinsic qualities that might be found in greater quantity in someone else. We do not abandon friendships simply for the promise of new friends, even if the latter promise us greater quantities of what we love and admire in our current friends. Anyone who is willing to do so is insincere in their professions of friendship, and perhaps fails to understand what friendship amounts to. Moreover, close relationships are often valued not only for what they give us at the moment, but also for their distinctive histories. In at least some cases, the fact that you are not currently getting pleasure or even value from a relationship is not sufficient reason to abandon it, if there is a valued historical trajectory behind that relationship (although we typically also need the promise of that historical value being realised again).³

Cohen suggests that one reason for holding a conservative bias is the importance of maintaining our sense of identity, not in a metaphysical but a “vaguer but very important sociocultural sense” of the term (2011: 226, fn6). Although he is deliberately unclear about what this amounts to, at a minimum it seems to involve our self-conception; our sense of personal identity is bound up with the particular values and attitudes we take towards the things that are currently important to us. The thought of waking up one day not caring about certain important people or commitments does seem to raise some concern of that person no longer being oneself in some important sense. So perhaps if certain radical changes threaten these central evaluative stances, they threaten an important sense of self that would make it worthwhile to seek enhancement in the first place; if the person who enjoys the goods of an enhanced life isn't ‘really’ me, what attraction could enhancement hold?⁴ This is by no means to suggest that every change ought to be resisted; on this view of self-conception, even quite radical changes might be germane to revealing who you ‘truly’ are.⁵ The thought is simply that *if* you conceive of certain relationships as partly constituting your identity, as many people seem to, the thought that radical

³ Kolodny (2003) argues that love cannot be explained by any non-relational feature of an individual.

⁴ C.f. Williams (1973).

⁵ See e.g. Levy (2011), responding to concerns outlined in Elliott (2003).

enhancement might change those attitudes will make it less attractive than it appears from a more external perspective.

In fact, some of Agar's discussion makes this idea clearer. He also draws a distinction between a purely metaphysical account of identity, and an "evaluative approach". An evaluative approach is concerned not with the question of what allows us to persist as the same person over time, say, but with "what makes our continued existence meaningful or valuable", quite apart from whether we are the same metaphysical individual (2013, 57).

Pugh et al. (2013) suggest that something like Cohen conservatism may ground a particular objection to certain forms of self-enhancement on the grounds we have reason to preserve our own nature (since we are a source of value) as it is, even if the change would be an improvement. Still, they note that some changes, including enhancements, can respect existing value, while others might be necessary to preserve existing value.⁶ As such, 'Cohen-conservatism' cannot generate any general objection to enhancement, though it might give us some guidelines about how to proceed: if correct, we should not undergo enhancement when it would undermine our very nature.

Similarly, Agar suggests that certain forms of radical enhancement would undermine our nature, by making us no longer human. As I will outline in more detail below, his view is that our human values and commitments are valuable to us *as humans*; we thus have reason to preserve what he calls 'species-relative' values (2010, 179-198; 2013, 17-32). Agar offers a couple of science-fictional examples to support this view. In *Invasion of the Body Snatchers*, human beings are replaced by alien 'pod people'. The pod people value, above all else, converting other humans to pod-personhood. And they do not value various aspects of human existence, such as emotions, that are valuable to us. Agar makes two claims. First, it is not objectively bad to be a pod person: pod people are satisfied with their lot, and glad (if that is the right word) to be rid of their troubling emotions. But it is also true, second, that it would be a mistake for a human to *become* a pod person, even knowing that once the transformation had occurred, they would have reason to be glad of it.

Before I continue with the central argument, it is worth explicitly outlining one element of Agar's argument that I will not be addressing further. In my view, the emphasis on *species*-relativism is unnecessary for the force behind Agar's argument. Consider the young radical who worries that she will grow up to be like her father, who is a ruthless capitalist. Ruthless capitalism may have various flaws, but it is surely not so radically inhuman as to fit into a species-relativist argument. Yet the young radical might have reason to resist this change if she can, or at least to set things up so that satisfying

⁶ See Powell, 2015; Danaher, Forthcoming for other defences of enhancement from a conservative perspective. Buchanan (2012, 26-51) also offers various considerations for why enhancement may serve a preservative effect in an environment that is constantly changing.

her later capitalist interests is more difficult, even if that will frustrate her when she is older. What seems crucial here is the level of *personal* importance that someone attaches to a value, not its species-typicality. Conversely, if someone (somehow) identifies very strongly with pod people – they are emotionally closed off, deeply troubled by the emotional reactions they do have, and wish to be rid of the last vestiges of them – it might not be so bad for them to be taken by pod people.

Note that this is a different claim than the thought that if someone wants to get rid of their emotions, they should try to do so. As Agar argues, emotion is a crucial part of our existence, and most ways of attempting to get rid of your capacity for emotions would backfire terribly. The crucial thing about becoming a pod person is that you can *safely* remove your emotions with (at least given the right set of value assumptions) no adverse side effects. Agar insists that there is no truly objective evaluative standpoint from which to compare human and pod-person values; I am suggesting that it may also make little sense to try to compare different individual values according to how ‘human’ they are. It might be that, as it happens, all humans do broadly subscribe to human values; but there seems nothing necessary about that possibility.

It is important to be specific about the implications of this suggestion. After all, even if there is nothing necessary about any particular human identifying with what we might identify as a ‘human value’, it might still be the case that for *most* humans, statements about human values are relevant to their prudential decisions. My suggestion, then, is not that Agar’s discussion of human values is without purpose. Rather, my suggestion is that what makes his argument compelling can be restated in simpler terms of personal value. While statements about human value may capture something important for many people, it is, in my view, statements about personal value that make for the most plausible defences of a conservative bias. Hence my decision to focus only on such questions.

One of Agar’s central arguments against certain radical enhancements is the claim that we will lose access to various ‘internal goods’, based on the views of Alasdair MacIntyre (e.g. 2007), associated with human activities (2013, 28-32; 42-46). There are various ‘external’ goods – such as prize money, or glory – to be associated with a particular activity, such as playing chess. Enhancement of our cognitive powers will enhance our capacities to gain these goods, at least so long as our opponents are not enhanced at a similar rate. But there are other goods that are more intimately connected with chess playing; and a person whose radically enhanced brain could win various external goods more efficiently might lose some of the ‘internal’ goods that a human chess player currently realises. Agar uses this distinction to argue that we must assess enhancements from the perspective of two ideals: the “objective” and the “anthropocentric”. From the perspective of the objective ideal, the value of an enhancement maps directly onto its degree of prudential value; that is, so long as an intervention really does improve our ability to perform some task that we value prudentially, it is of some value (though that value may be outweighed if it diminishes our capacity in some other area). But the overall value of

an enhancement cannot be determined by the objective ideal alone. Agar insists that we must also look to the anthropocentric ideal, which relates to the intrinsic value of a capacity, and so to its internal goods.

It may seem that this is an argument that cannot be so readily relativized to the individual. As Agar argues (2013, 42-46), we are able to project ourselves onto even exceptional performances, so long as they are recognisably human, imagining ourselves realising various achievements even if we could not physically or mentally do so; once we get beyond a human threshold, however, we are far less able to do this. But while that may be true of certain goods – and I will not argue the case either way – it seems clearly false of personal relationships.

The internal/external goods distinction is applicable to relationships: there are some goods that come from relationships that are purely external, such as having someone to share tedious chores such as cleaning the house or keeping track of finances. Other goods may be seen as internal both in a human, and in a more personal sense. Take companionship. For certain kinds of relationships, companionship seems to be an internal good; it is part of what it means to have a close friendship or romantic relationship with a person that you offer one another companionship. At this generic, human level, it might be the case that a person could realise companionship more efficiently with another partner than their current one, just as you might get more out of a chess game with one opponent than another. But this will not convince many people in committed relationships to look around for ways to realise that internal good more efficiently; even if I could plausibly realise companionship more efficiently with someone else, I am committed to realising it with *this* person. The good of companionship is not only internal to relationships in a generic way, as are the internal goods of chess; it is also internal to particular relationships in a personal way.

So my discussion will proceed with a modified version of Agar's view, that attaches the worry to *personal* values, rather than species-relative values. This is admittedly a major modification of Agar's view; but it is not, in my view, a change that affects the force of the particular argument that I am considering.

Cohen's view may give us another reason to reject radical enhancement, which fits with Agar's concerns in another way. As well as respecting sources of value for their distinct nature, we also have reason to respect particular sources of value because of our personal history with them. When someone becomes a pod person, they do not only destroy a specific source of value – themselves – they also seem likely to destroy various valuable relationships that they have. If I become a pod person, I will no longer care about my partner, except for wanting to turn her into a pod person too. My relationship with my partner is valuable in part because of the personal history I have with her. But Agar's arguments suggest that certain forms of radical enhancement will lead to a significant revision in my values, including in my attitude towards my partner and our relationship as a source of value. The next section outlines this

argument as applied to radical cognitive enhancement, raising some criticisms but offering moderate support. Section 3 then explains why even if the argument against radical cognitive enhancement is somewhat successful, these concerns do not translate to radical lifespan enhancement.

2. Agar's concerns – Radical cognitive enhancement

Agar claims that both radical cognitive enhancement (RCE) and radical anti-ageing enhancement (RAE) would change our fundamental interests and values. The predicted result on which I will focus is that we would be unlikely to feel the same commitments in our existing relationships. I will predominantly discuss the case of romantic relationships, although I also consider connections with other kinds of relationships, such as friendship.

This section discusses Agar's account of RCE, and suggests some problems that render his criticism rather weaker than he supposes. Nonetheless, I claim that he still raises a plausible concern for RCE, even if it is not as worrying as he supposes. However, I suggest in the following section that whatever plausibility this case has does not transfer to RAE.

Agar imagines various radical cognitive enhancements that could alter our capacities to appreciate complexity, subtlety and to take greater pleasure in more mentally taxing activities. However, he thinks that they will also reduce our toleration for the activities we currently enjoy. Agar suggests two worrying post-RCE scenarios with respect to a prospective enhancee's relationship. The first (2010, 185-186) is that she undergoes enhancement and her partner does not. The enhanced individual would lose enthusiasm for many of the common interests she shares with her partner that maintain their relationship, because they would be too simplistic for her to enjoy, at least in the same way as before. Perhaps the relationship between the cognitively enhanced and unenhanced would be akin to that between adults and children. We are clearly able to relate to children on many levels and indeed to enjoy fulfilling relationships with them; but those relationships are importantly different from the adult relationships that also enrich our lives. The kinds of activities through which you engage with children cannot centrally form the central basis of an adult life, or an adult relationship. Undergoing RCE when your partner does not, says Agar, will alter the character of your relationship in a way that destroys its status as the kind of source of value it currently is. As Kolodny (op cit, 165) suggests, certain kinds of relationship may depend on "viewing one's friend or lover as one with equal standing". If radical enhancement widens the gap too much, that may no longer be possible; as such, although you might continue to love your partner, that love might no longer be recognisable as a love among equals.

Agar (2013, 69) also suggests that there is a discrepancy in your fundamental interests pre- and post-enhancement, and that this means you will suffer "alienation" from your unenhanced past. If you take a purely instrumental view of enhancement – as many proponents appear to – then you might want to enhance your cognitive capacities, but you would also have a strong interest in maintaining various

psychological links – such as your memories, values and beliefs - with your pre-enhancement self. But following enhancement, you would “have a weaker interest in maintaining those particular connections”. However committed you are before enhancement to keep connections with your past life, that motivation will be considerably weakened following RCE.

The second scenario (2010, 186) is that both members of a couple undergo RCE. Although they would both be capable of enjoying the same activities (because their cognitive capacities would still be roughly on a par), Agar thinks it likely that their respective enhancements would take their interests in very different directions, again undermining the common ground that cements their relationships. In both cases, Agar thinks that we have strong reason to resist such changes, even if we were confident that we could find a new relationship with someone who was more in tune with our new interests, and even if that new relationship would be overall better for our enhanced selves than our current relationship is for us now. The fundamental connection with Cohen’s concern is obvious: the promise of greater value is not sufficient to make a change rational when it is destructive of a central current source of value (understanding destruction here not as literal annihilation of an object, but a radical change in its nature i.e. *destruction of the thing as it is now*).

I’ll now suggest that this argument may be too quick, and that it thus has rather less force than Agar supposes. Agar seems to want to apply the argument to all individuals,⁷ and for it to provide us with an very strong injunction against radical enhancement. I’ll argue that it can do neither of these, but should still give many of us a significant pause for thought before engaging in RCE.

I have suggested that the appeal of Agar’s concern may relate to the role in our own self-conceptions played by our involvement in and attitudes towards particular relationships. But you might deeply value your relationships, and yet see the promises of radical enhancement as obviously outweighing the cost of losing those relationships. In particular, those who do not find relationships to be an important part of their self-conception, as Section 1 suggested, are less likely to see the end of even very important relationships as signalling that enhancement would fundamentally change ‘who they are’. Moreover, changing who you are – in the sense of your self-conception – is not necessarily to be resisted. You might not at all like who you are, and so embrace the kind of radical attitude change that Agar predicts.

A second limit on Agar’s criticism of RCE concerns his empirical projections. He offers no positive reasons to think that a couple who both enhance to the same degree are likely to be pulled in very different directions. After all, even if our intellectual capacities have considerable influence over our

⁷ I make this assumption simply because Agar argues, *passim*, that ‘we’ should fear and reject radical enhancement. Since he makes no reference to particular circumstances, values (other than ‘human values’), I assume this ‘we’ is supposed to be global.

passions and personalities, they are not the only influence. So even if the particular activities that we now find engaging would come to bore us following RCE, I think we need some reason to suppose that we are unlikely to pursue the same kinds of pastimes. For instance, a music-lover might abandon her interest in Bach in favour of music of such depth and complexity that we cannot currently imagine it (187); but it seems likely that she will still be interested in *music*. So if her relationship with her partner is built around a shared love of music, why not think it likely that that interest will persist?

Similarly, even if your partner doesn't enhance, it is not clear that you would necessarily have nothing to share with them. For instance, in many cases where one member of a couple suffers from progressive dementia (an example Agar (2010, 184) thinks offers some parallels with RCE), the character of the relationship is significantly transformed; but there is no suggestion that the relationship is changed in a way that destroys its value. Cognitive capacity is not the only source of shared interests. Highly intelligent people can enjoy quite simple activities,⁸ so long as that is not their only source of stimulation; even if poetry is superior to push-pin, that doesn't mean that poetry aficionados cannot take pleasure in the occasional game. It is certainly true that as our cognitive capacities develop, we are no longer satisfied with a steady diet of the same activities that kept us entertained when we were younger, and this trend might continue with radical enhancement. So an unenhanced person who insisted on her enhanced partner only doing things that she would also enjoy would create problems in a relationship. But if the enhanced partner had space to exercise her newfound intelligence, perhaps the relationship could survive unequal enhancement if the couple also shared other interests. There are plenty of activities that can be enjoyed by enhanced and unenhanced alike, such as sport, sex, conversation, food, and travel.

On the other hand, a comparison with dementia may be overstretched. In the case of dementia, one individual may need the other as a carer and is also less clearly aware of the gulf between them. So there is both a motivation for the relationship to continue in some form, and less obvious resistance to a partnership of equals transforming into an unequal partnership. Further, as Agar notes, even if a relationship *can* survive dementia, it is clearly something that we would resist if we could. And even if it is true that some relationships can survive with significantly different levels of cognitive ability, Agar might note that the *level* of cognitive difference following genuinely radical enhancement will be significantly greater than we have seen before. So there is still a distinct risk here.

These considerations reduce the force of Agar's argument from an absolute injunction against RCE to a warning about the kinds of risks we might face when enhancing. At its most plausible, the conservative

⁸ Indeed, Hauskeller (2011) doubts that we can "believe that what makes Mozart great is entirely comparative, that there is nothing of intrinsic value in his music". If this is right, our alienation from current sources of value may be less extreme than Agar supposes.

bias gives us only *pro tanto* reasons to avoid cognitive enhancement. If the benefit is great enough, then these reasons may be outweighed by the benefit. But since Agar is worried only about *radical* changes, and these only apply when the enhancement is especially powerful, i.e. when its benefits are great, perhaps by the time the worry is relevant, the benefits will be strong enough to overwhelm it.

Nonetheless, these caveats do not undermine the concern entirely. Many people do care about retaining their particular personal relationships; that RCE poses a potential threat to that goal does seem to constitute a reason, albeit not a decisive one, to resist such interventions. If Agar's predictions retain some plausibility then, even if they are not sufficient to overwhelmingly rule out radical cognitive enhancement as a reasonable endeavour, they may present an additional risk that potential enhancees should consider carefully.

Further, they do not seem to affect an important further argument that Agar offers against RCE. As he puts it (2013, 76-77), "a consequence of accepting that it is prudentially good to enhance is that one could be subject to a prudential requirement to continue enhancing *ad infinitum*". The effect of this seems to be that we will never reach a point of stability. Agar compares this to the process of maturing from childhood: as children grow up, they leave behind various evaluative stances and become, to some extent, alienated from their childish views. But at the end of this process – ideally – is adulthood: an extended period of relatively evaluative stability. Such a period will never emerge if we undergo endless cognitive improvement. This argument becomes particularly powerful if RCE is undertaken along with radical anti-ageing enhancement (2010, 182-189); the prospect of a life that is open-ended from the perspective of ageing presents the possibility of genuinely endless cognitive enhancement.⁹

An anonymous reviewer has raised a potential objection to this idea. The conservative bias suggests that when choosing between a present source of value (V1) and a future source (V2), you should (often, if not always) choose V1. The reason is that V1 is in some sense unique, or at least incommensurate with V2. But this seems to suggest that we should *never* abandon V1, whereas both Cohen and Agar seem to think that we may sometimes prudentially opt for a novel source of value. In particular, Agar's suggestion that it is not prudentially wise to take a 'Peter Pan' pill, which freezes us in childhood and prevents our childish values from evolving, is an assertion of this kind (2010, 187-188).

The reviewer further suggested that Agar's response of appealing to the value of mature interests is not particularly convincing. They suggested that what seems problematic in the cases Agar outlines – such as the loss of a valuable relationship – is that the change is sudden. That is, if the same change were to occur over a longer period, we would not regard it as problematic: for instance, if two lovers drift apart

⁹ Again, this concern is echoed in Hauskeller (2013, 177) who worries that "There will...always be the possibility of something being even better than what we've got".

slowly over the course of a few years – perhaps even due to the fact that one of them loses interest having undergone RCE – this is not to be regretted in the same way as a sudden change might be.

If the force of Agar’s examples depends on the suddenness, rather than the degree, of change, how might this affect his argument? First, it might seem to cause a problem for his claim that what matters is the fact that something is presently valuable; if the loss of present value is only to be regretted if it occurs suddenly, it is hard to see how it can be the loss itself that is the problem, rather than its suddenness. Second, Agar would have to make a further empirical assumption: namely, that the kind of alienation and loss of interest that would occur after RCE would be sudden, not gradual. That is not an obvious assumption: radical changes in capacity might take some time to affect your preferences and sense of self, particularly as they might take some getting used to.

It is worth noting in the face of such concerns that Agar’s point is not that we must resist change at all costs. In particular, his discussion of ‘mature interests’ does not preclude those interests changing. The difference Agar sees (2010, 185) is that the things we lose interest in as adults “do not become completely meaningless to us” in the way that childish interests might. Adulthood, for Agar, is a period marked not by stasis, but by *relative* stability and finality compared with childhood: stability, in that significant changes happen less often, and finality, in that it involves “the physical, cognitive and emotional resources to arrive at final and decisive plans about one’s life” (2013, 72).

At the level of mature interests, then, it may seem that Agar would be in agreement with the aforementioned analysis; at the most general level, Agar thinks that it is good for us to have some stage at which we develop more secure and (to some degree) final interests. Indeed, Agar even speaks somewhat approvingly about a more gradual movement towards the goals of radical enhancement, a movement that “wouldn’t prevent us from relating to our former selves, our children, and our fellow citizens” (2010, 197). Perhaps one reason for this, which Agar doesn’t mention, is that a gradual change allows people – partners in a relationship, say – to affect the course of one another’s development in a way that cannot occur with sudden change.

For Agar, then, one of the valuable aspects of such gradual change is that it makes us less likely to lose our concern with things that currently matter to us, including our personal relationships. It is also possible that he would agree that it is rather less tragic if, say, a couple drifts apart slowly over time, than if one or both suddenly find that they have lost interest.

But doesn’t this suggest that what matters is *not* the loss of particular sources of value, but their sudden loss? After all, if their loss *per se* were to be regretted, surely we ought to resist even gradual change. Recall, however, that one fear Agar has about radical enhancement is not only that a relationship will end, but that you will come to regard it as “completely meaningless”. When we think about two people

drifting gradually apart, we may think of people coming to realise that they no longer have the same connection that they once did; but if we add the thought that when they think of their past relationship they find it meaningless, and simply cannot identify with it at all, then I think that there *is* something quite tragic about that. The important distinction here is between, on the one hand, whether a change is gradual or sudden and, on the other hand, whether it is radical or moderate. A gradual series of moderate changes that *adds up to* a radical change might be something to resist even if no individual change is itself regrettable.

Consider the following case: a young couple, who are deeply in love, become involved in politics. Over time the realities of political life draw them further apart, and by the end their relationship is one of mutual convenience, projecting family stability to the electorate. It may be that no sudden change occurred, nor that the couple regrets how things have turned out at the end of the transition. They may look back on their youthful love as naïve and idealistic. Despite the change being gradual, and approved of at the end, it seems to me that if they could foresee its occurrence they would have strong reasons to resist it at the beginning. At the beginning of their political careers they stand, unwittingly, to lose something of immense value to both of them.

In discussing the move from childhood to adulthood, Agar makes the following suggestion: “There is something a bit tragic about growing up”. This suggests that Agar does regard it as regrettable to abandon current sources of value for novel ones. It is just, he might say, that the gain of a stage of mature interests is sufficient to justify that sacrifice. Once we *reach* such a stage, however, similar sacrifices are not prudentially justified; indeed, on the radical enhancement model as Agar sees it, we would be sacrificing such a mature stage, not gaining one.

So this argument seems to me a further reasonable concern about RCE, perhaps particularly in combination with RAE. Still, rather than opposing radical enhancement altogether, what it may suggest is that if we value some level of evaluative stability – as many of us surely do – then we may need to draw a line at some point and call an end to radical cognitive enhancement, *even if* that line is to some extent arbitrary. Agar might reasonably ask why that line should not be drawn before undergoing *any* radical enhancement. My answer is that it might be, for some people. But there is no prudential requirement that the line be drawn there.

Compare the following analogy: you are walking down a beach with some friends, and various other groups unknown to you. This beach has a very special property: while it starts out fairly pleasant, it always gets slightly better the further you go down it. The sand is more comfortable, the water somewhat better for swimming, the beach bars serve nicer drinks and food (somehow, it also gets closer to the car

park for when you want to leave). It is also infinitely long;¹⁰ there is always prudential value to going just a little further. But if you go on forever, you will never get to enjoy the beach (other than the enjoyment of walking down a nice beach).

Some groups might decide not to get drawn into this mess at all: ‘We will sit here, where we entered the beach.’, they might say, ‘It’s good enough’. Others might get trapped into walking forever. But most of us could reach a point where we say that we have had enough of walking, and this bit of beach will do just fine. We are capable in many situations of not getting trapped into a never-ending cycle of improvements by drawing lines that we recognise are somewhat arbitrary. While I agree with Agar that endless improvement is a risk of RCE, it is not obviously a risk that cannot be managed.

Agar does consider this possibility, but suggests (2010, 188) that “the human pattern” is more valuable than a pattern that stops arbitrarily at some stage of cognitive development, because we do not currently view adulthood as simply another stage of development. As such, a life that stabilises in what we now call adulthood is better. But this seems to me to lack the force of most of his other appeals to the conservative bias.

The conservative bias, at its strongest, is a worry that we will lose something that is currently of value in the pursuit of something that is of more (objective) value. It is not simply an appeal to the way we see things now. The issue thus rests on whether it is valuable to have some period of evaluative stability, or whether what is valuable is to have the particular period of stability that humans currently have. If the latter, this is a worry both for endless enhancement, and *any* radical enhancement, since we would lose out on our current adult stage as a period of stability. If the former, this is still a worry for endless enhancement, but it does not give us any reason to pick the ‘human’ pattern over some other pattern, so long as that contains a period of evaluative stability. The fact that we currently see what we call adulthood as that period does not, to my mind, decide the question either way. Further, my view is that the value is to be found in having some stability, not stability at the particular stage that we have it. Some people take longer than others to settle down: emotionally, financially, socially. For some of them, this is distressing. But for others, this mode of life is attractive. Conversely, some people who settle down early may wish that they had had a longer period of uncertainty, while others settle quite happily near the start of adulthood. The idea that there is an overarching ‘human’ pattern seems to belie these differences.

¹⁰ Perhaps you think that certain features of a beach will eventually become such that they cannot be improved. If so, imagine that the improvements continually approach, but never quite reach, perfection, i.e. that the significance of the differences between each stretch become smaller as one goes on.

Finally, it is also worth noting that this argument predominantly targets RCE. Undergoing RAE simply *exacerbates* the problem, insofar as it exists at all. Since my central argument is that the concerns Agar raises are somewhat reasonable when raised against RCE, but not against RAE, we might conclude that they simply represent a further point in the case against RCE, given the attractiveness of RAE. Conversely, they do not really seem to have much force against RAE, since the problem emerges with the possibility of continual cognitive enhancement whether or not we have radically extended lives.

3. Why radical anti-ageing enhancement is different

This section suggests that, even if there is some force to the worry that RCE will lead to a kind of alienation that is opposed by the conservative bias, this worry does not extend to RAE. Agar also offers a number of moral arguments against RAE (2010, 127-132; 2013, 113-136). I will not address those here. For all I argue, then, it may be that radical enhancements that prevent us from ageing are immoral, even if they are prudentially reasonable.

In discussing RAE, Agar adopts the language of proponents such as Aubrey de Grey (2004) by talking of such enhancements as producing ‘negligible senescence’, i.e. the state not merely of slowed ageing, but the kind of radical effect on our bodies such that it is, from our perspective, as if we were not ageing at all. Agar’s worry from a prudential perspective (he also raises several moral objections, which I will not address) is based on the claim that death would be much worse for the negligibly senescent, because they would have far more to lose in dying (2010, 116-117).¹¹ He proposes that an individual’s death is bad in proportion to the good life of which it deprives them.¹² Someone who has undergone RAE will thus lose a great deal more prospective good when they die at any particular age than an unenhanced person at the same age.

This greater cost will impact our everyday life because many things we do now carry a risk of death. When I go to the cinema, for instance, there is some small risk that I will be run over as I cross the street, or that the building’s roof will collapse on me. We can calculate the expected benefit of an activity by multiplying its value if it occurs with the likelihood of it occurring, and do the same with expected cost. The expected benefit of my cinema trip following RAE stays the same, but the expected cost rises steeply because although the roof collapsing on me is no more likely, it is much worse for me if it happens. If my life is extended sufficiently, the expected cost of the trip will outweigh the expected benefit. This applies to a host of ordinary pastimes. Agar thus suggests that the negligibly senescent

¹¹ Interestingly, in his more recent work, Agar does not pursue this line of argument, instead developing the moral case. Still, since its only mention (2013, 113-114) is to refer back to the argument in Agar (2010), I will assume that Agar considers that prior discussion to be relatively complete, and that he continues to endorse it.

¹² For a more detailed explication of this kind of view, see Bradley (2009).

would become cautious shut-ins, avoiding the risks of interacting with others or engaging in common activities.

Agar suggests that the person who undergoes RAE will thus become detached from many current sources of value; though he does not explicitly link his discussion of personal relationships to RAE, it is clear that this extreme level of isolation and detachment would have a significant effect on our relationships. According to Agar's predictions, the long-lived would avoid many of the ordinary activities that, while mundane, go towards maintaining personal relationships, such as taking a "drive to the local cinema to watch a movie" (116) because of the increased costs of death, no matter how unlikely it is. They would replace these activities with safer ones, or ones that risked more gradual declines in health, such as smoking, since these will be amenable to the kinds of health interventions involved in RAE (121-122). So even if the long-lived did not change their attitude towards their loved ones directly, as in the case of RCE, Agar's description makes their relationships seem unsustainable.

As with RCE, Agar's empirical predictions are open to challenge. To borrow his example, when I take a trip to the cinema, I am simply not aware of likelihood of my dying while crossing the road, or of the roof falling in. We do not typically think about risk in the strictly rational way he discusses, and there is no reason to expect that RAE will change the cognitive biases that make us poor at assessing danger and odds, such as focusing on threats that are experientially salient, or out of our control.¹³ So it is not clear why those who have undergone lifespan enhancement will become radically more risk-averse, even if it would strictly be rational for them to do so according to Agar's theory of why death is bad.¹⁴

Agar might object that even if we are subject to biases, we are still capable of thinking about risk rationally to the extent of adopting new information. Following RAE, people will have new information about the risks involved in ordinary activities. Although there will be no new ways that these activities could go wrong, the actual risk involved is more significant, since we will have more to lose.

It is certainly true that we are capable of incorporating new information about risk, and changing our behaviour because of it. Since smoking was found to be linked to cancer, for instance, many people have given up smoking. So we cannot definitively rule out the possibility that those who undergo RAE will become more cautious. However, it is also true that many risks people are aware of do not seem to affect their behaviour all that much. As Bhattacharya and Simpson¹⁵ note, one of the features that seems to affect our tendency to incorporate risk-information into our behaviour is the probability of harm occurring. The information that was revealed when cigarettes were found to be linked to cancer was that an activity that had appeared neutral, and perhaps even beneficial, is in fact very likely to cause

¹³ See, e.g. Slovic (1987); Tversky and Kahneman (1974); Weinstein (1987).

¹⁴ Bhattacharya and Simpson (2014) offer for a similar criticism in greater detail.

¹⁵ *Ibid.*

harm. Moreover, cigarettes have come to be represented in many societies as *threatening*: we are bombarded with advertisements and medical warnings telling us about the dangers of cigarettes. So that danger is both absolutely quite likely to occur, and made socially salient.

People who undergo RAE may know on one level that their expected loss from going to the movies has increased. But the way in which this risk has increased is not of the type that is most readily incorporated in our decision-making. Unlike smoking cigarettes, going to the cinema will not become, from a subjective perspective, any more *likely* to harm you, and it thus remains a distant possibility. It will also not be the subject of advertising campaigns telling you that it is dangerous.

This is not to claim that there will definitely be no effect on people's attitudes to risk. But since Agar's argument is a *prudential* one, it is not enough to suggest that there will be just some effect. Rather, our decisions should surely be affected only if there is a risk of a *considerable* effect. Indeed, Agar's (2010, 114) explicit concern is that we risk becoming "completely dominated" by the fear of death; and it is this concern that seems most worrying in the context of personal relationships.

However, since we do not incorporate a great deal of risk-information into our decisions, it seems to me unwarranted to make such a strong claim supported only by the fact that we incorporate *some* novel information about risk. That evidence warrants only the weaker claim that there may be some effect on our decision-making; but *that* claim seems too weak to warrant the further prudential advice that we should eschew RAE. Somewhat ironically, then, it seems that Agar's prudential concern in this case is itself overly cautious.

Admittedly, if we underwent RAE in tandem with RCE, this might change things; given certain cognitive enhancements, we might become better at avoiding biased thinking, and of calculating survival odds. On the other hand, improved cognitive powers might make us aware that calculating the expected cost-benefit analysis of every activity is – if it will cause us to avoid doing things we enjoy and care about – not worth it in the long run. This is particularly pertinent if, as I argue in Section 4, those activities are connected to sources of particular value, such as our relationships.

Agar also offers some empirical extrapolation linking extensions of expected lifespan to risk aversion (118-119). But these are not well founded. For instance, he notes the fact that most modern people would not take the same risks as would, e.g. a fourteenth century peasant, and ascribes this increased risk-aversion to increases in lifespan. But another notable difference between the fourteenth and twenty-first centuries is the degree of choice available. Peasants had no choice over whether they fought for their feudal lord's army, or gave birth in dangerous conditions. It thus seems odd to say that they were 'willing' to take risks that we are not. His more modern example – the difference in the kinds of risks soldiers were willing to take in World War I as opposed to the recent war in Iraq – also involves a

multitude of differences, including the increasing professionalization of the armed forces, the control that soldiers have over the conditions in which they fight, and public awareness of those conditions.

Agar also claims that negligible senescence could involve a “twenty-year period in which life expectancies may improve by eight hundred and fifty years” (119) such that changes in risk-aversion will be much more severe than past changes, and so impact our psychology more significantly. It is certainly true that de Grey has claimed that the first person who will live to 1,000 is only twenty years younger than the first person who will reach 150, although this is not because our *lifespans* will increase dramatically in twenty years. Rather, the person who lives to 1,000 (A) will benefit from a series of incremental increases, each of which is sufficient to get them to the next step, whereas the person who dies at 150 (B) will just miss a crucial technological step (de Grey, 2004: 725).

However, Agar’s suggestion relates to life expectancies, rather than lifespans. It could be that so long as advances in lifespan are sufficiently predictable, our life *expectancies* may increase dramatically on the basis of predicted – albeit currently unavailable – technologies. This would rely on the idea that all incremental advances in lifespan will be predictable with sufficient reliability to make A’s life expectancy 1,000 even while her current lifespan is 150. There is no guarantee that this will occur; many technological advancements are not reliably predictable until quite shortly before their advent. Still, it is at least possible that advances will be predictable in the way that Agar’s suggestion requires.

Whereas I suggested that the problems with Agar’s predictions about RCE nonetheless sat atop a plausible concern, there is another, more fundamental issue with RAE. Cohen’s discussion gives us scope to see a clear difference between this case and the concerns raised in the context of RCE. Following RCE, Agar imagines the enhanced person coming to see their partner and relationship in a very different way, i.e. as trite and meaningless. In this case, the agent undergoes what I will call ‘direct’ changes in her preferences and attitudes with regard to her relationship. She no longer sees her partner as valuable (at least *qua* romantic partner). Since these attitudes partly constitute a relationship, we may say that this change is destructive of the relationship itself, or at least of its character. The specific relationship is a source of value, and so by Cohen’s lights knowingly undergoing a change that will be destructive of it is to ignore the conservative thought that we ought to preserve rather than replace sources of value.

With RAE, this does not happen even if Agar’s predictions are correct. Instead of coming to see their partner or relationship differently, the negligibly senescent person changes their view of the circumstances under which they engage with these sources of value. The negligibly senescent will be reluctant, even fearful, about physical interaction with their partner, or participating in the everyday pastimes that sustain their relationship. But this does not change how they feel about the other person. Rather than a direct change, what we see in this case is an ‘indirect’ change in attitude. She retains the

same feelings about her partner; her reluctance about the relationship involves external or circumstantial costs.

I can hopefully clarify this distinction by considering a simpler case. Petra decides that she would like to hunt Sumatran tigers. There are two ways we might dissuade her. We may point out that hunting endangered species is illegal. If this persuades her not to hunt, she might sensibly say, ‘I wish I could hunt tigers, if only it weren’t illegal’. In one sense, she no longer wants to hunt; we have persuaded her that the risk just isn’t worth it. But there is also a clear sense in which she still does want to hunt. There is nothing about hunting itself that puts her off; it is only that hunting under these circumstances is unattractive. On the other hand, we might convince her that shooting animals for kicks is cruel. But then she cannot give an analogous reply. It makes no sense for her to say ‘I wish I could go hunting, if only it didn’t involve the slaughter of animals’. That just is the practice of hunting. Now she has undergone a direct change in her attitude.

As I will suggest in the next section, an important element of this direct change is that she is no longer motivated to hunt, whereas in the case of an indirect change, she would hunt if she thought she could get away with it, and will do so if she wants it enough. I return to Cohen’s discussion to suggest that this difference is sufficiently fundamental to undermine Agar’s concerns about radical anti-ageing enhancement, even if we grant them in relation to radical cognitive enhancement.

4. Motivation

Cohen-conservatism is not only a claim about what we ought to do, but is also supposed to serve a descriptive role: Cohen (2011, 204) says that “everyone who is sane has something of this [conservative] disposition”. This suggests that there is an important difference in motivation between direct and indirect changes. Since indirect changes mean that I retain our attitudes towards an object of value (e.g. my partner and my relationship with her), I will also retain the relevant motivations to engage with them. So even if I do become more cautious as a result of RAE, I will retain strong motivations to engage in activities that support our relationships. Agar’s claim is that I will be unwilling to engage in everyday activities because I will run cost-benefit analyses and find them wanting; when the expected loss from my possible death during a cinema trip goes up, I will judge that the enjoyment I get from a film is not worth it. But this misses two important facts. First, when viewed in the context of something that maintains a valuable relationship, it is no mere trip to the cinema; its value must include in some sense the relationship that it sustains. Most people engage in some activities simply because it will make their partner happy, and sustain their relationship; there need be nothing intrinsically attractive about it

at all. So even if the expected costs outweigh the expected benefits from the film, they are less likely to outweigh the overall benefit of the trip.

Second, even if RAE does make us very cautious, it will not change our psychological needs, which for most of us include contact with others. Agar's further mistake is in assuming that each trip to the cinema will have roughly the same expected benefit. But even if I would avoid the first few trips because of the psychological mechanisms Agar suggests, later trips will take on a greater importance. We refuse opportunities for social contact all the time, in part because we know that more will come along later. When we go for some time without social contact, on the other hand, almost any opportunity begins to look attractive. So we should see that each trip will not be framed in the same way; for most people, the longer you go without social contact, the greater the expected benefit even from the same trip. Moreover, we do not only have a need for social contact in general, but for contact with specific people, who have special significance for us. As you turn down repeated trips, the cost of doing so increases with the weakening of your interpersonal ties.

Since RAE does not directly change our attitudes towards others, our psychological tendencies, or our basic needs, Cohen's observation should help us see that even if we become more cautious in the ways Agar predicts, it is unlikely to make us abandon our relationships as special sources of value because we will retain relevant strong motivations. The conservative worry, in this case, provides its own solution. Not so with RCE, as this causes direct changes in our attitudes. After RCE, you come to see your partner as not worth sticking with, and so come to see your relationship as no longer worthwhile. This means not only that you no longer value them in the same way, but also that you *lack the motivation* to overcome this obstacle. If the conservative bias is right, you have some reason to avoid RCE, but not RAE.

This distinction can help us to see some of what is at stake when we consider radical enhancement. We can distinguish between RCE and RAE by noting that the former directly changes the way we feel about certain sources of value, whereas the latter does not. One way, although not the only way, to represent this difference is to say that while radical anti-ageing enhancement might affect your preferences and even your character, radical cognitive enhancement seems more likely to affect your self-conception. Agar's subject of RAE is more fearful of engaging with his past sources of value, but they remain sources of value for him; in that sense, you can think about the subject of RAE as still recognisably you. According to Agar's predictions, the subject of RCE not only lacks motivation to engage with past sources of value, but is positively motivated to reject them. At least in terms of subjective self-conception, this makes it far harder to see the beneficiary of enhancement as the same person as the individual who is deciding whether to undergo enhancement in the first place.

5. Conclusion

Agar's concerns about the risk of radical enhancement alienating us from current sources of value are worth taking seriously. I have argued that Cohen's perspective on conservatism towards value, which shares a central strand of thought with Agar, can help us to see the limits of the latter's approach. His concerns about radical enhancement suffer from some empirical predictive problems in the case of both cognitive enhancement and anti-ageing enhancement. However, even if we grant fully his worries about radical cognitive enhancement, the same concerns do not translate to radical anti-ageing enhancement because the changes he predicts are of a fundamentally different kind. As such, a conservative bias cannot speak against radical anti-ageing enhancement in the way Agar supposes. There are, of course, other concerns relevant to whether such enhancements are permissible or even desirable, and these may be successful; but an appeal to evaluative conservatism is not such an argument.

Conflict of Interest: The author declares that they have no conflict of interest.

References

- Agar, N. (2004). *Liberal Eugenics: In Defence of Human Enhancement*. Oxford: Blackwell.
- (2010). *Humanity's End: Why We Should Reject Radical Enhancement*. Cambridge, MA: MIT Press.
- (2013) *Truly Human Enhancement: A Philosophical Defense of Limits*. Cambridge, MA: MIT Press.
- Bhattacharya, A. and Simpson, R.M. (2014). Life in Overabundance: Agar on Life-Extension and the Fear of Death. *Ethical Theory and Moral Practice*, 17(2), 223-236
- Bradley, B. (2009). *Well-Being and Death*. Oxford: Oxford University Press.
- Buchanan, A. (2012). *Better Than Human: The Promise and Perils of Enhancing Ourselves*. Oxford: Oxford University Press
- Cohen, G.A. (2011). Rescuing Conservatism: A Defence of Existing Value. In RJ Wallace, R Kumar, and S Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (pp. 203-230). Oxford: Oxford University Press.
- Cohen, G.A. (2012) *Rescuing Conservatism: A Defence of Existing Value (All Souls Version)*. In M Otsuka (Ed.), *Finding Oneself in the Other* (pp. 143-174). Princeton: Princeton University Press.

- Danaher, J. (Forthcoming). An Evaluative Conservative Case for Biomedical Enhancement. *Journal of Medical Ethics*.
- De Grey, A. (2004). Escape Velocity: Why the Prospect of Extreme Human Life Extension Matters Now. *PLoS Biology* 2(6).
- Elliott, C. (2003). *Better than Well: American Medicine Meets the American Dream*. New York: W.W. Norton & Company.
- Hauskeller, M. (2011). Human Enhancement and the Giftedness of Life. *Philosophical Papers*, 40(1), 55-79.
- (2013). *Better Humans? Understanding the Enhancement Project*. Durham: Acumen
- Kolodny, N. (2003). Love as Valuing a Relationship. *The Philosophical Review*, 112(2), 135-189.
- Levy, N. (2011) Enhancing Authenticity *Journal of Applied Philosophy* 28 (3) pp308-318
- MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory*, 3rd Edition. Indiana: University of Notre Dame Press.
- Powell, R. (2015). In *Genes We Trust: Germline Engineering, Eugenics, and the Future of the Human Genome*. *The Journal of Medicine and Philosophy*, 40(6), 669-695.
- Pugh, J., Kahane, G., and Savulescu, J. (2013). Cohen's Conservatism and Human Enhancement. *The Journal of Ethics*, 17(4), 331-354.
- Slovic, P. (1987). Perception of Risk. *Science*, 236(4799), 280-285.
- Tversky, A. and Kahneman, D. (1974). Judgement Under Uncertainty: Heuristics and Biases. *Science*, New Series, 185(4157), 1124-1131.
- Weinstein, N.D. (1987) Unrealistic Optimism About Susceptibility to Health Problems: Conclusions from a Community-Wide Sample *Journal of Behavioural Medicine*, 10(5), 481-5.