

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/281901506>

Moral Enhancement: Do Means Matter Morally, Reply to Simon Gaus

ARTICLE · AUGUST 2015

READ

1

2 AUTHORS:



Farah Focquaert

Ghent University

28 PUBLICATIONS 109 CITATIONS

SEE PROFILE



Maartje Schermer

Erasmus MC

77 PUBLICATIONS 539 CITATIONS

SEE PROFILE

Online Neuroethics Symposium on The Brains Blog

<http://philosophyofbrains.com/2015/08/25/neuroethics-symposium-on-focquaert-schermer-moral-enhancement-do-means-matter-morally.aspx>

Moral Enhancement: Do Means Matter Morally?

Reply to Simon Gaus

Farah Focquaert & Maartje Schermer (both authors contributed equally)

There is quite a lot with which we agree in the thoughtful and clear comments of Simon Gaus. We are glad that the objections and concerns he voices give us the opportunity to further clarify our position.

First, we would like to mention that we indeed refer to interventions as *deliberate* neuro-treatments or enhancements with the aim of morally enhancing an individual and not to all potential life events and/or (conscious or subliminal) experiences an individual may have throughout his/her lifetime and which may contribute to the development of his/her moral character.

Second, we indeed limit the notion of active interventions to those interventions that require an individual to actively participate and, as Gaus formulates it: *to be aware of how and what they are participating*. Such awareness might not always be as deep or profound, but the awareness is a required element. On our account, active moral enhancement interventions require an individual to practice moral behavior or appropriate emotional responses and/or engage in moral reasoning. Such active participation thus requires being aware throughout the intervention in order for someone to enhance his or her skills. Complex learning processes such as the development of moral reasoning skills require an individual's awareness and attention. We do not agree that effective moral education is a passive process. It seems a bit odd to think that enhancing one's moral reasoning skills can be the result of a by-and-large passive process. Having said this, we agree that indirect interventions can also be non-consensual and ill-formed, however, we argue that the risk of being non-consensual and/or ill-formed is greater or potentially more problematic in passive interventions compared to active interventions.

We therefore fully agree that on our notion of active interventions, 'moral enhancements that result from active interventions *just are* changes that result [at least in part] from critical reflection on how one is and how one out to be.' Passive interventions do not require the active reasoning processes or cognitive-emotional involvement on behalf of the participant and might therefore induce changes that are partially or fully concealed and thus unreflectively 'endorsed'.

Third, we find the comment on synchronic inconsistency very interesting and valuable. Self-reports from patients after deep brain stimulation for Parkinson's disease do seem to suggest that personality changes can go unnoticed and be immediately incorporated into one's self without experiencing a disruption without one's pre-operative self. Gaus suggests that even if a passive intervention radically changes your whole psychological make-up, in such a way that you achieve a new synchronic consistency (your first order desires are completely in line with your second order volitions and they are not inconsistent or conflicting) you could still notice a diachronic discontinuity by reflecting on your past behavior, decisions etc. This would of course also depend

on what the intervention would do to your biographic memory; and it could also be that the new self would denounce the former self, repress these memories, and deny any continuity, and would try to 'forget' about his former self. We believe this observation is correct, and interesting. We further believe that, first, this diachronic discontinuity can be so great and occur so abruptly that it leads to narrative disruption which can be psychologically uncomfortable, puzzling and upsetting for people. Hence, we argued that people should be warned of this beforehand and should be counseled and coached afterwards to deal with this. Also, such counseling could be a very good place and starting point for reflection and for rational endorsement. People could denounce their former self but could also try to create continuity in their narrative again – perhaps also by acts of repentance or repair.

**Active participation in one's own moral enhancement: it's either rare or worthless
Comment on "Do Means Matter Morally?"**

Simon Gaus

Humboldt-Universität zu Berlin

Farah Focquaert and Martje Schermer distinguish between *direct*, or brain-targeting, and *indirect*, or mind-targeting, means of moral enhancement. The authors' starting point is a tension: On the one hand, it seems that direct means of moral enhancement tend to be significantly more morally problematic than indirect, traditional means. On the other hand, however, mere differences in directness simply do not seem to be morally relevant.

The authors offer an elegant solution: The indirect/direct distinction, they argue, partially maps onto another contrast, namely that between moral enhancements that require *active* participation and involvement by the subject, and those that allow or force the subject to remain *passive*. And this latter contrast *is* morally relevant, the authors argue, because passive moral enhancements carry with them greater risks of reducing the enhanced person's autonomy, of leading to feelings of self-alienation and of leading to severe identity-changes that the individual does not reflectively endorse.

I am very sympathetic to the general approach. I agree that the *active/passive* contrast does seem to be far better suited for normative relevance than the *indirect/direct*-distinction. The authors' argument that many forms of direct moral enhancement threaten to create the above-mentioned problems is compelling, and implementation of safeguards along the lines the authors suggest indeed seem morally required. I worry, however, that direct means of moral enhancement are not, in general, any more likely to run into the mentioned problems, and more in need of safeguards against these problems, than indirect means. In what follows, I focus on two of the three threats the authors discuss – threats to autonomy and to the possibility of critical assessment of changes. I will argue that in both of these cases, either direct *and* indirect means of moral enhancement count as passive, or whether a means of moral enhancement is passive or active does not make a difference for the threats the authors identify. In either case, direct interventions would be no more likely to threaten autonomy and the possibility of critical assessment of changes than indirect interventions.

1. *Threats to Autonomy*

Focquaert and Schermer suggest that “[p]art of being an autonomous individual involves having the freedom and ability to withdraw from interventions and treatments that we no longer endorse at any given time” and argue that direct interventions are more likely to threaten the targeted individual's autonomy than indirect interventions. As I understand it, the authors offer two distinct arguments for this claim.

Here's what I take to be the first argument. Indirect interventions tend to require active participation by the targeted individual. Therefore, it is usually within the power of the targeted individual to reject or stop an indirect intervention at any time – simply by refusing to participate. Direct interventions, in contrast, do not tend to require a significant degree of active

participation, and so they might in principle be carried out or continued even against the explicit wishes of the individual that is being enhanced.

However, even if indirect interventions tend to require active participation, and even if the targeted individual could refuse to do what constitutes the required active participation, this power to refuse does not amount to the power to reject or stop the intervention in any interesting sense. Here's why: The power to stop the intervention, in the sense in which having that power is relevant for autonomy, is the power to intentionally stop the intervention because one has decided that one does not want it to occur. However, in order to be able to do so the individual must be *aware* of the fact that she is actively participating in an indirect intervention aimed at moral enhancement. If she is doing something that constitutes active participation, but this is not transparent to her, she cannot, in the sense relevant here, stop the intervention by refusing to participate.

Now, we could limit the notion of *active interventions* to those interventions that require the individual to actively participate *and be aware of in what and how they are participating*. But on that notion of active interventions, many or most forms of traditional indirect interventions aimed at moral enhancement would count as *passive*: moral education and providing role models, for instance, do not require and arguably typically do not involve this level of self-consciousness on the side of the targeted individuals.¹ Quite generally, indirect interventions aimed at moral enhancements seem to work best if they are not presented with the aim of morally enhancing the audience.² So if the indirect/direct contrast is to at least roughly correspond to the active/passive contrast, active interventions do not require that the targeted individuals are aware of actively participating in an intervention aimed at their moral enhancement. And if they aren't aware of participating, their power to refuse to participate does not translate to a power to reject the intervention.

I now turn to the second argument. According to Focquaert and Schermer, indirect interventions tend to be temporally extended processes whose effects occur gradually and accumulate over time. This allows the individual to give (or withhold) consent in a fine-grained and informed way: she can consent (or withhold consent) on specific effects or specific parts of the procedure, and she can do so after having become familiar with the earliest, comparatively weak effects of the first step in the procedure. Direct interventions, in contrast, tend to be less temporally extended – they might be single intervention-events – which means that the consent the individual can give is far more coarse-grained (it has the complete process as its object) and far less informed (the consent on later stages is not informed by experiences of earlier stages, because the complete consent is given in advance).

Let's first take a look at the claim that consent to indirect interventions is better-informed. The idea was that this is so because consent to indirect interventions comes in different stages, and consent to later stages is informed by the experience of earlier stages of the procedure. This presupposes, however, that (a) individuals are aware not only of gradual changes, but also of the fact that the changes are an effect of a specific indirect intervention – otherwise, they cannot link their experience of earlier stages to the question of whether to consent to continuing the intervention; and (b), that the effects of later stages will be sufficiently similar to the effects of earlier stages of the intervention – otherwise, having experienced earlier stages does not put one in a superior epistemic position with regard to consenting to later stages.

¹ Focquaert and Schermer explicitly state that “[m]ost ‘traditional’ means to attain moral enhancement, such as education, are indirect.” (6)

² Cognitive therapy – the example the authors mention in this context – might be an exception. However, even then, it surely is just that – an exception, rather than the rule.

Both of these presuppositions are, I believe, problematic. The objection against (a) is identical to my objection against the first argument – people usually do not know that they are subjects of traditional, indirect enhancements. Moreover, given that indirect interventions are extended over time and that human beings change over time even without being subject to indirect interventions, it will often be impossible to know whether certain changes are due to indirect interventions or part of what normally happens to oneself when one ages.

Let's turn to presupposition (b). (b) looks plausible for interventions that aim at influencing specific character traits or solving specific problems – for, in that case, it makes sense to think of the gradually appearing effects as, so to speak, going in the same direction (being less socially inhibited, not feeling as strong an urge to smoke, etc.). But this does not seem to be how moral education, or any other traditional, indirect means of moral enhancement, works. Later effects of moral enhancement are often not 'more of the same' in any interesting way. So, for traditional, indirect interventions aimed at moral enhancement, there is no reason to think that having experienced earlier effects of a given indirect intervention puts us in a significantly more informed position with regard to consenting to future effects of that intervention.

What remains is the idea that consent to indirect interventions is usually more fine-grained than consent to direct interventions (unless safeguards are put in place) because indirect interventions consist of various steps each of which can be the object of consent. I believe that this is true, but not particularly relevant. First off, if what I've said about the nature of ordinary forms of moral enhancement is true, traditional indirect interventions largely occur without the explicit or implicit consent of the subjects of enhancement. If that's true, the potential for more fine-grained consent is at best a hypothetical virtue. But even that is doubtful. For, holding fixed the degree of information, why should a single, irreversible decision with major consequences be any less autonomous than multiple, subsequent decisions that, in sum, have the same consequences? I see no reason to accept this. In fact, the opposite might be true: when I decide in advance about the whole package, I can at least be sure that no part of my decision is a result of the very changes that I am deciding on. This is not so if the decisions are gradual: in that case, most of those decisions will be at least partly the result of the changes that result from earlier stages of the intervention.

This concludes my discussion of Focquaert and Schermer's suggestion that direct interventions carry a higher risk of threatening a person's autonomy. To sum up: I have argued that indirect interventions are just as likely to be non-consensual and ill-informed, and that the existing differences in temporal extension do not in themselves matter for the involved person's autonomy.

2. *Narrative Identity and Reflective Endorsement*

Focquaert and Schermer suggest that direct interventions are more likely to involve identity changes that the individual herself does not reflectively endorse. This, they say, is because "... the directness of [direct interventions aimed at moral enhancement], which leaves the person a passive recipient of these changes, rather than an active agent... might lead individuals to unreflectively accept or even welcome certain traits that would not be similarly endorsed by their pre-interventional/pre-enhanced self".

As I understand it, the problem the authors point out is a *combination* of unreflective endorsement *and* conflict with the pre-interventional self: it's not just that some changes might be unreflectively endorsed, or that they would not have been endorsed by the pre-interventional self, but that they would not have been endorsed by the pre-interventional self *and* are accepted by the

post-interventional self without proper reflection. This, it seems to me, gets things exactly right: it does seem to be particularly problematic to not even critically assess changes that one's earlier self would have rejected outright.

However, I am not entirely clear on what exactly the relation between the *directness* of an intervention, the *passivity* of the person who is subject to the intervention and the subject's inclination to "unreflectively accept or welcome" the new character traits is to be. Why is it that the directness and passivity of an intervention "might lead individuals to unreflectively accept or welcome certain traits"?

On a very strong view of what it is for an intervention to be active, the answer is readily available. Suppose that an intervention aimed at moral enhancement were to count as *active* just in case the individual's moral enhancement is achieved in part by her own critical moral reflection on how she is and ought to be. On this conception of active interventions, moral enhancements that result from active interventions *just are* changes that result from critical reflection on how one is and how one ought to be. Since active interventions, so understood, necessarily involve critical reflection by the individual, active interventions cannot induce changes that are unreflectively endorsed. Passive interventions, in contrast, do not involve critical reflection by the individual, and so they *can* induce changes that are unreflectively endorsed. On this view of the active/passive contrast, active interventions possess a built-in safeguard against unreflective endorsement that passive interventions lack.

However, this notion of active intervention requires that the individual not only actively participates in the intervention, but does so with an awareness of participating in an intervention aimed at her own moral enhancement. As argued above, this is too strong: many or most traditional, indirect interventions aimed at moral enhancement – in particular moral education and role-modeling – do not usually involve this level of self-consciousness on the side of the 'patient' and would thus come out as passive. If the author's notion of the active/passive contrast is to roughly map onto the indirect/direct contrast, the demanding view of activeness is not available. Consequently, the sketched motivation for why changes induced by active interventions are more likely to be reflectively endorsed is not, either. The upshot is that there is no direct link between the degree of active participation of the individual and her noticing and critically reflecting on the changes.

In a footnote, Focquaert and Schermer sketch a different line of thought. They suggest that direct interventions tend to affect the entire brain, and that directly changing an individual's moral dispositions "will affect one's entire belief system, i.e., in such a way that the individual in question does not necessarily experience a discontinuity as such" (FN 11). I would like to hear more about how the assumed broader effects of direct interventions are to inhibit experiencing discontinuity. As I understand it, the idea seems to be that we often experience diachronic discontinuity by experiencing synchronic inconsistency – some attitudes have changed, others have not, and the newly arisen conflict between old and new helps us realize that something has changed. I find this picture interesting and attractive.

However, surely synchronic inconsistency is not our only way of noticing and experiencing diachronic discontinuity. We also have memory of past actions, speech acts and convictions, and surely those memories are left untouched by successful direct interventions aimed at moral enhancement. Granted, we typically notice changes in how we think or act only if we happen to actively remember those past actions or convictions that indicate that something has changed – not every change that we could find evidence in our memories for is immediately transparent to us. But if direct interventions really do change a person's entire system of moral beliefs, than

many or most substantial past actions will fail to conform to the person's current system of beliefs, and so any instance of actively remembering one's past actions will indicate a severe change in one's belief system. The upshot is this: While interventions that affect every part of our normative outlook might not be noticeable in virtue of any newly arisen synchronic inconsistency, this, I would think, is more than offset by the fact that if all parts of our normative outlooks are changed, we will find divergence between past behavior and current moral outlook wherever we look.

3. *Conclusion*

I've argued that passive interventions aimed at moral enhancement, in any sense of "passive" in which direct interventions tend to be passive and indirect ones active, are neither more likely to threaten the individual's autonomy nor to impair the individual's ability to notice and critically reflect on changes resulting from the intervention. If that's right, threats to autonomy and the ability to critically assess changes cannot help vindicate the impression that indirect, active means of moral enhancement are to be preferred over direct, passive means.