sciendo

Ethics & Bioethics (in Central Europe), 2021, 11 (3–4), 217–230 DOI:10.2478/ebce-2021-0016

Moral bioenhancements and the future of utilitarianism

Francisco Lara[1]

Abstract
Utilitarianism has been able to respond to many of the objections raised against it by undertaking a major revision of its theory. Basically, this consisted of recognising that its early normative propositions were only viable for agents very different from flesh-and-blood humans. They then deduced that, given human limitations, it was most useful for everyone if moral agents did not behave as utilitarians and habitually followed certain rules. Important recent advances in neurotechnology suggest that some of these human limitations can be overcome. In this article, after presenting some possible neuro-enhancements, we seek to answer the questions, first, of whether they should be accepted by a utilitarian ethic and, second, if accepted, to what extent they would invalidate the revision that allowed them to escape the objections.

Keywords: moral bioenhancement, human enhancement, utilitarianism

Utilitarianism has survived in part because it has revised its postulates with the intention of presenting itself as a realist theory. To demand strictly the best for all was to expect human beings to behave like moral saints. Thus, the utilitarians eventually came to believe that the theory should accommodate flesh-and-blood agents, who would do what was right in consequentialist terms, but within their cognitive and motivational limitations. Recognition of these limitations led them to maintain, paradoxically, that it is often in everyone's best interest for the agent not to behave like a utilitarian agent. Instead, they prescribed following deontological rules. This allowed them to respond to much of the strong criticism that had previously been levelled at them for their demanding and counter-intuitive prescriptions and, in so doing, to keep the theory afloat.

But, in the near future, this theoretical resource of utilitarianism may cease to be an effective life vest. Current neurological and biotechnological advances may make biological interventions that enhance our capacities, including those related to morality, a reality. At that point, human beings could no longer be blamed for those limitations that prevented them from fully meeting the demands of utilitarianism. If such "moral bioenhancements" were one day feasible, would they not require a thorough revision of utilitarianism to bring it back into line with an (updated) realist conception of human nature? And if that nature were to be less limited, shouldn't the theory now demand, based on its commitment to the optimum, that the agent forget the rules and do in each situation what has the best consequences for everyone? Wouldn't this entail the obligation to behave "monstrously" in the search for the common good, forgetting, if necessary, about personal rights and obligations? These are the questions I pose in this article. To do so, first, I will show how in the structural revisions of utilitarian theory the recourse to certain limitations or weaknesses of the human being has been crucial and how this has allowed the theory to survive the strong criticism that has been levelled at it. In the second section, I will review the current possibilities of using biotechnology to avoid or alleviate those human limitations that prevented the utilitarian agent from fully applying the principles of the theory. In the last section I will argue for some basic requirements that bioenhancement interventions should meet in order to be considered ethically acceptable, I will test whether the interventions necessary to enhance the utilitarian agent meet them and, if so, whether this would strip the utilitarian theory of its hitherto fruitful remedy.

[1] University of Granada (Spain); flara@ugr.es; ORCID: 0000-0002-9049-8986

Ought implies can: Human limitations as a theoretical resource

At first sight, there is something strange in an ethical theory that starts from the criterion that what is right consists of always seeking what is useful, when morality is usually conceived as a sphere of behavioural regulation in accordance with non-negotiable principles, whose fulfilment should be alien to consequential considerations. This initial impression of strangeness has been accompanied, in the academic sphere, by strong criticism of the theory, arguing that from the utilitarian conception of what is right it is impossible to respect, to an adequate degree, both personal autonomy and duties, as well as individual rights.

On the one hand, the consequentialist scheme that underlies this utilitarian conception of what is right,[2] due to its neutral and global perspective, in which there is no place for what is permissible because everything that is not optimal is forbidden, prevents us from assigning preference to our commitments, projects and loved ones. Devoting time and resources to oneself or one's kin at the expense of the general good would constitute an unacceptable licence for any consequentialist agent (Williams, 1973, pp. 116–117; Railton, 1984). On the other hand, if the ultimate goal is to achieve the best outcomes, any action that allows us to do so would be justified, including those that do not respect important individual rights. At the time, this possibility aroused the most gruesome imagination of some critics. Utilitarians were accused of allowing, because of their consequentialist theory of rightness, such heinous and ruthless actions as falsely incriminating an innocent person in order to prevent riots and lynchings (McCloskey, 1957; Foot, 1967; Nielsen, 1972); blowing up the obese person blocking the mouth of the cave as the only way to prevent his companions from drowning when the tide comes in (Nielsen, 1972); persuading a frightened old woman by twisting her grandson's arm to hand over car keys to take injured people to hospital (Nagel, 1986, p. 176); or even dismembering someone in order to save several sick people with their organs (Foot, 1967; Thomson, 1976). [3]

It is not surprising, then, that this consequentialist conception of moral obligation as something empirical and relative and the consequent criticism of the counter-intuitiveness of its normative implications led to predictions about the short future of the theory. Thus, Bernard Williams (1973, p. 150), for example, already maintained in the early seventies that the day could not be far off when we would no longer hear of utilitarianism.

But such predictions have not been accurate. The extensive literature and specialised community to which utilitarianism has given rise shows that it continues to be one of the protagonists of the current ethical debate.[4] What is the reason for this survival of utilitarianism, contrary to the understandable predictions?

A first explanation could come from the same consequentialist scheme that was used by many to criticise it. If everything is a question of achieving the best results, any morally difficult or dilemmatic situation can be resolved by doing the least bad thing. To the attraction, for some, of this pragmatism, we must add the fact that the theory incites, more than others, activism and social conscience. In contrast to deontologism, which is based on what is prohibited, consequentialism opts for what is obligatory, for a broad conception of moral responsibility, for which all kinds of consequences (immediate and long-term, direct and indirect) deriving from both actions and omissions count.

---

[2] Utilitarianism is basically composed of two theoretical components. A welfare conception of what is good and a consequentialist conception of what is right. Thus, if for the latter, our obligation is to do that which brings about the best results for all, considered impersonally, for its conception of the good, the value of the results must be measured in terms of welfare.

[3] On utilitarianism's inability to take individual rights seriously, see also Fried (1978, pp. 81–105), Glover (1977, pp. 73–75), McCloskey (1957) and Williams (1973, pp. 97–100).

[4] Proof of this are the expert societies to which this theory has given rise, among which the International Society for Utilitarian Studies and the Ibero-American Society for Utilitarian Studies stand out. In addition, these societies publish the specialised journals *Utilitas* and *Telos*, respectively.

However, it is evident that this resolutive and mobilising character of consequentialism does not explain by itself the survival of utilitarianism, since, as I said above, the criticism of utilitarianism is derived from it as well. What will allow utilitarianism to really free itself from criticism are certain structural modifications in its theory. The most significant of these is a differentiation between the criterion of moral correctness and the decision-making procedure we derive from it. This allows one, in principle, to maintain that, even when the criterion of correctness tells us that the right thing to do is to behave optimally, this does not mean that we always have to decide what to do by calculating the consequences of possible actions.[5]

Now, in order not to depart from the consequentialist scheme, this permission not to calculate at all times must have an empirical foundation, it must be justified by virtue of the limitations. One of these limitations is purely cognitive because it is often not within our reach to know what the consequences of our choices will be. But there are also motivational limitations. It is common for us to interpret data in a self-interested way. Thus, for example, we are prone to believe, contrary to available information, that acting in our own interest, or in accordance with some cultural prejudice, coincides with what is best overall. In other cases, emotions prevent us from appreciating what is right or weaken our will to be governed by it.

Therefore, if by our limited nature, we cannot behave as agents who always choose the optimal action, it makes no sense to defend strict consequentialism as a decision-making procedure, even if our normative criterion is still to do what is optimal. It will often be best for the agent to be guided by consequentialist evaluations not of the particular action options before him, but of the general patterns of behaviour - rules, predispositions, virtues, institutions, etc. - to which those behavioural options would be subsumed. Thus, if one considers that, for example, predispositions or virtues are the factor to be evaluated in terms of consequences, one should not calculate at each moment which particular action has the best results; one should only perform that action which conforms to that character trait (loyalty, honesty, aversion to inflicting pain, gratitude, spontaneity, etc.) which, if regularly followed by all, produces the best results (Driver, 2001; Hurka, 2001; Jamieson, 2007; Bradley, 2005).

This strategy of deciding by evaluating the consequences of general behavioural patterns, however, poses the difficult challenge of articulating the regular following of such patterns with the consequentialist spirit that must infuse them. How can we harmonise the defence of non-contingency that comes with following internalised rules or predispositions without considering the consequences of doing so in each situation, with the consequentialist non-complacency that demands always being attentive to changing circumstances in order to achieve the optimum? What about those situations in which, considering all relevant aspects, we very reliably believe that we will get the best result by doing the opposite of what our rules and predispositions prescribe?

The most successful way to meet this challenge to consequentialism is by resorting to a two-level ethical theory of deliberation. On one basic level, the agent would be governed by these general guidelines of conduct, considered as if they were unrenounceable; and on the other, critical, the agent would consequentially justify both the existence of such guidelines and the possible exceptions to following them (Hare, 1981). Thanks to this theoretical resource, it can be argued that, even if we normally behave according to general guidelines, we do not lose coherence by accepting that such guidelines cannot always govern our decision about what we

[5] This differentiation of functions can, in principle, be defended by any normative ethics that seeks to implement its abstract principles. See Feldman (2012, p. 151). However, it has been the utilitarians who have explored this avenue the most, especially since Bales (1971). This has been done, for example, by Adams (1976), Ellis (1981), Railton (1984, pp. 100–106, 113–117), Parfit (1984, pp. 24–29), and Pettit (1986, p. 194). It should be noted, however, that even in a less explicit way, this differentiation of functions has been present in the utilitarian tradition since its origins. See, for example, Bentham (1789), chap. IV, sect. VI, Mill (1861, p. 64), and Sidgwick (1907, p. 413).

should do. And they will not govern it in exceptional situations where, either because of some control of the main decisional factors, or because of the very bad conditions, we can, in consequentialist terms and with probable success, resolve conflicts between guidelines or choose new guidelines for new stages. In such situations we would be making decisions from the critical level.

With this modification of the theory, utilitarian consequentialism can now respond to criticism. On the one hand, it can justify special obligations towards relatives by pointing out the best consequences of not calculating at all times whether or not to respect them. For example, it can appeal to the happiness and lower psychological cost of normatively directing people according to a predisposition, in line with the natural inclination of parents to put the interests of their children before the general welfare (Hare, 1981, pp. 136–137). Moreover, if they are not required to constantly calculate their relationships with relatives, it will be easier to maintain a socially beneficial institution, the family. It would be difficult to maintain ties, for example, between parents and children if the latter realised that the former were caring for them purely in the general interest, not because they felt a predilection for them.

The same would be true of respect for individual rights. Internalising this respect has better consequences than constantly calculating its desirability. Firstly, because due to our cognitive limitations, one can hardly be sure that one is in that exceptional situation in which it would really be advantageous to violate rights. For in order to undertake something normally as harmful as, for example, killing, torturing or enslaving an innocent person, one has to have very strong, and unusual, evidence that doing so will ultimately be more advantageous.

Therefore, in the face of that usual lack of certainty, it will almost always be better to do what is usually in everyone's interest, in this case, to respect rights. And it is usually best for all because, in addition to protecting its direct beneficiaries, the prescription of such respect would allow for sufficient trust among agents to enable cooperation and, ultimately, the very existence of society (Harsanyi, 1977, pp. 12–16; Hodgson, 1967, c. 2; Brandt, 1979, pp. 271–277; Johnson, 1991, c. 3, 4, 9; Pettit, 1986, pp. 450–451; 1988, pp. 51–53). Only if rights, such as to life or property, are guaranteed outright, is it in the interest of agents to covenant with others and to honour the covenant, despite the personal sacrifices that the covenant entails.

In short, there are consequential reasons to demand not to calculate normally but to respect obligations and rights, and even for these rules to be internalised in such a way that people are reluctant to violate them, except in exceptional situations. But these consequentialist reasons for not behaving as a consequentialist are only grounded by virtue of the aforementioned cognitive limitations and that natural tendency to bias, which leads agents to put their own self-interest and sentiment before the common good. It is the recognition of these limitations and their normative implications that has allowed utilitarianism to be revised in order to survive. But what would happen if advances in neuroscience and biotechnology were to overcome these limitations, and should utilitarianism demand that such advances be used to morally improve human beings? Would the theory then be forced to return to its original defence of an extremely calculating moral agent?

Possible bioenhancements of the utilitarian agent

The first thing to do would be to ask about those biotechnological advances that could modify individuals to make them better consequentialist agents. To do so, I will elaborate on the limitations that served to justify the well-worn two-level theory, and then outline the neurological interventions available to us now or in the short term to address these limitations.

Enhancing our ability to foresee the consequences of our actions

In principle, utilitarianism, because of its consequentialist conception of what is right, is a more fallible theory than others. Although we can sometimes anticipate the effects of our actions, we

will never be certain of them. Moreover, the more distant and global the effects, the less certain. Thus, an act that seems to lead to good results can always lead to disaster. It is true that, after all, life is full of risks and uncertainties and, even so, we do not stand still, we do things on the basis of more or less reasonable forecasts. But such forecasts based on uncertain calculations, which may serve to make prudential decisions, will be of little help to a behavioural domain in which the right thing is intended. Can someone be said to have done the right thing when doing what *seems* optimal produces lousy results? From a consequentialist perspective, can this be said of the individual who, in 1938, threw himself into the river to save a stranger who later turned out to be Adolf Hitler? (Smart, 1973, p. 59).

The above-mentioned distinction between the criterion of moral rightness and the decision-making procedure is also very useful to get out of this impasse. For, even if the right thing to do is still only that action whose consequences *actually* maximise the good, the way of deciding may be designed in accordance with human limitations to require only that action which conforms to *the probable expected consequences, in the short or medium term*. In the words of Ingmar Persson (2008), it would be something like distinguishing between what we should do, maximising actual value, and what we should try to do, maximising expected value.[6]

Even so, this does not exempt the agent from *trying* to maximise expected value by acquiring as much certainty as possible about the consequences of his choices, using all the means at his disposal. He must therefore make a mental effort and inform himself "externally" as much as he can; but if he can improve himself "internally" to improve the degree of certainty of his deliberations, he should do so as well. What does biotechnology offer today for this purpose?

Among the smart drugs or nootropics, Modafinil (Provigil) stands out, as it appears to have neither side effects nor chemical addiction (Myrick et al., 2004). It is a neurostimulant commonly used for sleep disorders and attention deficits and has recently been shown to improve working memory by inhibiting rapid adaptive response. It allows for an equally rapid, but much deeper response due to better use of certain areas of memory (Sanderg, 2011, p. 74). This tremendous mental functionality and its effects in costlessly extending wakefulness and attention make this drug an ideal tool for making better consequential decisions in complicated situations that require sustained concentration and, at the same time, lucidity in the use of memorised information. An alternative technology that is also showing good results in terms of possible cognitive enhancement with respect to consequential foresight is brain stimulation. Studies show how transcranial magnetic stimulation, in particular, could increase or decrease the excitability of the cortex, thus changing its plasticity levels, and thus improving, among other things, working memory (Fregni et al., 2005).

Strengthening the will with greater impulse control

Doing what one believes to be right is not always easy. This phenomenon was known to ancient philosophers as weakness of the will (*akrasia*). This human limitation was also among those that were adduced to justify why human beings cannot be asked to behave as strictly consequentialist agents. However, insofar as this limitation sometimes has a biological origin, it could be partly rectifiable. The weakness of the will often responds to a dysfunction similar to that suffered by people suffering from some kind of addiction. I refer to the inability to delay gratification. When these people find immediate incentives to perform an action, they lose sight of the reasons, supported by indirect and long-term incentives, against performing that action. Recently this thesis has been confirmed by physiological studies of even more impulsive

[6] This possible solution to the fallibility problem of consequentialism, integrating actual and probable outcomes, would settle a long-standing internal debate between those who, following Mill and Bentham, favoured a probabilistic or subjective consequentialism (Hudson, 1989; Jackson, 1991; Howard-Snyder, 2005), and those who thought that what is right is what, in fact, objectively, produces the best outcomes (Bales, 1971; Brink, 1986; Railton, 1984; Smith, 1988).

individuals, such as psychopaths. They show both increased activity in the ventral striatum region of the brain, which is responsible for our inclination towards immediate rewards, and a reduced connection of this area with the prefrontal cortex, which is associated with future-focused decisions (Hosking et al., 2017). But impulses may have a deeper physiological cause, which in addition to explaining addictive or unrestrainable behaviour, better accounts for weakness of will. They may be due to an inadequate functioning of the serotonergic transmission system. When this happens, the orbitofrontal cortex of the brain, the area where emotions are regulated, receives inadequate doses of serotonin and, as a consequence, impulses can sometimes not be controlled, nor emotional reactions to provocation or temptation regulated. Therefore, if a person with this dysfunction were to take a selective serotonin reuptake inhibitor (SSRI), a substance used to combat certain psychological disorders, they could slow down the uptake of serotonin (also balancing dopamine and noradrenaline) and thus control their impulses and strengthen their will to act according to those intentions and reasons that are more consistent with their values and the relevant data (Stahl, 2006).

Counteracting the tendency towards self-interest

Many actors do not pursue the best consequences for all concerned because this entails a significant personal effort and/or a sacrifice of self-interest for the common good. But this could also be remedied, in part, by neurology; in particular, by the discoveries made about the intervention in the nervous system of the oxytocin, a hormone that also functions as a neurotransmitter and is partly responsible for the predisposition of some mammals to mate bonding and offspring care (Hasting et al., 2014). It has been artificially synthesised and has been shown to enhance certain empathic abilities important for perspective-taking in humans when administered nasally or intravenously. Thus, for example, it facilitates the identification of emotional states when looking at photos that only show the surroundings of the eyes (Domes et al., 2007), makes the subject look more closely into the eyes of the other (Guastella et al., 2008), improves the recognition of positive emotional expressions (Marsh et al., 2010), or helps autistic people to better understand affective language (Hollander et al., 2007). Furthermore, scientific evidence shows that the involvement of this substance in neural systems can explain the altruistic motivation to avoid the suffering of others (Bartels et al., 2004; Insel & Fernald, 2004; Dolen et al., 2013), and that when administered to humans, they sacrifice more for others and become more trusting, reciprocal and generous (Kosfeld et al. 2005; Reyes & Mateo, 2008; Morhenn et al., 2008; Rodrígues et al., 2009; Barraza 2010; Zak et al., 2004; 2005; 2007). In enhancing sociability and promoting cooperation, it has been shown that SSRIs, by increasing serotonin levels, may also have an important effect (Wood et al., 2006; Tse & Bond, 2002).

Ethical requirements for bioenhancements

It follows from the above that we already have biotechnological resources to alleviate, in part, the biological limitations that prevented individuals from behaving as full utilitarian agents. But does that mean that we should use such resources? I will argue in what follows that, from an ethical point of view in general, and from a utilitarian point of view in particular, at least the following two requirements should be met to answer this question in the affirmative.

1. *The principle of security*

Despite the theoretically positive effects of the above interventions, they could be risky for health because of possible side effects. It should be borne in mind that most of the neurotransmitters or brain areas affected by the stimulation substances or techniques normally serve many functions. Thus, for example, serotonin, in addition to having effects on social behaviour, is involved in other processes such as learning, vision, sexual behaviour, sleep, appetite, pain or memory, and may cause unintended changes (Crockett, 2014).

In addition, other adverse effects may occur if we are not able to modulate its application according to differences between individuals and different application contexts. Thus, oxytocin has been shown to have positive effects on empathy only in certain contexts (Akitsuki & Decety, 2009; Bartz et al., 2011; Bos et al., 2012). Thus, for example, such effects would occur when the individual needs to interact with acquaintances, trusted ones, or family members, but would seldom appear in situations of competition (Shamay-Tsoory et al., 2009), uncertainty (Declerck et al., 2010), institutional inefficiency (Zak, 2008) and interaction with strangers (De Dreu et al., 2010). Similarly, the empathic effectiveness of oxytocin appears to depend heavily on the peculiarities of the individual. Thus, its effects would be meagre or negative in the social sphere for those who have less ability to put themselves in the place of others (Abu-Akel et al., 2015); for those who, whether for genetic reasons (Rodrigues et al., 2009) or for manifested behaviour (Barraza, 2010) are less willing to show empathy; for those who are the most socially adept (Bartz et al., 2010); for those who have been brought up with less parental care (Carter, 2003); for persons with aggressive tendencies (DeWall et al., 2013); or simply for men rather than women (Hurlemann et al., 2010).

Given these difficulties in implementing biotechnological interventions in humans and the possible negative welfare implications of not taking them into account, it should be a first ethical-utilitarian criterion that bioenhancements should not be implemented until there is a high degree of certainty about their efficiency and safety.

2. *The requirement of open intervention*

But it is not enough for biotechnological interventions to be safe. They should also be "open" in the sense that they aim to directly modify moral skills, but not character traits, motivations or behavioural patterns (Shaw, 2014; Schaefer, 2015; Earp et al., 2018). In other words, changes in the nature of the agent cannot involve imposing substantive values on the subject by inducing ways of thinking and feeling that are clearly determined by those values. On the contrary, they should only aim at giving the subject mental tools with which to critically determine his or her own values.

What are the basic reasons for requiring bioenhancements to be open? First of all, it is a matter of simple operationality. A "closed" bioenhancement, in which the aim is to modify directly, and in isolation, certain aspects of the moral personality of human beings, is very unlikely to increase their morality. There would be no point in changing certain values or attitudes if these are not controlled or modulated by higher-order, usually deliberative, capacities that allow the agent to flexibly apply such tendencies in response to relevant reasons and the context in which they are used (Earp et al., 2018). This is similar to what Focquqert and Schermer (2015) argue, for whom it is not enough to improve the emotions that lead to better actions if this does not correspond to a proper functioning of moral reasoning. Harris (2011) gives an example to show how biotechnologically diminishing aggression might not be morally desirable. He asks us to imagine the situation of an individual attacking a terrorist who is about to bomb a plane. If that individual had previously undergone an SSRI intervention to reduce his aggression, he would probably not have been able to save the passengers. With this example, Harris wants to underline that true moral behaviour is a complex phenomenon that must prepare you to act correctly in different scenarios and that for this to happen, it is important to be able to reason morally. Thus, pretending for its own sake to reduce the force of our impulses, as we suggested he could do with an agent in order to make him more strictly consequentialist, might incapacitate the agent to act decisively, and even unreflectively, in situations that (rationally) demand such a course of action. The same would be true of empathy, which, as we saw above, is context- and individual-related and, to be morally effective, would also require modulation in its use (Earp et al., 2018, pp. 169–171).

But it is not only a question of interventions that aim to directly modify behaviour and attitudes being unsuccessful for moral enhancement. They are also inconsistent with their objective. That is, a bioenhancement that claims to be moral cannot be closed. Even if it were ultimately successful, making the enhanced individual's behaviour conform to moral standards, the intervention would be more an example of social control than a moral enhancement in the strict sense of the word, because it would be carried out with the aim of directly changing their behaviour and not their capabilities. We would have made individuals behave better morally, but not because they had chosen to do so, but because someone or something, the enhancer or the technology in question, has made them behave that way. The mistake would be in pretending to morally enhance individuals by short-circuiting something that is essential to morality: freedom and behaviour in accordance with self-assumed reasons (Schaefer, 2015; Focquaert & Schermer, 2015, pp. 145ff).

Moreover, closed enhancements would also go against this autonomous and critical conception of morality because they would undermine a crucial element for the development of personal autonomy: dissent. If closed enhancements are based on the assumption that certain attitudes or behavioural patterns, together with the values that underlie them, are morally correct, it is to be expected that these same attitudes and patterns will be the ones that are tried to be implanted in all the subjects to be enhanced. This would lead to a homogenisation of people (Schaefer, 2015) and a consequent lack of reference points for individuals to contrast their views, thus making moral progress, understood in fundamentally critical terms, considerably more difficult.

These reasons for preferring that interventions should always be open are compatible with a certain type of utilitarianism. The origin of this would be in J. S. Mill's (1859; 1861) defence of the ideal utilitarian agent, who, from the greatest possible autonomy, ends up identifying the collective good with his personal good. In its more modern version, utilitarianism is interested in personal autonomy from a conception of well-being dependent on human freedom to choose. The ultimate good cannot consist, for this perspective, in mere pleasurable enjoyment or in a satisfaction of interests conditioned by an imposed situation. The subjective or experiential aspects of the subject must be complemented by the demand for objective capacities to autonomously choose one's own way of life (Sen, 1985; 2009).

In short, I believe that there are important reasons, even from a utilitarian perspective, to demand that moral enhancement interventions should only be carried out if, in addition to being safe, they do not infringe on the autonomous capacity to deliberate. We must now ask ourselves whether the aforementioned bioenhancement interventions, which could alleviate those human deficiencies that led to the revision of utilitarianism, would fulfil this requirement.

On the one hand, enhancing cognitive capacities to increase the degree of certainty about the probable consequences of our actions could be defended as an open intervention. The more efficient use of our memory does not in itself entail a change of personal identity in any particular direction. It is simply a matter of enhancing skills with which to make better judgements when making moral decisions in which the interests of all concerned are at stake.

The other two possible bioenhancements would, however, be more controversial. Changes in impulsivity and, above all, in a person's capacity for empathy could be seen as closed enhancements. They could lead to a substantive change in the values, identity traits and thus motivational disposition of the person being treated. For a person who is intervened to be less impulsive, or another who is treated to be more empathetic, is likely to end up being something he or she was not before: more calculating or more altruistic, respectively.

However, changes in such attitudes could also be directed so that the agent changes certain skills that, having much to do with morality, affect understanding rather than motivation. Thus, by increasing the subject's willpower so that, by not allowing himself to be so easily led by temptations, he can be consistent in his actions with his consequential reasons, what we would

ultimately achieve is to make him more autonomous. The intervention would then be understood as an aid so that, by controlling his emotions, the agent can harmonise more what he feels with what he thinks, so that, in short, he can be freer and more reflective in his behaviour. In short, he would be freed from his subjection to the arbitrariness of emotions.

On the other hand, if by enhancing empathy, we give the agent the option of better perceiving the harm that his actions would cause, either by putting himself in the place of the other or by seeing him from a position of fairness, we would be making it easier for him to judge the consequences of our decisions on the well-being of all.[7]

The decisive question would then be whether it is possible to establish any criteria to differentiate when these bioenhancements in impulsivity and empathy involve a substantive change in values or motivation (closed intervention) and when they only increase the capacities for understanding and deliberation (open intervention). Vincent (2011) does not believe that this is possible. He suspects that such bioenhancements, in one way or another, ultimately always involve some modification of personal identity, of what is authentically constitutive of a person. However, Vincents scepticism only makes sense if we start from an essentialist conception of identity. It is possible, however, to understand personal authenticity as something under construction or self-discovery. If we understand it in this way, the reduction of impulsivity, for example, could be seen as a liberation from impediments to becoming more authentic. Indeed, as Bolt and Schermer (2009) have shown, it is this sense of liberation and personal authenticity that is experienced by ADHD patients taking Ritalin. Similarly, it can be assumed that the increased empathy would allow modified individuals to have important information to redirect their own lives. By making it easier to put ourselves in the shoes of others, we could learn more about how others see us and, with this information, enhance our understanding of ourselves, which is essential for discovering or building our authentic selves.

However, this does not mean that some of these interventions cannot be an attack on personal identity, even when, in line with the defence of autonomy postulated here, we understand it as self-discovery or self-creation. It is a matter of degree. A constant and/or profound influence on impulsivity or empathy could be the cause of a motivational or attitudinal change that is not, in itself, open to different values or ways of life; it could mean, ultimately, the imposition of an identity not determined by the agent himself. It seems to me that the key here is to keep in mind a distinction between two tiers in the human deliberation. In the first would be the desires, interests, knowledge or predispositions that one has before one has meditated on them. These elements are then interpreted, in a second tier, with flexibility and rationality, with the agent conforming to them, or not, according to reasons based on their contextual applicability, but also on prudential or constitutive aspects of the agent's own personality.

Bearing this distinction of deliberative tiers in mind, the determining factor for ethical approval of a bioenhancement will be the degree to which the values of the first instance are modified. If the modification is so constant or profound that it alters the idiosyncratic principles of the agent, we would be limiting or even annulling this second tier of reflection, which is so decisive for personal autonomy and critical spirit. In this case, the intervention would be unethical because we would be imposing on the subject an evaluative perspective in the assimilation of which he or she has not participated.

[7] In the field of human enhancement, there are positions that seek to change the capacities of human beings to the point of placing them on a higher ontological plane. In one of the most recent critiques of these transhumanist positions, S.B. Levin (2021) calls for a non-biotechnological progress of the human being based on the acquisition of a holistic and virtuous perspective of the good life. The requirement of open intervention would, however, imply a commitment to a non-transhumanist justification of moral bioenhancements. It would be a guarantee that these interventions never exceed the limits imposed by our basic conception of human nature. Indeed, bioenhancement might even enhance human nature. It would make us more autonomous and thus more disposed to the holistic and virtuous vision claimed by Levin.

Towards a new revision of utilitarianism?

It follows from the above that the three bioenhancement interventions considered, provided they are safe, could, in principle, be open and therefore acceptable from an ethical, as well as utilitarian, point of view. Even those more related to motivation, such as impulse control or empathy, can be carried out to a degree in which personal autonomy and integrity are enhanced. We must now ask whether these acceptable bioenhancement interventions would invalidate the appeal to human limitations that led utilitarianism to a two-level decisional theory and thus to escape the strong criticism levelled against it.

Presumably they would not invalidate it. For two reasons. First, because ethically permissible enhancements would not turn the enhancing agent into a "moral saint", excellently calculating what is right and behaving accordingly. And, secondly, as a consequence of the first, because the theory, not having excellent moral agents as targets, has no other structural option than the two-level decisional theory. Let us look at it.

Why do I argue, on the one hand, that the so-called bioenhancements will not turn humans into excellent consequentialist moral agents? First, because the cognitive enhancements, although significant, would be insufficient. The intervened individual would have an enhanced knowledge of the consequences of his possible acts, but he would not have it all his own way. His progress would be in terms of a more versatile and extensive memory, less intellectual fatigue and increased concentration. It is true that this would allow him to increase his degree of prediction of the consequences of his actions, but he would still be far from omniscient. Not having a crystal ball, there will still be unintended consequences that can turn your claim to do the right thing into an unexpectedly negative event.

Second, enhancing agents would not be moral saints because requiring enhancements to be overt would not allow for such a significant alteration of motivation as to ensure that the agent identifies with the utilitarian agenda. For example, the impulse control achieved could not, by itself, in any case, amount to a substantive alteration of identity. Therefore, for those who are emotionally determined in their way of being, after the intervention, there would still be the possibility that, having to choose between what they know to be right and what they desire, they would let themselves be led by the latter. They would progress with this and other open interventions in that they would have a greater (consequential and cognitively empathic) knowledge of what is right and that their final decision would not be driven by strong impulses or blind immediate gratification. However, there will always be the possibility that individuals will put their own interests before those of all concerned.

On the other hand, why would these partial advances in moral agency not be sufficient to dispense with the decisional levels theory of utilitarianism revisited? Why would it still be in everyone's best interest for consequentialist agents to behave as if the (useful) behavioural patterns had no exceptions (even though they know that the patterns may have exceptions if they switch to the other level)? I understand that this would still be a necessity because the target agent of the theory, though morally enhanced, remains incapable of knowing and wanting to do what, *at any given moment*, has the best consequences for all. It is true that the enhanced agent, in comparison with the non-enhanced agent, can design their behavioural patterns in a way that is more consistent with consequentialism, introducing more exceptions in those patterns, with more possible situations in which they have the licence to change at the critical level. Their greater knowledge of consequences will allow them to do this without much psychological cost. But this greater agential versatility will also have its significant limits. It will remain true that for agents to learn and use behavioural guidelines easily, these cannot contain so many exceptions in their formulation, nor be so open to revision. Moreover, keep in mind that such guidelines, in order to be followed with conviction as if they admit of no exceptions, should be assumed by the enhanced agent as defining his or her personality, even

to the point of generating remorse in case of non-compliance. To this end, these must be firm and general, without being habitually open to revision.

In conclusion, given these persistent human limitations, it is foreseeable that, from a utilitarian perspective, it is still better that new utilitarian agents, enhanced by safe and open interventions, behave according to a two-level decisional scheme. Presumably, however, bioenhancements would allow for a higher frequency of occasions when pattern revision should be considered and a higher rate of success, from a consequentialist perspective, both in such revisions and in resolving conflict between patterns.

Acknowledgement
This article was written as a part of the research project *Digital Ethics. Moral Enhancement through an Interactive Use of Artificial Intelligence* (PID2019-104943RB-I00), funded by the State Research Agency of the Spanish Government (AEI/10.13039/501100011033).

References

ABU-AKEL, A. et al. (2015): Oxytocin increases empathy to pain when adopting the other – but not the self-perspective. In: *Social Neuroscience*, 10(1), pp. 7–15.
ADAMS, R. M. (1976): Motive utilitarianism. In: *Journal of Philosophy*, 73(14), pp. 467–481.
AKITSUKI, Y. & DECETY, J. (2009): Social context and perceived agency affects empathy for pain: An event-related fMRI investigation. In: *Neuroimage,* 47(2), pp. 722–734.
BALES, R. E. (1971): Act-utilitarianism: Account of right making characteristics or decision-making procedure? In: *American Philosophical Quarterly*, 8(3), pp. 257–265.
BARRAZA, J. (2010): *The physiology of empathy: Living oxytocin to empathic responding*, Dissertation. Claremont Graduate University, Proquest.
BARTELS, A. et al. (2004): The neural correlates of maternal and romantic love. In: *Neuroimage*, 21(3), pp. 1155–1166.
BARTZ, J. et al. (2010): Oxytocin selectively improves empathic accuracy. In: *Psychological Science,* 21(10), pp. 1426–1428.
BARTZ, J. et al. (2011): Social effects of oxytocin in humans: context and person matter. In: *Trends in Cognitive Sciences,* 15(7), pp. 301–309.
BENTHAM, J. (1789/1961): *An introduction to the principles of morals and legislation*. New York: Doubleday.
BOLT, I. & SCHERNER, M. (2009): Psychopharmaceutical enhancers: Enhancing identity? In: *Neuroethics*, 2, pp. 103–111.
BOS, P. et al. (2012): Acute effects of steroid hormones and neuropeptides on human social-emotional behaviour: A review of single administration studies. In: *Frontiers in Neuroendocrinology*, 33(1), pp. 17–35.
BRADLEY, B. (2005): Virtue consequentialism. In: *Utilitas*, 17(3), pp. 282–298.
BRANDT, R. (1979): A *theory of the good and the right*. Oxford: Oxford University Press.
BRINK, D. O. (1986): Utilitarian morality and the personal point of view. In: *The Journal of Philosophy*, 83(8), pp. 417–438.
CARTER, C. (2003): Developmental consequences of oxytocin. In: *Physiology and Behaviour,* 79(3), pp. 383–397.
CROCKETT, M. J. (2014): Moral bioenhancement: Neuroscientific perspective. In: *Journal of Medical Ethics*, 40(6), pp. 370–371.
DE DREU, C. et al. (2010): The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. In: *Science,* 328(5984), pp. 1408–1411.
DECLERCK, C. et al. (2010): Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. In: *Hormones and Behaviour*, 57(3), pp. 368–374.

DeWALL, C. et al. (2013): When the love hormone leads to violence: Oxytocin increases intimate partner violence inclinations among high trait aggressive people. In: *Social Psychological and Personality Science,* 5(6), pp. 691–697.
DOLEN, G. et al. (2013): Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. In: *Nature,* 501(7466), pp. 179–184.
DOMES, G. et al. (2007): Oxytocin improves 'mind-reading' in humans. In: *Biological Psychiatry,* 61(6), pp. 731–733.
DRIVER, J. (2001): *Uneasy virtue*. Cambridge: Cambridge University Press.
EARP, B., DOUGLAS, T. & SAVULESCU, J. (2018): Moral enhancement. In: L. Johnson & K. Rommenfanger (eds.): *The Routledge handbook of neuroethics*. London: Routledge, pp. 166–184.
ELLIS, B. (1981): Retrospective and prospective utilitarianism. In: *Nous*, 15(3), pp. 325-339.
FELDMAN, F. (2012): True and useful: On the structure of a two-level normative theory. In: *Utilitas*, 24(2), pp. 151–171.
FOCQUAERT, F. & SCHERMER, M. (2015): Moral enhancement: Do means matter morally? In: *Neuroethics*, 8(2), pp. 139–151.
FOOT, P. (1967): The problem of abortion and the doctrine of the double effect. In: *Oxford Review*, 5, pp. 5–15.
FREGNI, F. et al. (2005): Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. In: *Experimental Brain Research*, 166(1), pp. 23–30.
FRIED, C. (1978): *Right and wrong*. Cambridge, MA: Harvard University Press.
GLOVER, J. (1977): *Causing death and saving lives*. London: Penguin Books.
GUASTELLA, A. J. et al. (2008): Oxytocin increases gaze to the eye region of human faces. In: *Biological Psychiatry,* 63(1), pp. 3–5.
HARE, R. M. (1981): *Moral thinking*: *Its levels, method and point*. Oxford: Clarendon Press.
HARRIS, J. (2011): Moral enhancement and freedom. In: *Bioethics*, 25(2), pp. 102–111.
HARSANYI, J. C. (1977/1978): Rule utilitarianism and decision theory. In: H. W. Gottinger & W. Leinfellner (eds.): *Decision Theory and Social Ethics*: *Issues in Social Choice*. Dordrecht: D. Reidel Publishing Company, pp. 3–31.
HASTINGS, P. D. et al. (2014): The neurobiological bases of empathic concern for others. In: K. Killen & J. Smetana (eds.): *Handbook of moral development*. Hove, East Sussex: Psychology Press, pp. 411–434.
HODGSON, D. H. (1967): *Consequences of utilitarianism*. Oxford: Clarendon Press.
HOLLANDER, E. et al. (2007): Oxytocin increases retention of social cognition in autism. In: *Biological Psychiatry,* 61(4), pp. 498–503.
HOSKING, J. et al. (2017): Disrupted prefrontal regulation of striatal subjective value signals in psychopathy. In: *Neuron*, 95(1), pp. 221–231.
HOWARD-SNYDER, F. (1997): The rejection of objective consequentialism. In: *Utilitas*, 9(2), pp. 241–248.
HUDSON, J. L. (1989): Subjectivisation in ethics. In: *American Philosophical Quarterly,* 26(3), pp. 221–229.
JACKSON, F. (1991): Decision-theoretic consequentialism and the nearest and dearest objection. In: *Ethics*, 101(3), pp. 461–482.
HULERMANN, R. et al. (2010): Oxytocin enhances amygdala-dependent, socially reinforced and emotional empathy in humans. In: *Journal of Neuroscience,* 30(14), pp. 4999–5007.
HURKA, T. (2001): *Virtue, vice, and value*. Oxford: Oxford University Press.
INSEL, T. R. & RERNALD, R. D. (2004): How the brain processes social information: Searching for the social brain. In: *Annual Review of Neuroscience,* 27(1), pp. 697–722.
JAMIESON, D. (2007): When utilitarians should be virtue theorists. In: *Utilitas*, 19(2), pp. 160–183.

JOHNSON, C. D. (1991): *Moral legislation*: *A legal-political model for indirect consequentialist reasoning*. Cambridge: Cambridge University Press.
KOSFELD, M. et al. (2005): Oxytocin increases trust in humans. In: *Nature,* 435(7042), pp. 673–676.
LEVIN, S. B. (2021): *Posthuman bliss? The failed promise of transhumanism*. Oxford: Oxford University Press.
MARSH, A. et al. (2010): Oxytocin improves specific recognition of positive facial expressions. In: *Psychopharmacology,* 209(3), pp. 225–232.
McCLOSKEY, H. J. (1957): An examination of restricted utilitarianism. In: *Philosophical Review*, 66(4), pp. 466–485.
MILL, J. S. (1859): *On Liberty*. New York: Norton.
MILL, J. S. (1861): *Utilitarianism*. Oxford: Oxford University Press.
MYRICK, H. et al. (2004): Modafinil: preclinical, clinical, and post-marketing surveillance – A review of abuse liability issues. In: *Annals of Clinical Psychiatry*, 16(2), pp. 1001–1009.
MORHENN, V. B. et al. (2008): Monetary sacrifice among strangers is mediated by endogenous oxytocin release after physical contact. In: *Evolution and Human Behaviour,* 29(6), pp. 375–383.
NAGEL, T. (1986): *The view from nowhere*. Oxford: Oxford University Press.
NIELSEN, K. (1972): Traditional morality and utilitarianism. In: *Ethics*, 82(3), pp. 219–231.
PARFIT, D. (1984): *Reasons and persons*. Oxford: Oxford University Press.
PERSSON, I. (2008): A consequentialist distinction between what we ought to do and ought to try. In: *Utilitas*, 20(3), pp. 348–355.
PETTIT, P. (1986): Social holism and moral theory. In: *Proceedings of the Aristotelian Society,* 86, pp. 173–197.
PETTIT, P. (1988): The consequentialist can recognise rights. In: *Philosophical Quarterly*, 38(150), pp. 42–55.
RAILTON, P. (1984/1988): Alienation, consequentialism and the demands of morality. In: S. Scheffler (ed.): *Consequentialism and its critics*. Oxford: Oxford University Press, pp. 93–133.
REYES, T. & MATEO, J. (2008): Oxytocin and cooperation: Cooperation with non-kin associated with mechanisms for affiliation. In: *Journal of Social, Evolutionary, and Cultural Psychology* (*Special issue of the Proceedings of the 2nd Annual Meeting of the North Eastern Evolutionary Psychology Society*), 2(4), pp. 234–246.
RODRIGUES, S. et al. (2009): Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. In: *Proceedings of the National Academy of Sciences of the United States of America,* 106(50), pp. 21437–21441.
SANDERG, A. (2011): Cognition enhancement: Upgrading the brain. In: J. Savulescu, R. Meulen & G. Kahane (eds.): *Enhancing human capacities*. New Jersey: Wiley-Blackwell, pp. 71–91.
SCHAEFER, G. (2015): Direct vs. indirect moral enhancement. In: *Kennedy Institute of Ethics Journal*, 25(3), pp. 261–289.
SEN, A. (1985): Well-being, agency and freedom. In the Dewey Lectures 1984. In: *Journal of Philosophy*, 82(4), pp. 169–221.
SEN, A. (2009): *The idea of justice*. Cambridge, MA: Harvard University Press.
SHAMAY-TSOORY, S. et al. (2009): Intranasal administration of oxytocin increases envy and schadenfreude (gloating). In: *Biological Psychiatry,* 66(9), pp. 864–870.
SHAW, E. (2014): Direct brain interventions and responsibility enhancement. In: *Criminal Law and Philosophy*, 8(1), pp.1–20.
SIDGWICK, H. (1874/1907): *The methods of ethics*, 1st ed. London: Macmillan.
SMART, J. J. C. (1973): An outline of a system of utilitarian ethics. In: J. J. C. Smart & B. Williams: *Utilitarianism: For and against*. Cambridge: Cambridge University Press, pp. 3–74.

SMITH, H. M. (1988): Making moral decisions. In: *Nous*, 22(1), pp. 89–108.
STAHL, S. (2006): *Essential psychopharmacology: Neuroscientific basis and practical applications*. Cambridge: Cambridge University Press.
THOMSON, J. J. (1976): Killing, letting die, and the trolley problem. In: *The Monist*, 59(2), pp. 204–217.
VINCENT, N. (2011): Capacitarianism, responsibility and restored mental capacities. In: L. Klaming & B. van den Berg (eds.): *Technologies on the stand: Legal and ethical questions in neuroscience and robotics*. Nijmegen: Wolf Legal Publishers, pp. 41–65.
WILLIAMS, B. (1973): A critique of utilitarianism. In: J. J. C. Smart & B. Williams: *Utilitarianism: For and against*. Cambridge: Cambridge University Press, pp. 77–150.
ZAK, P. J. (2008): *Moral markets: The critical role of values in the economy*. Princeton: Princeton University Press.
ZAK, P. J. et al. (2004): The neurobiology of trust. In: *Annals of the New York Academy of Sciences*, 1032, pp. 224–227.
ZAK, P. J. et al. (2005): Oxytocin is associated with human trustworthiness. In: *Hormones and Behaviour,* 48(5), pp. 522–527.
ZAK, P. J. et al. (2007): Oxytocin increases Generosity in Humans. In: *Public Library of Science ONE,* 2(11), p. e1128.