Authenticity and the Ethics of Self-Change

Alexandre Erler
Lincoln College, Oxford

A dissertation submitted for the degree of Doctor of Philosophy (DPhil) in Philosophy

Faculty of Philosophy
University of Oxford
Trinity Term 2013

AUTHENTICITY AND THE ETHICS OF SELF-CHANGE

Alexandre Erler
Lincoln College, Oxford
Thesis submitted for: DPhil in Philosophy
Trinity Term 2013

ABSTRACT

This dissertation focuses on the concept of authenticity and its implications for our projects of self-creation, particularly those involving the use of "enhancement technologies" (such as stimulant drugs, "mood brighteners", or brain stimulation). After an introduction to the concept of authenticity and the enhancement debate in the first part of the thesis, part 2 considers the main analyses of authenticity in the contemporary philosophical literature. It begins with those emphasizing *self-creation*, and shows that, despite their merits, such views cannot adequately deal with certain types of cases, which require a third option, "true self" accounts, emphasizing *self-discovery*. However, it is argued that in their existing versions, accounts of this third sort are also unsatisfactory.

Part 3 of the thesis proposes a new account of the "true self" sort, intended to improve upon existing ones. Common problematic assumptions about the concept of the true self are critiqued, after which a new analysis of that concept is presented, based on seven different conditions. Two specific definitions of authenticity, respectively emphasizing self-expression and the preservation of one's true self, are provided, and its relation to various associated notions, such as integrity or sincerity, are examined.

Finally, part 4 looks at the implications of the previous parts for the enhancement debate. In particular, it discusses the prospect of technologically enhancing our personality and mood dispositions. Do such interventions always threaten our authenticity, as some worry? A negative answer is provided to that question. Various potential pitfalls hinted at by the inauthenticity worry are discussed and acknowledged. It is, however, argued that such enhancements could still in principle be used in a fully authentic manner, and that they have the potential to bring about genuine improvements in our mood but also to our moral capacities and our affective rationality more generally.

TABLE OF CONTENTS

AUTHENTICITY AND THE ETHICS OF SELF-CHANGE

1 INTRODUCTION

1.1 Authenticity, self-transformation, and the enhancement debate

As Charles Taylor argued twenty years ago in his book *The Ethics of Authenticity*, many people in contemporary Western culture are committed to the ideal of living an authentic life (Taylor, 1991). A life that is inauthentic is essentially considered a failure. The notion of authenticity is taken to provide normative guidance about various aspects of life, including how to present oneself to others, what key choices to make, for instance regarding one's career, and also about whether or not to change oneself when doing so can be expected to bring social or economic rewards for oneself. Authenticity, it is typically assumed, demands that we express our real feelings and opinions, and more generally that we present ourselves to others as we really are; that we live our life in accordance with our deepest values and preferences, and that we take responsibility for our choices and actions, rather than letting others steer the course of our existence or thoughtlessly modelling ourselves on them. In many cases, authenticity is taken to require that we resist external pressures to shape ourselves a certain way. To agree to change ourselves in such circumstances, it is said, would constitute a form of self-betrayal. By contrast, being authentic means remaining somehow "true to ourselves" and declining to change ourselves in the relevant ways. Considering for instance the world of popular culture, actresses and singers who decline to boost their careers by undergoing cosmetic surgery are often praised for remaining true to themselves (a phrase typically used interchangeably with "being authentic"). As Debra Gimlin puts it, "[i]f not in feminist

theory, then in popular culture, there lies an implicit notion that the benefits of plastic surgery as somehow inauthentic and, therefore, undeserved" (Gimlin, 2002, p.79). Rock bands that refuse to tone down their image and musical style at the request of record companies are similarly praised for their authenticity, whereas those who comply are condemned as "sellouts". Finally, authenticity is also regarded as an important notion for other purposes than decisions regarding the private conduct one's own life. In the healthcare context, for instance, it is typically assumed that the degree to which we ought to respect a patient's choice, say to refuse treatment, depends at least in part on whether or not that choice is authentic. If there are grounds for thinking that it is not, this may be taken to mean that it is permissible to override it.

The present dissertation will assume that it is indeed intuitively plausible to regard authenticity as an important value, and will ask what the normative implications of authenticity are for our projects of self-transformation and self-creation. One class of self-transformation projects that I will be interested in, and one which I shall focus on in part 4, concerns projects that involve the use of so-called "enhancement technologies". I will be using that phrase to refer to a broad spectrum of procedures including, for instance, cosmetic surgery, or the use of steroids in sports, but also (and most importantly for my purpose) the various means available today of directly intervening in a person's brain with a view to improving mood, cognition, or changing personality, even when the person does not meet the criteria governing the attribution of any disorder or disability at those different levels. Such means include psychotropic drugs, as well as more invasive procedures like deep

brain stimulation.[1] A common view is that when the effect of such interventions is simply the preservation of health or "normal functioning", they count as treatments, but that they will constitute enhancements if they improve the relevant capacity beyond that "normal" level (see e.g. Juengst, 1998, and Daniels, 2000). It is typically assumed that interventions the aim of which is treatment, such as reconstructive surgery following an accident, are unobjectionable (though one might of course be concerned that they are not equally accessible to all), but that the use of enhancements, such as "purely" cosmetic surgery destined to give someone a more sensual appearance, is ethically more problematic.

The rise of enhancement technologies within the last few decades has thus opened up new ways for us to shape ourselves as we desire, but in doing so it has also generated much moral debate. A number of different ethical concerns have been raised about the use of such technologies: it has been suggested for instance that they might exacerbate social inequalities, or that they might destroy our ability to appreciate life as a gift. Among such concerns, the worry that enhancement technologies threaten our authenticity occupies a prominent place. From now on, I shall speak of *the authenticity objection* to refer to that particular worry. It is encountered both in casual conversation and in the philosophical literature on this topic. We should immediately note, however, that there is not one single way of spelling out the authenticity objection, but several, as we shall see in more detail in what follows. Also, whereas I shall understand the authenticity objection to imply that

[1] Of course, we are already familiar with more traditional ways of improving such human features, like psychotherapy or meditation, and these are typically not perceived as ethically problematic. While some might want to extend the label "enhancement technologies" to refer to such procedures, it seems to me more standard practice to use this label to designate interventions that fundamentally involve industrially produced artifacts, such as pills or electrodes. In what follows I will therefore reserve the label for these more "high-tech" and controversial procedures, even though I shall also discuss more traditional means of improvement.

all use of enhancement technologies is morally problematic, it is also possible to raise concerns about authenticity in relation only to *some* specific cases of enhancement use. My contention in this dissertation will be that the authenticity objection, due to its sweeping nature, is implausible, but that it is nevertheless fully legitimate to object, on grounds of authenticity, to certain particular types of self-transformation involving the use of enhancements. Some of these types of cases, which I will undertake to describe in part 4, have been somewhat neglected by contemporary bioethicists.

In their 2003 report *Beyond Therapy*, the members of the President's Council on Bioethics appear to have one possible variant of the authenticity objection in mind when they write that

> In seeking by these means [i.e. enhancement technologies] to be better than we are or to like ourselves better than we do, we risk "turning into someone else," confounding the identity we have acquired through natural gift cultivated by genuinely lived experiences, alone and with others. (The President's Council on Bioethics (U.S.), 2003, p.300)

This worry about a possible threat to our identity from enhancement technologies is echoed by Carl Elliott, the bioethicist who has probably written most extensively about the authenticity objection, in passages like the following:

> What is worrying about so-called "enhancement technologies" may not be the prospect of improvement but the more basic fact of altering oneself, of changing capacities and characteristics fundamental to one's identity. … [Deep] questions seem to be at issue when we talk about changing a person's identity, the very core of what the person is. Making him smarter, giving him a different personality or even giving him a new face—these things cut much closer to the bone. … They mean, in some sense, transforming him into a new person. (Elliott, 1999, pp.28-9)

The concepts of "identity" and "new person" being appealed to in this context are, as we shall see, ambiguous, and stand in need of further clarification. For the moment, let me stress that Elliott is somewhat ambivalent towards the authenticity objection in his writings and it is not clear that he wants to fully endorse it, yet the worries he describes in the passage just cited seem shared by a significant number of people in our society.

A related question that is sometimes raised in the context of the enhancement debate is whether the use of some particular enhancement technology really constitutes a "true" or "authentic" enhancement. Erik Parens, for instance, has argued that we should move beyond the polarization of the debate between "pro-" and "anti-" enhancement writers, and focus instead on asking what counts as a true, as opposed to a phony enhancement (Parens, 2011). Though this is in a sense a question about authenticity, it does not ask whether enhancement technologies threaten *our* authenticity. Rather, it concerns the nature of *the interventions* themselves: do they count as "true" enhancements, and do they deliver what they promise? This question makes sense (and the answer cannot just be, "enhancement technologies must by definition produce true enhancements, otherwise we would not call them that way!"), because we need to distinguish different senses of "enhancement". Julian Savulescu, Guy Kahane and Anders Sandberg have distinguished for instance between *functional* enhancement, i.e. enhancement of a particular human function or capacity such as memory, and *human* enhancement, which they take to mean that the intervention in question tends to increase the person's well-being (Savulescu et al., 2011b, pp.3-8). Another relevant sense of enhancement which I believe could be distinguished here would entail that the intervention in question allows the person to lead a *better life*.

Unlike Savulescu and colleagues, I think this is distinct from "human enhancement" in their use of the phrase: an intervention can in principle make your life go better *for you* (i.e. promote your well-being) without necessarily allowing you to live a *better life*. For instance, your life post-enhancement might be questionable from an ethical perspective even though it would be in your interest to live it. Yet that is a minor point. What is clear is that it is human enhancement in the sense given by Savulescu and colleagues, and possibly the additional sense I have distinguished, rather than functional enhancement, that is most relevant to Parens's question about what counts as true enhancement. Now if the use of some particular enhancement technology were to fail to promote our well-being, or to help us live a better life, or if it even proved harmful in this regard, that is how we ought to describe it. Saying that it had made our life inauthentic would seem an inappropriate way of stating the problem[2] – unless e.g. the idea was that our well-being had been diminished because our life had been made inauthentic by the intervention. But even then, the question of its impact on our authenticity would precede the issue of the authenticity of the enhancement itself, and should be considered separately (unless one wanted to identify well-being with authenticity, an option that doesn't appear particularly plausible). Though the question whether some particular intervention constitutes a "true" or authentic enhancement in a sense other than the functional one (as opposed to the question whether it threatens our authenticity) is certainly important, I will not be able to discuss it in much depth, as I don't have the space to deal with such a huge issue as the nature of well-being. I can say, however, that insofar as the view I shall defend does recognize authenticity as an important ideal worth striving for, it does imply that if some enhancement compromises the authenticity of our life, it will typically not

[2] To be fair to Parens, he does not himself claim that this is the same issue as the nature of "true" enhancement.

make it a better life, but will rather have the contrary effect, in which case it will thus not count as a "true" enhancement in that sense.

Much of the discussion of the potential threat to authenticity from enhancement technologies has been focused on what psychiatrist Peter Kramer famously called "cosmetic psychopharmacology", i.e. the use of psychotropic drugs by people who do not meet the criteria warranting a diagnosis of psychological disorder, with the purpose of improving their cognitive ability, mood, or personality. The extent to which cosmetic psychopharmacology is a real phenomenon today is not entirely clear: reliable data are not easy to come by, to which we should add the problem of deciding exactly which cases of psychotropic drug use are to count as "cosmetic" and which are not. That said, some genuine cases of cosmetic psychopharmacology do seem to have been documented. One example is the use of psychostimulants like Ritalin (methylphenidate) and Provigil (modafinil) by students and academics, particularly in the United States, not for treating any of the disorders that such drugs were basically designed to treat (such as Attention-Deficit Hyperactivity Disorder) but rather for the purpose of working longer hours and improving focus, e.g. in preparation for exams.[3] Other possible examples are given in Peter Kramer's book *Listening to Prozac*, in which he describes the spectacular personality changes (and cognitive benefits, such as enhanced clarity and quickness of thought) he reports to have observed in several of his patients whom he initially treated for psychological conditions like depression or obsessive-compulsive

[3] For instance, an informal survey run by the journal *Nature* about Ritalin, Provigil and beta-blockers reports that "[o]ne in five respondents [from 60 different countries] said they had used drugs for non-medical reasons to stimulate their focus, concentration or memory" (Maher, 2008, p.674). In relation specifically to college students, another survey by Teter and colleagues found a lifetime prevalence rate for illicit stimulant use of 8.3 % among respondents. The main reported motives were to help with concentration and studying (Teter et al., 2006).

personality disorder, but who asked to remain on the drug even once their original symptoms had been removed (Kramer, 1994). Formerly tentative, shy and vulnerable people are described as becoming confident, resilient, decisive, and more successful in their professional and personal lives once they are put on the drug. However, it isn't fully clear to what extent it is appropriate to talk of enhancement, rather than just treatment, in relation to Kramer's cases. While he occasionally suggests that some of his patients failed to meet the strict diagnostic criteria for any particular disorder, on which account we might want to call them "healthy", he also describes a number of them as exhibiting symptoms akin to depression, and as having "dysthymia", a condition that the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) characterizes as a chronic state of depression. True, Kramer describes changes of such a magnitude that he takes them to go beyond normal functioning, making the patients not just well again but "better than well" (ibid., p.x). Still, this claim is difficult to assess, as it is hard to know what the drug may have accomplished in those patients by simply lifting their depression and bringing them to a "normal" state. The fact that they may never have experienced such a state before going on medication doesn't yet demonstrate that what they experienced was not just treatment but enhancement, given the possibility that they may have been chronically depressed since early on in their lives. Couldn't it be that Prozac simply allowed their "normal", healthy personality, energetic and outgoing, to emerge at last?

The question whether Kramer's cases involve enhancement rather than just treatment isn't crucial for our purposes, however. What is more important from the perspective of authenticity is whether Prozac changed an aspect of these patients' true self (a notion I shall endeavour to clarify in the remainder of this dissertation), or

whether it rather allowed their true self to be expressed by removing an internal obstacle. What Kramer says suggests that the former is at least sometimes true: some people diagnosed as "dysthymics" may simply have what we might call a melancholy personality. Their depression might be a stable characteristic, not an aberration in their psychological life, and it might not, either, result from an independently identifiable physiological dysfunction, such as a deficiency in iron levels. If such possibilities were ruled out, there would be a good case for regarding dysthymia as part of the person's true self. What may be more important to highlight is that the cases of radical self-transformation described by Kramer, even if they are indeed cases of enhancement, are currently anecdotal and that systematic studies of "cosmetic" antidepressant use have not found any such dramatic effects (see e.g. Repantis et al., 2009). More modest effects were, however, reported in a 1998 study in the *American Journal of Psychiatry*: the authors found that administering the antidepressant Paxil to healthy subjects led to lower results on measures of hostility, and increased pro-social behaviour (Knutson et al., 1998).

That said, it now seems more appropriate to talk more broadly about "cosmetic neurology", or neuroenhancement, rather than cosmetic psychopharmacology when referring to the enhancement of cognition, mood or personality by technological means, since psychoactive drugs are no longer the only available option in this regard. There is thus some evidence that various forms of brain stimulation, currently used for the treatment of neurological conditions like Parkinson's disease, can also have an enhancing effect on cognition and mood among healthy individuals, allowing us to foresee that they might at some future point be used specifically for enhancement purposes. They include invasive procedures, like

deep brain stimulation (Synofzik and Schlaepfer, 2008), but also non-invasive ones such as transcranial direct-current stimulation (Cohen Kadosh et al., 2012). Moreover, as our understanding of the workings of the brain continues to advance within the coming decades, we can expect to see an improvement in the safety and effectiveness of neuroenhancers, enabling us to produce even more spectacular and more precise changes than those described by Kramer. This is likely to increase the attractiveness of cosmetic neurology to the general public. Rigorous discussion of the ethical issues raised by cosmetic neurology, including the authenticity objection, thus seems appropriate at the present time.

Interestingly, several of Kramer's patients (usually women) who requested to get back on Prozac once the medication had been stopped (or its dose lowered) claimed that they "no longer felt themselves" without the drug. The best-known of Kramer's examples is a woman called Tess. Despite a tough childhood marked by sexual abuse, followed by a failed marriage to an alcoholic husband, Tess, Kramer tells us, had been remarkably successful in her professional life, having risen to a high position in a large corporation while still taking care of her clinically depressed mother. Tess herself initially came to see Kramer for depression, having already been through a prolonged period of psychotherapy. After he put her on a first antidepressant called imipramine, she appeared to have recovered, yet she still showed the vulnerable and subdued disposition that had characterized her throughout her life. She thus felt upset by the negotiations she had to conduct with the aggressive leaders of a workers' union on behalf of the company she worked for. Her personal life was especially unfulfilling: Kramer tells us that Tess "considered herself unattractive to men" and that her only relationship for the past few years had been

with a married man who eventually rejected her in favour of his wife. Every time his name was mentioned in the course of a consultation, she started crying. However, once Kramer prescribed Prozac to ensure that her depression would not return, Tess suddenly started displaying a social confidence, energy and unshakeable resilience she had never shown before. The negotiations with the union, we are told, were no longer a daunting prospect for her: "[s]he was less conciliatory, firmer, unafraid of confrontation". Her romantic life underwent a complete revolution. Suddenly she had as many as three dates a week-end, and far from crying on hearing the name of her former lover, she said she no longer thought about him at all. Having observed this spectacular improvement in Tess's well-being, Kramer eventually discontinued the drug. Yet a few months afterwards, Tess, though she no longer showed any depressive symptoms, asked if she could get back on Prozac, claiming she was "not herself" anymore without it (Kramer, 1994, pp.7-10). As Kramer notes, if taken at face value, such a statement would have the implication that Tess's authentic self was only able to finally emerge once she had been put on Prozac, and that in the absence of such medication, she would have lived her whole life without ever being herself (ibid., pp.18-19). However, this will not count as paradoxical if we assume that Tess's true self had been stifled by chronic depression throughout her life, and that it was only allowed to emerge once Prozac had lifted her depression. Yet if we assume with Kramer that the drug actually did more than just remove Tess's depression, many people – both philosophers and non-philosophers – will find her case troubling and fear that her authenticity may in fact have been compromised, even if her own self-perception says the contrary. Carl Elliott, as we have seen, raises the worry that enhancement technologies (including neuroenhancers) will alter a person's "identity". Yet elsewhere in his writings he also suggests another possible construal of the

authenticity objection, no longer referring to the problematic status of changing a person's identity, but rather implying that the new personality traits produced by neuroenhancers are somehow "fake", or at least, in some sense, not really "ours". He thus writes, regarding the use of Prozac for "cosmetic" purposes:

> It would be worrying if Prozac altered my personality, even if it gave me a better personality, simply because it isn't *my* personality. This kind of personality change seems to defy an ethic of authenticity"... What could seem less authentic, at least on the surface, than changing your personality with an antidepressant? (Elliott, 1998, p.182, 186)

That said, Elliott goes on to note that the question of the implications of cosmetic psychopharmacology for personal authenticity is a tricky one, mentioning the fact that some users experience their use of medication as authenticity-promoting. Still, earlier in the same paper, he offers yet another way of developing the authenticity objection, implying that enhancement technologies might compromise our ability to appropriately respond, through our feelings and behaviour, to our life circumstances. He asks us to imagine a case of the following sort:[4]

The accountant's case. An accountant living in Downers Grove, Illinois, suddenly comes to himself one day and asks himself: "Jesus Christ, is this it? A Snapper lawn mower and a house in the suburbs?". He has come to feel deeply alienated from his way of life. He talks to his psychiatrist about his problems, who diagnoses him with depression and, at his patient's request, puts him on Prozac. The accountant soon feels much better, not alienated anymore but now at peace with himself and his run-of-the-mill life. (Elliott, 1998, p.180)

Elliott's suggestion here is that there is something deeply problematic, from

[4] For ease of reference, all the main cases I will be discussing are listed separately, in the order in which they appear, in the Appendix at the end of this dissertation.

the perspective of authenticity, about such a way of addressing the accountant's sense of alienation. A natural way of understanding this would be to say that there might be a different way of life that would make the accountant feel fulfilled and provide him with a real sense of purpose (perhaps a life devoted to some noble, charitable cause), in which case that is the life he ought to be living if he can. This kind of life, contrary to his current one, would be truly *his own*. Quelling the accountant's sense of alienation with the help of Prozac, on the other hand, would destroy his awareness of his predicament and any motivation he might have to make appropriate changes to his life. Yet it appears that Elliott (and as I shall explain, I believe this makes his position less plausible) actually wants to go further than that. He wants to say that feelings of alienation are an appropriate response, not just to a bland conformist lifestyle like that of the suburban accountant, but to a fundamental fact about our condition as 21st-Century Westerners: namely, we find it hard, if not impossible, to believe any more in a transcendent framework that might tell us what constitutes a good, meaningful life for us. "It is not just questioning the givenness of one's own form of life; it is questioning whether *any* form of life can have the kind of justification that you feel you need. It is a sense that all our ethical and epistemological practices are up for grabs" (Elliott, 1998, p.180). And later on: "The novel says, of course you're depressed. Take a look around you; it would take a moron not to be depressed" (ibid., p.183).

Concerns about authenticity are not only relevant to the enhancement debate, but also to ethical dilemmas relating to mental disorder. A widely discussed example is the treatment of Attention-Deficit Hyperactivity Disorder (ADHD) with psychostimulants like Ritalin or Adderall – though this discussion arguably overlaps

with the debate about cosmetic neurology, given that a number of cases in which a person is diagnosed with ADHD and put on medication lie in a grey area where it is not clear whether or not the person should be regarded as genuinely "ill". Several recent studies have addressed the issue of ADHD treatment and its relation to the authenticity of the patient. Ilina Singh has looked at the pharmacological treatment of young boys with ADHD, and has suggested that their mothers often justify their decision to medicate their child by arguing that their son's "real" self is the one on medication, i.e. that medication restores the boy's authenticity, compromised by the disorder (Singh, 2005, p.40). Maartje Schermer and Ineke Bolt have interviewed adults with ADHD, asking them about their experience of medication and their view of its relation with their sense of identity. In a somewhat similar vein to Kramer's patients, some of the subjects interviewed reported feeling *more* themselves on medication. Others, by contrast, felt less authentic. Accordingly, members of the latter group either chose to discontinue medication, or if they remained on it for the sake of the social and professional advantages of doing so, still felt that the medication had robbed them of a valuable part of their identity – which is why some of them would go off medication when they had the opportunity, e.g. outside of the job environment (Bolt and Schermer, 2009). Such studies call into question some preconceived ideas one might have about the relation between disorder and the authentic self, and thus about the implications of authenticity for the choice to treat or not to treat. While a focus on conditions like depression might suggest that treating a disorder must necessarily restore the person's authentic self which was masked by pathology, the ethical debate around ADHD treatment shows that such an assumption can be questioned in the case of some disorders.

As I have mentioned, my aim in this dissertation will be to assess the soundness of concerns about authenticity in specific relation to the use of enhancement technologies. While rejecting the authenticity objection as too sweeping to be plausible, I will nevertheless argue that the worries about authenticity voiced by a number of people concerning these technologies do touch on important normative considerations that haven't yet received the attention they deserve. I shall add, however, that talk of authenticity in this context sometimes seems to rely on other ethical concepts such as self-respect, the need to preserve existing value, and resistance to intolerant social norms. Showing this, however, will require a substantial amount of preliminary conceptual work. I will start by considering some common pre-philosophical ideas associated with the concept of authenticity. In part 2, I will look at philosophical accounts of authenticity in the contemporary literature, and distinguish three main ones. I shall argue that the third one, which I refer to as the true self approach, allows to capture some of the key concerns raised about authenticity in relation to enhancement, concerns that are left out by the other two analyses, largely focused on the concept of autonomy. Yet I shall also show that the true self analysis in its existing forms is unsatisfactory. In part 3, I will offer my own version of the true self analysis, intended to improve upon existing ones. This will require me to defend the concept of the true self, often criticized as a misleading "folk" notion. I will propose two definitions of authenticity, both of which I take to be relevant to the enhancement debate: first, authenticity as valuable self-expression, and secondly, authenticity as the preservation of valuable aspects of our authentic self. Thus equipped, I shall focus in part 4 on the question of authenticity and enhancement technologies. Unfortunately, I shall not have the space to consider all the interventions that raise important concerns about authenticity, such as cosmetic

surgery. I shall focus on the domain that has largely started that debate and is one of those that generate the most controversy, the technological enhancement of personality and mood. My main argument will be that it is in principle possible to use such interventions in authentic, and ethically sound ways, yet that other potential uses do pose a threat to our authentic selves, and that this gives us a reason to avoid and discourage them. The main lines of that analysis should be transferable to other types of enhancement technologies.

1.2 Pre-philosophical intuitions about authenticity

Let us therefore begin with the task of characterizing authenticity. Considering what can be said about the notion prior to examining the accounts that philosophers have proposed about it, we may start by noting that the term "authentic" can be applied to a variety of different things: first, inanimate objects, such as when we speak of an authentic Monet, meaning that the painting in question was really produced by Monet himself, not by a copyist or fraudster. A coffee brand can also market its product as having an authentic taste – that is, as really tasting like coffee from Columbia, rather than having a very different, artificial taste falsely presented as the real thing. These uses correspond to the definition of authenticity given by the *Oxford English Dictionary* as "[t]he quality of being authentic...as being what it professes in origin or authorship, as being genuine". Another sense of "authentic" given by the *OED* is "being in accordance with fact...being true in substance": this is what we mean when we say for instance of someone's testimony that it is authentic (Oxford English Dictionary, 2012). But we also describe people, their lives, choices, actions, and mental states (e.g. their emotions, moods or desires) as authentic or

inauthentic. Particularly in the context of the enhancement debate, people's capacities, and their performances in various areas of human endeavour (such as sports) are also assessed in terms of their authenticity. It is this particular category, that of agents and what they do, think, and feel, that I will be interested in when I discuss the question of authenticity.

As I have suggested above, describing someone or their actions as authentic is often meant to indicate praise: to say for instance that a certain musician is an authentic artist typically counts as a compliment, suggesting that the person in question has a genuine passion for music and that she follows her own artistic vision in her creations, rather than complying with external pressures (e.g. from her sponsor or record company). Sometimes, however, the term "authentic" is also used in a more neutral manner. That is particularly the case in the context of discussions in bioethics about whether someone's decision (to refuse treatment, for instance) is "authentic", namely whether it really emanates from the person rather than e.g. from external pressures or from a mental pathology. The implication is often that if the person's decision is authentic, we ought not to override it since it expresses what she really wants. If the decision to refuse treatment is inauthentic, on the other hand, no such respect is due to it anymore, as it does not emanate from the *person* herself, but rather from her pathology, or her pushy relatives, etc. And describing a decision as authentic in this sense does not entail that we regard it as good or praiseworthy in any way. It is perfectly possible to simultaneously condemn it as irrational – all we then mean is that it accurately reflects the person's beliefs, preferences and set of values. Finally, people sometimes colloquially speak of "an authentic psychopath", for instance, in which case the term is simply used as a synonym for "real", and adds emphasis to the

statement being made about the person in question.[5]

The core idea behind the concept of authenticity as most people understand it in contemporary Western society, and which we have already encountered above, is that the authentic person is somehow being "true to herself", which is contrasted with being "fake", "phoney" or insincere. Authentic people, it is assumed, do not dissemble. They accurately represent to others what their actual beliefs, feelings and preferences are. To the extent that they are being true to themselves in the choices they make and the actions they perform, these choices and actions count as authentic. Accordingly, an authentic life is that of someone who chooses and acts authentically with enough consistency, and whose most momentous choices, in particular (e.g. which career to pursue), are authentic. This idea of being true to oneself is usually not spelt out in much detail, and as we shall see later this becomes a source of difficulties when one tries to figure out whether or not some particular person is being true to herself. But for the moment we can at least offer a rough idea of what people have in mind when using the notion: for example, if I love the Arts and a bohemian lifestyle more than anything else, many people will argue that being true to myself means shaping my life in accordance with this set of preferences, even if doing so elicits the disapproval of my parents, who had hoped that I would be perpetuating the family tradition of a career in banking.

As regards authentic mental states, such as feelings, emotions, moods and desires, they are usually defined in contrast to merely feigned ones, in keeping with the remarks above linking authenticity to sincerity. Yet they are also sometimes

[5] The last sense of "authenticity" given by the *OED* is thus the quality of being "real, actual". See ibid.

identified with “natural” states, as opposed to “artificial” ones. “Artificial” can of course be used as a synonym for feigned, but it can also be, rather, taken to mean that the state in question was produced by some artifice, as opposed to spontaneous. There are many such artifices, more or less sophisticated. At the simplest level, we have techniques of mood and emotion management: examples include counting one’s blessings and avoiding unfavorable social comparisons in order to boost our happiness level, or deliberately avoiding thinking about a funny joke one had just heard in a situation where amusement would be out of place, such as a funeral. The practice of advertising is a familiar means of inducing “artificial” desires in people for items they otherwise might not have thought of desiring – as opposed to the objects of “natural” desires such as those for food or shelter. Another way to artificially induce emotions or moods in someone would precisely be to use enhancement technologies, existing or prospective. There is a widely shared assumption that improving oneself, e.g. increasing one’s level of self-confidence, by “artificial” means such as pills, is inherently inauthentic, whereas doing so by supposedly more “natural” means, such as psychotherapy or coaching, is compatible with authenticity. This idea is a rather questionable one, as it is not clear why drugs, for instance, should count as artificial means whereas psychotherapy should not, since both are human inventions. Yet the contrast does often seem to be drawn that way. Also, as mentioned at the beginning, authenticity, besides involving being honest about who we are and what we believe and value, is often taken to mean not changing oneself in circumstances where it might be tempting to do so, e.g. because of the social and economic advantages at stake for oneself. That is why the concept of authenticity crops up so often in the debate about the use of enhancement technologies.

A link is often drawn between the idea of authenticity and that of spontaneity. Spontaneous reactions and feelings, it is often said, reflect the authentic self. More reflective and controlled ones, by contrast, are often misleading about who one is, even though they tend to be more in keeping with our ideal image of who we would like to be. In those cases, one might say that we are – consciously or not –repressing the declarations of our true self in order to make us more acceptable to ourselves and to others, to embellish ourselves. That is why some hold that the person whose inhibitions have been removed under the influence of alcohol is showing her authentic self to a greater degree than the more timid person she is when sober. The true self is the one that is expressed when fear and other inhibitory factors are removed. This view, however, is not unanimously accepted: some also tend to accept the fact that a person was drunk as an excuse for bad behaviour, on the assumption that she was not really herself while under the influence.

Finally, we may note that the concepts of authenticity and the true self are not unique to Western culture: they are also present, for instance, in East Asian culture. However, studies have found some differences in terms of how exactly these different cultures think of the true self. Westerners, for instance, tend to view authenticity as requiring one to express a set of traits in a stable and consistent manner across different social contexts, whereas East Asian culture allows that the true self the person ought to express may exhibit inconsistency between different social contexts (though consistency within one specific context is still required). [6] To the extent that the concept of the true self is meant to refer to the features that fundamentally define who a person is, it is closely related to the notion of "identity", which I shall discuss

[6] See e.g. Schlegel et al., p.487.

at length in a later section. Here again, some cultural differences have been pointed out in terms of how people conceive of a person's identity, with some (non-Western) cultures placing much more importance on social role for the definition of that identity (Rorty and Wong, 1990, pp.27-8).

In what follows I will be concerned with the authenticity of our lives, choices and actions, as most accounts of authenticity seem to be focused on these particular notions, but I shall also consider the authenticity of psychological features like moods, emotions, and personality traits, a relevant issue when discussing "cosmetic neurology".

1.3 The popular ideal of authenticity

In *The Ethics*, Charles Taylor argues that a great number of people in contemporary Western culture are committed to the ideal of living an authentic life. He describes this ideal as "the moral ideal...of being true to oneself, in a specifically modern understanding of that term" (Taylor, 1991, p.15). As it is enacted around us, he says, the ideal goes hand in hand with two other important notions: first, the search for self-fulfilment (Taylor talks about the "individualism of self-fulfilment"),[7] and secondly, what he calls "soft relativism", a view he describes as involving both a metaethical and a normative claim. The first, metaethical claim is that "[e]verybody has his or her own 'values', and about these it is impossible to argue". The second, normative claim says that we ought not to challenge other people's values or life projects even when we dislike or disapprove of them. These things are their choice, and we ought to respect that choice (ibid., pp.13-14). In what follows I will be

[7] Ibid., p.14.

especially interested in the link Taylor correctly points out between authenticity and self-fulfillment in the popular mind. I will not discuss the link he draws between authenticity and "soft relativism". Even though this claim by Taylor has some plausibility, first, empirical data would be needed to establish its truth (Taylor does not cite any studies but relies on anecdotal evidence and Bloom's remarks), and secondly, even if such a link does obtain, soft relativism seems to have so little to recommend itself that if our aim is to develop a plausible view of authenticity and its implications for the enhancement debate, there is no reason to consider taking soft relativism on board.

As Elliott shows in his book *Better Than Well*, the idea that we should find our true self and be "true to ourselves" has become something of a cliché in contemporary Western society, and it is a staple of the self-help literature (Elliott, 2003, chap.2). Elliott points out two features which he sees as central to the contemporary ethic of authenticity. The first one is that this ethic involves a new assumption that our lives are planned undertakings for which we are responsible, in replacement of, for instance, the assumption that life is "something that is entrusted to you and determined by God, so that the purpose of your life is to follow God's will" (Elliott, 1998, p.181). The second feature of the contemporary ethic of authenticity is the assumption it makes about the conditions for living a good life. "An ethic of authenticity", Elliott writes,

> says that in order to answer the question, "How should I live?" I will have to look inward, because there is no single universal way of living a meaningful life. The answer to this question will differ from one person to the next, and each person has to discover the answer for himself. "You have to find yourself," we sometimes hear. "You have to find your own way." "You have to be true to yourself." Each "self" is different and unique; for a life to be a good life, a meaningful life, a life

> properly oriented toward the good, we have to get in touch with ourselves. (Ibid., pp.181-2)

Taylor's claim about the pervasiveness of the ideal of authenticity in contemporary Western culture does not need much arguing for. It is now a familiar fact that many people cite "being true to themselves", or simply "being themselves", an idea often used interchangeably with that of being authentic, as a key principle offering guidance in how they are to live their lives. It is appealed to in justification of various momentous decisions, whether it be leaving a job or ending a relationship: a person might say that the job did not reflect her true calling, or that the relationship was holding her back, not allowing her to further develop as an individual. Also, during the past few decades, the idea of being true to oneself has become one of the most widespread slogans within Western popular culture. It has been used to advertise all sorts of goods from soft drinks to mobile phones. Pop musicians such as Nelly Furtado proclaim their determination to "stay true to themselves", while self-help gurus promise potential customers that their programs destined to develop self-confidence or other skills will not turn them into someone they are not, but will on the contrary allow "the real them" to shine through at last. Such anecdotal observations are corroborated by the work of some philosophers from the "experimental" school, who claim to have found results indicative of the influence of the ideal of being yourself on people's ethical reasoning (see e.g. Knobe, 2005).

There is a large body of empirical evidence showing that people tend to believe in the existence of a true self, which they understand as a set of stable traits that help explain a person's behaviour. This true self is taken to include the "core", most fundamental features of the person, and is often thought of as some kind of essence

(see Riis et al., 2008, and Newman et al., under review). This idea of an essence suggests that the true self tends to be viewed as an entity that persists through the life course and cannot be changed – we can only decide whether to manifest it in our behaviour or to repress it. When it comes to identifying the true self within each of us, this essentialist understanding of the true self implies that this is to be achieved through a careful work of introspection and self-examination. However, it should be noted that this assumption is not universally accepted. Some people, rather than seeing their "true self" as something already there to be discovered and to which they ought to conform if they are to live authentically, identify their true self with their *ideal* self, the self they most wish to have and are working towards developing. Elliott cites for instance the example of Sam Fussell, a shy student turned bodybuilder, who in his memoir *Muscle* describes his personal transformation as a process of becoming *himself* (Elliott, 2003, p.37). This contrast, as we shall see in the next section, reflects two different approaches to authenticity in the contemporary philosophical literature as well.

2 THE NATURE OF AUTHENTICITY: PHILOSOPHICAL ACCOUNTS

2.1 Two contrasting frameworks for understanding authenticity

In *The Ethics*, Taylor describes the popular ideal of authenticity with which we are familiar as originating crucially from the ideas of certain late 18th-Century thinkers. He mentions Rousseau as a precursor, and after him, the movement he describes as Romantic expressivism (Taylor, 2007, p.475), which he particularly associates with Herder and his idea that each of us has his/her own way of realizing

our humanity – his or her own particular "measure", in Herder's words.[8] Building on this core idea, here is how Taylor fully spells out the ideal of authenticity:

> There is a certain way of being human that is *my* way. I am called upon to live my life in this way, and not in imitation of anyone else's. But this gives a new importance to being true to myself. If I am not, I miss the point of my life, I miss what being human is for *me*. This is the powerful moral ideal that has come down to us. It accords crucial moral importance to a kind of contact with myself, with my own inner nature, which it sees as in danger of being lost, partly through the pressures towards outward conformity, but also because in taking an instrumental stance to myself, I may have lost the capacity to listen to this inner voice. (Taylor, 1991, p.29)

Furthermore, the ideal as Taylor describes it does not merely state that I ought to find my own way in life, rather than taking the facile path of conformity. It also presupposes what Taylor calls the principle of originality: "each of our voices has something of its own to say... Being true to myself means being true to my own originality". (Ibid., pp.28-9)

It is natural – though we shall see, not necessarily true to Taylor's intentions – to interpret Taylor's characterization of authenticity as heading in the direction of the authenticity objection. In the wake of a distinction proposed by Erik Parens (Parens, 2005), Neil Levy has contrasted two main ways of thinking about authenticity in relation to the question that occupies us: what he calls the "self-discovery" vs. the "self-creation" view of authenticity. I personally prefer to talk of "models" or frameworks here, as different specific views can fall on the same side of Levy's distinction. The self-discovery model is in line with the Romantic legacy taken up by Taylor: on this model, Levy writes, authenticity means being true to oneself, in the

[8] "Jeder Mensch haat ein eigenes Mass, gleichsam eine eigne Stimmung aller seiner sinnlichen Gefuhle zu einander" (Herder, 1877-1913, p. 291). Quoted by Taylor, 1991, p.127 n.22, and 1989, p.375. Taylor's translation is as follows: "Each human being has his own measure, as it were an accord peculiar to him of all his feelings to each other".

sense of "listen[ing] attentively to an inner voice which calls on us to be human in a way that is distinctively *ours*" (Levy, 2011a, p.3). But, Levy adds, there is also an important rival model of authenticity, omitted by Taylor in the *Ethics*, a model associated with Jean-Paul Sartre and, more recently, with the work of David DeGrazia. The self-creation model denies that authenticity consists in conforming our life to the dictates of a pre-given self. Indeed, supporters of that model typically deny that we have such a thing as a pre-given self. On that model, authenticity rather means confronting the fact that it is up to us to freely decide what shape our life should take, and in striving to mould ourselves in accordance with our deepest desires and values. Different advocates of the self-creation model may disagree as to the extent to which we are free to shape ourselves as we wish. DeGrazia thus takes care to distance himself from the radical suggestion made by Sartre in some of his writings that who we become is solely determined by the free choices we make, unimpeded by any other factors, whether our genetic endowment or anything else (DeGrazia, 2000, pp.36-7). But these authors agree in their rejection of the idea of a pre-given self that might offer practical guidance.

We need, however, to get clear about what this notion of a pre-given self means exactly. In his book *Neuroethics*, Levy defines the pre-given self as "a self that is innate in us" (Levy, 2007, p.105). In the recent paper cited above, he also identifies such a pre-given self with an "essence" (Levy, 2011a, pp.311-12), a term suggesting that such a self involves a set of traits so fundamental that they cannot be modified without putting an end to the existence of their possessor.[9] Now while some

[9] DeGrazia takes a similar line when he criticizes the authenticity concern for being "beholden to the rather implausible romantic notion of a 'true' self whose defining traits are independent of the individual's choice", a notion which he says is "most intelligible if construed as a person's *essence*". See DeGrazia, 2005, pp.233-34 (italics in original).

supporters of the self-discovery model may well be committed to one (or both) of these conceptions of the pre-given self, it does not seem correct to assume that espousing that model *necessarily* commits us to either of these. As Levy himself notes, Carl Elliott, one of the main authors associated with that model in the context of the enhancement debate, often talks about the "true self" (for reasons I shall explain, I also prefer using that phrase rather than speaking of a "pre-given self") and accepts that idea as valid, yet makes it clear that he does not understand it as an essence (Elliott, 2003, p.49). And the authenticity objection could hardly be taken seriously if it assumed that the traits targeted by enhancement technologies, and the modification of which strikes us as problematic, must be inborn. Such an assumption would make it virtually impossible to regard the modification of physical appearance, personality or mood via such technologies as inauthentic, since people are typically not born with such traits in their adult form, the form they will usually have when the person considers enhancing them. One might perhaps argue that a person's genotype is indeed inborn, and suggest that the pre-given self might be identified with it, but this would clearly not work. Indeed, the traits just mentioned are not just as it were written in our genes. Genetic determinism is a simplistic and mistaken view. Virtually all human characteristics, including personality traits, are the product of a complex interaction between genes and environment.[10]

I believe it is more helpful for our purposes to see the idea of the pre-given or true self as part of a certain approach to the question of what constitutes a person's *identity*. The sense of "identity" I have in mind here is *narrative* identity, as opposed to *numerical* identity, which is the sense relevant to philosophical discussions of so-

[10] See e.g. Lobo and Shaw, 2008. Regarding personality traits in particular, see e.g. South & Kruger, 2008.

called “personal identity”. Numerical identity concerns what Marya Schechtman has called the reidentification question: what is it that makes me one and the same individual at two different points in time? Narrative identity, on the other hand, has to do with what Schechtman calls the characterization question: which set of characteristics (actions, beliefs, values, character traits...) makes me the particular person I am (Schechtman, 1996, pp.74-6)? A Cartesian answer to that second question, for instance, would appeal to the idea of continuity of immaterial substance. That does not rule out the possibility that some features might be relevant both to my narrative and numerical identities. For instance, the fact that I was born of such and such parents might sometimes appropriately figure in an answer to the characterization question, but many authors would also regard it as tied to my numerical identity: it seems plausible to think that I would be a numerically distinct individual had I been born of different parents. Still, it is important to keep the characterization and reidentification questions separate.

I propose to understand the pre-given, or true self, as referring to a set of traits that contribute to defining who a person is *whether or not the person identifies with them*, i.e. whether or not she endorses them and is happy to regard them as definitive of who she is. I say that these traits *contribute* to defining the person, because I wish to acknowledge that they typically do not provide a complete account of a person’s identity. Indeed, while we have seen that it is not plausible to require the features constituting the pre-given self to be either inborn or essential, it might still plausibly be suggested that such features must have their grounding in *intrinsic* properties of the person, e.g. in the particular structure of her brain. Features like personality traits, for instance, appear to fit that description. However, this is not true of other features like

social roles (e.g. friend, head of state), which are fundamentally grounded in certain relations to others, but which we nevertheless often take to form part of a person's narrative identity.

Some might perhaps argue here that the concept of a *pre-given* self does suggest, if not that the features constituting it are inborn, at least that they must still be ones that the person *has not chosen*, and that the analysis I have just proposed leaves that out. Even if they are right, however, we should first note that their view cannot, either, turn the notion of the pre-given self into a complete account of narrative identity. Indeed, while people's social roles, for instance, might not always be chosen (Louis XIV, for instance, did not choose to become King of France when he was four years old), they often are, in which case they are still not "pre-given" in any plausible sense. And secondly, if we are to understand the notion of the pre-given self as fundamentally entailing that its constitutive traits were not chosen by the agent, it seems to me that the notion will unduly limit the applicability of concerns about authenticity, in the context of the enhancement debate and elsewhere. Indeed, while the features that authenticity, as the self-discovery model understands it, requires us to preserve will often be ones that we did not deliberately shape, it need not always be so. Consider for instance someone pressured by her new social circle to abandon the honest or temperant disposition she had spent years developing. Given this, I will henceforth speak of the *true* self rather than of the pre-given self, in order to emphasize that it is not a necessary property of its constitutive traits that they were not at all the product of the agent's choice. The key fact about those traits, as I have said, is that they contribute to defining who their possessor is, regardless of whether she now happens to identify with them or not. One could in principle work hard to acquire

a certain trait, yet change one's mind after acquiring it and fail to identify with it any longer. Even then, the trait in question could still be part of the person's true self, as I understand the notion.

Proponents of the self-creation model reject the idea of a true or pre-given self even as a partial account of narrative identity. DeGrazia, for instance, argues that if someone does not identify with a particular trait that is properly attributable to her, this trait will not count as part of her (narrative) identity.[11] Taylor himself actually seems to echo such a view in some of his writings – hence the caveat I have given about interpreting the passages from *The Ethics* quoted above as defending the self-discovery model.[12] The fundamental point of contention between the self-discovery and self-creation models thus seems to me to be this: the former model allows for the possibility that some of our features (those that constitute our pre-given self) might determine what constitutes an authentic life for us whether or not we identify with them, whereas the self-creation model rules this out. I will henceforth rely on this criterion as a way of distinguishing the two models.

It is usually assumed that the self-discovery model speaks in favour of the authenticity objection about enhancement technologies, while the self-creation one speaks against it. Yet Levy, as we shall see, argues that this assumption is mistaken: none of the models need be understood as fundamentally supporting the objection. To throw some more light on this issue, I will now examine the various contemporary

[11] DeGrazia, 2000, pp.37-8. As I understand DeGrazia, he takes "identifying" with some trait to mean *endorsing* it, and not merely recognizing it as a key part of who we are. Indeed, he writes for instance that "who we are has everything to do with what we value" (ibid., p.38). I follow his understanding of the term here.

[12] See for instance Taylor, 1989, p.27: "My identity is defined by the commitments and identifications which provide the frame or horizon within which I can try to determine from case to case what is good, or valuable, or what ought to be done, or what I endorse or oppose".

accounts of authenticity, and consider how exactly they relate to the authenticity objection. Three main ways of understanding the notion can be distinguished in the current philosophical literature: first, the view of authenticity as wholeheartedness; secondly, DeGrazia's view of authenticity, which incorporates existentialist elements; and finally, what I shall call "true self" accounts. The first two families of accounts belong to the self-creation model, while the third is equivalent to the self-discovery model.

2.2 Authenticity as "wholeheartedness"

The view of authenticity as wholeheartedness is often attributed to Harry Frankfurt (see e.g. Litton, 2005, p.66; and Cottingham, 2010, p.10). Even though he usually does not use the term "authenticity" himself, it seems that such a view can reasonably be derived from some of his writings. In an essay on autonomy and behaviour control, Gerald Dworkin provides the grounding idea behind this view when he describes authenticity as a person's second-order identification with her first-order desires (Dworkin, 1976, pp.24-5). To this, the "Frankfurtian" view – as I shall henceforth call it – adds that the identification in question must be "wholehearted", in the sense of not involving any ambivalence or inconsistency at the second-order level (see Frankfurt, 1988a, 1988b). The concept of wholeheartedness, as Frankfurt understands it, prevents the possibility of reaching contradictory verdicts about authenticity in cases where an agent feels ambivalent towards some first-order desire or preference. On the Frankfurtian account, the authentic agent is one who acts upon desires with which she wholeheartedly identifies. Choices and actions involving such wholehearted identification with the desires and preferences that guide them will qualify as authentic choices and actions – they will represent where the person really

stands. The Frankfurtian account of authenticity can equally be viewed an account of *autonomy*: the two notions are equivalent on this view (see e.g. Loughrey, 1998, and Oshana, 2006, p.22).

Let us now consider a series of examples that should throw more light on the nature of the Frankfurtian view and its implications. The first one is mentioned by Frankfurt himself in 'Identification and Wholeheartedness' (Frankfurt, 1988b, p.164):

Akratic. Akratic is addicted to smoking. She would like to quit but has been unsuccessful so far. At a party, when someone offers her a cigarette, she yields to her desire to have one even though she does not identify with it – in fact she would prefer not to have it.

There is a plausible sense in which Akratic's choice to have a cigarette is inauthentic, and the Frankfurtian account can make sense of it: her choice is not truly *hers*, as she clearly does not endorse the first-order desire that determines it. Contrast her case with the following one:

Rebel. Rebel, just like Akratic, is offered a cigarette at the party, which she accepts. Yet she is an unrepentant smoker who deems it perfectly rational to sacrifice some of her life expectancy for the sake of a more pleasurable life (and has no second-order attitudes that conflict with this outlook). Accordingly, she wholeheartedly endorses her desire to accept the cigarette.

Rebel's choice to have a cigarette, unlike Akratic's, will count as authentic on

the Frankfurtian view, which again seems correct.[13] I am assuming that the relevant identification with the first-order desire that causes one's action needs to occur *at the time of acting* for the action to count as authentic. This is what Frankfurt's writings suggest, even though he does not explicitly say so himself. One can think of other possible ways of developing the Frankfurtian line: one might for instance argue that actions should always be assessed as authentic or inauthentic *in relation to a specific point in time*, and not absolutely speaking. According to this line, if I wholeheartedly identify with the desire that guides my action at t1 (the time of acting), but no longer do so at t2 (say, ten years later) because my personal convictions have changed quite significantly, then my action is authentic relative to t1, but inauthentic relative to t2. It makes no sense to ask whether my action was authentic or inauthentic, full stop. Such a time-relative conception of authenticity, however, appears rather counterintuitive. Arguably, Akratic and Rebel's choices to accept the cigarette are, respectively, inauthentic and authentic, full stop. Also, the truth of that statement does not seem to depend on the future evolution of their higher-order attitudes. Consider the following variation on Rebel's case:

Ex-rebel. Ex-rebel also accepted the cigarette she was offered at the party, which she attended while in her twenties. At the time, she used to adhere to a rebellious, pro-smoking attitude. This is no longer the case, however: she abandoned this mindset in her mid-thirties, as her priorities in life had gradually changed with the years.

It seems plausible to think that Ex-rebel's choice to accept the cigarette at the party

[13] To be fully accurate, this verdict will only hold if we make some further assumptions: e.g. that Rebel's higher-order endorsement of her desire for the cigarette has not resulted from autonomy-undermining influences such as brainwashing. I will have more to say about this in the next section, but for the moment let us simply assume that such influences are absent.

she attends while in her twenties is authentic and that this remains true even after she has changed her outlook on life, because what matters for the authenticity of her choice at the party is the second-order attitude she *then* had. (Ex-rebel herself might accept this: "even though my philosophy of life is now quite different", she might say, "I agree that my behaviour at the party truly reflected the person I was at the time".) If that is correct, we should also reject the alternative suggestion that what matters for the authenticity of a person's action is whether she *now* identifies with the first-order motive that guided it. Such a view would have the implausible consequence that Ex-rebel's past choice to take the cigarette now counts as inauthentic, since she no longer identifies with it. Equally implausible is the implication that a particular past action that initially counted as authentic can later become inauthentic, or the other way round – since the view concedes that Ex-rebel's choice was indeed authentic at the time of the party. It thus seems more plausible to regard an agent's *synchronic* identification with her first-order desires as providing the criterion of authenticity on the Frankfurtian view.

We may note that the sufficient *stability* of Ex-rebel's initial higher-order attitudes is arguably crucial to the authenticity of her choice to accept the cigarette. We are assuming that these attitudes last for a significant period of time and only start to change as she reaches her thirties. The importance of stability can be brought out by contrasting Ex-rebel's case with the following one:

Influenceable. At the party, Influenceable, a repentant smoker like Akratic, finds herself surrounded by friendly hedonists who encourage her to join them in their smoking and to share their *carpe diem* philosophy. Temporarily falling under their sway, she accepts the cigarette she is being offered, and at that time feels no scruples at all about her choice – after

all, you only live once, she tells herself. However, a few days later, contemplating again her decision to abandon her efforts to quit smoking, Influenceable finds that she can no longer endorse it. She now feels she was misled by the carefree spirit that prevailed at the party.

It seems more plausible to declare Influenceable's decision to take the cigarette inauthentic, even though, at the time of acting, she did not disown her desire to take it. One way of justifying this verdict on the Frankfurtian view would be to argue that even if we accept the presence of a second-order attitude of endorsement of that desire at the time of acting, this attitude will likely fail to be wholehearted. Indeed, we can plausibly assume that Influenceable then still has (presumably without attending to them) other second-order attitudes incompatible with that first one, such as an identification with her desire to proudly announce to her doctor, at her next visit to his surgery, that she has finally stopped smoking. (Suppose Influenceable does not like to lie about such matters.) And, one might add, it is more plausible to assume that the special situation in which she found herself changed at most a limited number of her higher-order preferences, perhaps only one of them (i.e. that relating to her desire to take the cigarette), rather than *all* the preferences related to that particular one. I.e., it is not plausible to think that her higher-order preferences all remained fully integrated even during the brief time during which she felt and acted out of character.

Secondly, even if we assume that there was in fact no conflict between Influenceable's higher-order attitudes at the time when she accepted the cigarette, one might still plausibly argue that a reasonable degree of *stability* in one's attitude of endorsement is required for that attitude to count as wholehearted, or even for it to be attributable at all to the person in question. On that basis, one might maintain that Influenceable's second-order attitude of endorsement is too short-lived either to count

as wholehearted, or to be properly attributable to her at all. Rather, it might be said, when she takes the cigarette she is just temporarily confused about what her higher-order preferences really are on this issue.

Suppose, however, that Influenceable does not come to disapprove of her choice and the motive behind it until quite some time later (let us refer to the agent in this variant as Influenceable*). She spent, let us assume, all her University years in the company of these hedonist friends, and went along with their philosophy of life until she graduated and moved to a new city, where she started socializing with people more concerned about their health. As a result, she takes a critical look at her former hedonistic mindset, and ends up rejecting it. It may then well be that her case should be seen as similar to Ex-rebel's, and that her choice to accept the cigarette at the party should count as authentic. However, there might also be grounds for resisting that verdict, depending on how exactly we construe Influenceable*'s case. For instance, it may be that she was *self-deceived* when she told herself, throughout her University years, that she agreed with her friends' hedonistic outlook. Perhaps she entered the University environment with the belief that one ought not to compromise one's long-term health prospects for the sake of frivolous pleasures like those of smoking, but was subjected to strong peer pressure to think differently. And while she may not have brought herself to sincerely agree with her hedonist friends on this issue, she may at least have successfully persuaded herself that she did agree with them, repressing her inconvenient feeling that their outlook was in fact irrational.

Now if we suppose that Influenceable* has been deceiving herself about her real views throughout her time at Uni, it seems appropriate to consider her decision to take

the cigarette at the party inauthentic, no matter for how long she may have convinced herself that she agreed with her hedonist friends. After all, self-deception has been treated as the polar opposite of authenticity by one of the thinkers whose name is most often associated with the latter concept, existentialist philosopher Jean-Paul Sartre. Sartre contrasts authenticity with bad faith, a concept closely related to that of self-deception, which describes a state that can occur in at least two different forms: when one regards oneself as pure transcendence and refuses to acknowledge that the actions one has freely performed in the past contribute at all to defining who one is – Sartre gives the example of a homosexual who, even when faced with evidence of his homosexuality, persists in denying that he is a homosexual in any sense of that idea (Sartre, 1957, pp.63-4); or when one "constitutes oneself as a thing" by denying one's essential freedom, suggesting instead that one is, for instance, an alcoholic in the same way a tree is a tree, and that there is nothing one can do about it (ibid., p.65). By contrast, authenticity for Sartre means not taking any of these routes of escape from the anguish that he identifies with the awareness of our essential freedom, but facing up to it.

The idea that decisions founded on bad faith or, more generally, on self-deception are thereby inauthentic is an important and plausible one. Yet nothing, it seems, prevents a supporter of the Frankfurtian account of authenticity from incorporating this idea, by simply arguing that wholeheartedness cannot co-exist with self-deception. Surely, one cannot wholeheartedly endorse a first-order desire if one does not genuinely have that attitude of endorsement, but has merely deceived oneself into believing that one did endorse the desire in question.

The Frankfurtian view does not support the authenticity objection to enhancement. It will of course declare inauthentic any particular use of enhancement technologies where the agent is not wholehearted. Yet since it seems perfectly possible for someone to wholeheartedly endorse her first-order desire to enhance herself, the Frankfurtian account will not extend the verdict of inauthenticity to *all* enhancement use.

2.3 David DeGrazia's view

It does appear that the Frankfurtian account of authenticity can adequately deal with the various cases we have considered so far. Nevertheless, other possible cases reveal its limitations. Consider the following one:

Indoctrinated. For ten years Indoctrinated has belonged to a sect that subjected its members to an intensive process of brainwashing. During these ten years he has had a constant desire to obey the demands of the sect's guru, including those of a financial nature. He has always fully identified with this desire, and this higher-order attitude did not conflict with any others he had.[14] Indoctrinated's relatives, however, eventually manage to take him away from the sect's environment for two weeks. They push him to critically think about his situation and talk to several "outsiders". He finally concludes that the sect has manipulated him, exploiting his desire to belong to a community. He severs all links with the sect, cancels his standing order to donate money, and no longer endorses the desires on which he had been acting for ten

[14] What about the likelihood that he may still have had a desire not to act on first-order desires resulting from manipulation? This desire, it might be argued, would contradict his endorsement of his desire to obey the guru, preventing the latter from being wholehearted. The problem is that the higher-order desire just described seems too general to undermine the wholeheartedness of Indoctrinated's commitment to the sect. To think the contrary would commit us to the conclusion that if I desire never to act on a first-order desire that would lead me to violate the requirements of practical wisdom, no action of mine that isn't fully virtuous can ever be wholehearted, autonomous, or authentic on the Frankfurtian view. Such a conclusion seems implausible.

years. He now wishes that he had never been lured into joining the sect.

It would not seem legitimate to argue that Indoctrinated's endorsement of his obedience to the sect was not stable enough to be wholehearted. Indeed, it is as stable as Ex-rebel's contrarian outlook in the example described above, which we have suggested is compatible with her being wholehearted, and thus making an authentic choice, when accepting the cigarette at the party. Also, we are assuming that Indoctrinated was not self-deceived during that period – he sincerely believed what his guru and the other sect members told him. As it stands, the Frankfurtian account must therefore also declare authentic Indoctrinated's choices and actions while under the influence of the sect, since by hypothesis they meet the wholeheartedness condition. Yet it seems more appropriate to deny that his choices and actions during that period were really authentic – not just because he came to regret them afterwards, but because even then they didn't reflect his *considered* preferences. Underlying this line of argument is the intuition that an agent's second-order preferences have to be formed under at least sufficiently good conditions, if they are to allow his choices or actions to count as authentic.

An important contemporary account of authenticity that purports to capture this intuition is that of David DeGrazia. His account builds on the idea, central to the Frankfurtian view, that authenticity requires a second-order identification with the first-order desires that guide one's actions. However, for DeGrazia this identification, even if wholehearted, will not suffice for either authenticity or autonomy. Also, rather than equating the two notions, he regards autonomy as a necessary (but not sufficient) condition of authenticity. He characterizes autonomous action in the following way:

[Person] A autonomously performs intentional action X iff (1) A does X because she prefers to do X, (2) A has this preference because she (at least dispositionally) identifies with and prefers to have it, and (3) this identification has not resulted primarily from influences that A would, on careful reflection, consider alienating. (DeGrazia, 2005b, p.102)

Conditions (1) and (2) roughly correspond to authenticity in the Frankfurtian sense, with the wholeheartedness condition left out (strictly speaking, they are therefore equivalent to authenticity in Dworkin's sense). Condition (3) adds a constraint on the formation of the second-order preferences relevant to authenticity: they should survive a process of critical reflection on their causal history. Yet DeGrazia's account is not just a revamped version of the Frankfurtian analysis of authenticity as autonomy. DeGrazia sets another condition for authenticity besides autonomy: honesty, in the sense of accurate presentation (both to others and to oneself) of who one really is.[15] This honesty requirement is reminiscent of Sartre's contrast of authenticity with bad faith, even though it is more wide-ranging, given that it makes authenticity incompatible with self-deception more generally (just like the Frankfurtian view), and also with deceiving *others* (in which respect it differs from the Frankfurtian view). DeGrazia's conditions for authenticity are therefore more demanding than those set by the account of authenticity as wholeheartedness. As a typical example of inauthentic behaviour, DeGrazia thus cites "a preppy East Coast schoolboy who suddenly dresses down and adopts a southern accent in order to impress a girl from Louisiana" (DeGrazia, 2005b, p.108). In addition to his emphasis on honesty, DeGrazia's insistence that self-creation can be a fully authentic enterprise further suggests the presence of an existentialist element in his view of authenticity, in

[15] See e.g. DeGrazia, 2005, p.112: "any self-creation project that is autonomous and honest is *ipso facto* authentic".

addition to the Frankfurtian core.

To come back to the cases previously described, DeGrazia's account and the Frankfurtian one both imply that Akratic's choice to accept the cigarette at the party is inauthentic. Indeed, she does not identify with her desire to have a cigarette, which violates both the wholeheartedness requirement and DeGrazia's second condition for autonomy. The two views also seem to lead to the same conclusion in Influenceable's case, namely that her decision is inauthentic, though they will do so for different reasons. A supporter of the Frankfurtian account might stress the transient nature of the agent's endorsement of her desire for the cigarette, as well as the likelihood that she is self-deceived or that her endorsement actually conflicts with some of her higher-order preferences, and add that such things are incompatible with wholeheartedness. DeGrazia's view, on the other hand, suggests that the problem with Influenceable's choice is about autonomy, but understood in a slightly different way than Frankfurt's: once she gets the opportunity to reflect critically on the influences behind her choice, she disowns it, thereby violating DeGrazia's third condition for autonomy. As for Rebel, the two accounts might or might not reach the same verdict in her case, depending on the exact assumptions we make about it. If we assume (as I did) that no amount of careful reflection on the part of Rebel about the influences behind her commitment to a contrarian, radically hedonistic lifestyle would lead her to disown this commitment, then both DeGrazia's and the Frankfurtian accounts will agree that her choice to have the cigarette at the party is authentic. By contrast, if in fact she would conclude that she had initially adopted this outlook merely to get the approval of some of her rebellious friends as a teenager, and that this is no good reason to embrace it, then Rebel's choice to have the cigarette would no longer count

as authentic on DeGrazia's view, as it would violate his condition 3) for autonomy. However, it could still be authentic according to the Frankfurtian view, provided that her change of heart were merely hypothetical (it would only happen were Rebel to engage in careful reflection on the relevant influences) and that her endorsement of the desire that guided her action met the wholeheartedness condition. Similar remarks apply to Ex-rebel, whose choice to accept the cigarette at the party in her younger days can count as authentic on the Frankfurtian view whether or not we assume that it meets DeGrazia's third condition for autonomy, whereas obviously it will not necessarily count as such on DeGrazia's view. DeGrazia's preppy schoolboy, too, might count as acting authentically on the Frankfurtian view if he wholeheartedly endorses his desire to impress the girl even if this means deceiving her, whereas DeGrazia treats him as a clear example of inauthenticity. Finally, the two accounts also yield different verdicts about Indoctrinated's choices and actions while under the influence of the sect, which are inauthentic on DeGrazia's account but authentic on the Frankfurtian view. In cases where the two views diverge, it seems to me that DeGrazia's is the one delivering the more plausible verdicts. As I shall discuss later, it is not clear that DeGrazia's understanding of "critical reflection" does in fact fully incorporate the above-mentioned intuition about the need for autonomy-conducive higher-order attitudes to be formed *under appropriate conditions*. Still, as it stands his view does seem to yield the correct verdict in a case like that of Indoctrinated.

DeGrazia's view does not, either, support the authenticity objection to enhancement. As we have seen, he holds that any self-creation project will count as authentic if it is both honest and autonomous. Various forms of enhancement use can certainly meet those conditions. (We will look at some specific examples in the final

part of this dissertation.)

2.4 Four challenges to the self-creation model

The account of authenticity as wholeheartedness and DeGrazia's view both have something to be said for them. Each of these views can make good sense of our judgments about authenticity relative to a specific range of cases. However, there are cases of still another sort that many would think are not adequately dealt with by these two accounts. Consider the following four scenarios:

Unconfident. Unconfident is studying musical composition under the supervision of Professor Arnold S., the leading figure of a well-regarded musical school. Unconfident happens to have ideas that would really break new ground in contemporary music. However, he is not sure whether he ought to include them into his compositions, as they are unlike any of the music he can find around him. Also, he finds it more challenging to create coherent pieces out of his own audacious ideas than by sticking to the canons of S.'s school. Unconfident eventually decides to be "reasonable" and to take the latter path, staying away from any "risky" endeavours. The results earn him accolades from his mentor and others, although he ends up being yet another follower of S. who doesn't stand out very much in the musical world.

Opportunist. Opportunist is an aspiring composer in all respects similar to Unconfident, with the exception that he is not intimidated at all either by the boldness of his ideas or by the prospect of working to give them coherent shape. Yet his mentor S., a rather close-minded figure, is of a different opinion. He tells Opportunist that only by sticking to the musical style he himself champions will he be able to fully realize his artistic potential. As the time comes to submit his graduation piece, Opportunist reflects that if he complies with S.'s recommendations, his life as a musician is likely to be much easier, as he will then have the

support of an eminent authority in the musical field. Even though he regards S.'s views as backward, he decides to "play it safe" and scraps everything in his work that was personal and original so as to fit his mentor's artistic outlook.

Penitent. Penitent is a 20-year old gay man. He is also the member of a religious group who regards homosexuality as an immoral perversion. Accordingly, the group preaches that gays and lesbians should "purify" themselves from their desires, allegedly imposed on them by the devil, through spiritual exercises and "reparative therapy". Penitent sincerely accepts the creeds of his religion, including this one, and loathes himself for his sexual orientation. He decides to embark on the supposed purificatory path recommended by his community. After several years of painful work on himself, Penitent eventually manages to get his homosexual inclinations completely under control. He still occasionally experiences them, but never acts on them. He enters a heterosexual marriage, and while he doesn't feel sexually fulfilled, he nevertheless develops a strong bond with his wife, who supports him in his efforts to keep his gayness at bay.

Ex-gay. Ex-gay is in all respects similar to Penitent, with the difference that he is offered the latest form of reparative therapy, more effective than any of its predecessors, which allows him to completely abolish his homosexual desires and replace them with heterosexual ones. With great relief, he can now pronounce himself "cured", and enjoy all the benefits of a purely heterosexual relationship.[16]

These four agents are all engaged in particular self-creation projects: they respectively choose to shape their artistic career, and some aspects of their

[16] In imagining such a case I do not mean to suggest that any such "cure" for homosexuality currently exists. On the contrary, the available evidence suggests that, far from achieving what they promise, such interventions actually tend to cause psychological harm. In order to fully test our intuitions about authenticity, however, I shall assume that the procedure in this second version of the case does succeed in changing Ex-gay's sexual orientation.

psychology, in a way they consider desirable. (Ex-gay will presumably regard his self-transformation as constituting an enhancement, but I have deliberately chosen cases that do not involve the use of enhancement technologies as standardly understood, since these cases will interest me in the final part of this dissertation.) I believe that their self-creation projects (with the possible exception of Unconfident) are morally problematic, and that all four are undesirable overall. And I would argue that one significant reason why this is so is that the agent is open to the charge of inauthenticity.[17] I shall concede that introducing other concepts than authenticity is necessary to fully spell out what is problematic especially in Ex-gay's case, though I shall also argue that, even then, the charge of inauthenticity nevertheless has the merit of drawing our attention to certain ethical considerations that are often neglected, for instance, in the enhancement debate. Finally, if my four agents had resisted the pressure and/or temptation to mould themselves a certain way, and had done so for the right reasons (I shall discuss later on what this entails exactly), they would constitute models of authenticity.

The view that Unconfident and Opportunist are inauthentic artists will, I think, generate significant agreement. Remember what I said at the beginning about the view many take of musicians who "sell out", as well as Taylor's words in *The Ethics* about the danger of losing our capacity to listen to our "inner voice". It is true that people will likely reach different verdicts about Penitent and Ex-gay's cases depending on whether they are liberally or conservatively minded. Those of a liberal orientation who are also sympathetic to the self-discovery model will tend to agree with me that these two agents' self-creation projects go against their true selves, and

[17] There are of course other reasons, e.g. the reinforcement of discriminatory attitudes towards homosexuals.

that this makes them, if not worthy of blame, at least something to be regretted. The authentic thing for them to do, they will say, is to accept themselves as they are, to reject their community's hostile attitude towards homosexuality (though not necessarily all the other aspects of their religious outlook), and to live their lives as gay men. Most people might be unwilling to blame these agents for what they do, for we are likely to see them as helpless victims of harmful socialization in an intolerant community that led them to hate themselves, without giving them the opportunity to hear other viewpoints and to think for themselves. Opinions might be more divided if we assumed that their disapproval of homosexuality, as well as their other religious beliefs, had been acquired in a fully autonomous manner. Let us suppose that their family and friends are supportive and accept them for who they are. Each of them, however, takes a different path from that of his liberal relatives after socializing with a group of religious conservatives. Even then, it would seem harsh to place much blame, if any, on them for changing themselves as they do. While their choice does strike me as regrettable and morally disturbing, it is nevertheless motivated by sincere – though intolerant and harmful – religious convictions, and the psychological suffering they are experiencing due to the conflict within them seems to make them worthy of some degree of sympathy. Yet on the other hand, if they were to embrace their gayness, condemning their community's attitude towards homosexuals as intolerant and possibly facing rejection from them as a consequence, many of us would praise them for being authentic persons.

I accept that people with a more conservative mindset, such as members of religious groups similar to the fictional one I have described, would probably disagree with my verdicts about Penitent and Ex-gay. On the contrary, they might argue, each

of them is a model of authenticity, and he would actually have betrayed his true self had he done what I have described as the authentic thing.[18] While I shall mostly rely on my own value judgments about those cases in the rest of this dissertation, it is not crucial to the analysis of authenticity I shall present that my judgments should be the correct ones. What is more important for my purposes is that many people, both of a liberal *and* a conservative orientation, will agree with me that the question whether such self-creation projects are authentic or not cannot be resolved simply by considering whether the agent's decision is autonomous (in either Frankfurt's or DeGrazia's sense) or honest (in DeGrazia's sense). In other words, if I am right, many people will agree that we have to appeal to value judgments about the sort of lives Penitent and Ex-gay are choosing to live if we are to adequately assess their self-creation projects from the perspective of authenticity.

So can Frankfurt's and DeGrazia's accounts justify the charge of inauthenticity in these four cases? Yes, but only if we make certain specific assumptions about them. As we have seen, both accounts treat self-deception as antithetical to authenticity. It is possible that Unconfident should be deceiving himself about the value of his bolder ideas. Since they are a riskier bet than is sticking to S.'s style of composition, he may well be motivated to persuade himself that these ideas are no good, when in his heart of hearts he actually judges them superior to those of S. This is possible, but not necessary: Unconfident may e.g. simply be too timid to believe that his own ideas are better, in which case talk of self-deception is unwarranted. Also, if Ex-gay's case were a real-life one, self-deception would likely

[18] That people with conservative beliefs tend to make such judgments is suggested for instance by recent research in experimental philosophy on the concept of the true self, which I shall discuss myself in part 3 of this dissertation: see Newman et al., under review.

be involved. He would merely have persuaded himself that he had changed – eventually, the truth of his homosexuality would catch up with him. On this construal of Ex-gay's case, appealing to the idea of self-deception provides a persuasive justification for the charge of inauthenticity: he hasn't really changed. Yet I am assuming that he *has* really changed, and become straight. As for Penitent, he doesn't fool himself into thinking that he is no longer attracted to men – he just keeps his homosexual inclinations in check for the rest of his life.

What then? According to the Frankfurtian account, each of these self-creation projects will still count as inauthentic if, while not deceiving himself, the agent doesn't wholeheartedly endorse the first-order desires that guide his decision. Unconfident, for example, might actually feel ambivalent about his desire to "play it safe" in his compositions, being at the same time attracted to the idea of breaking new ground in music. Or he may not have realized that his identification with his desire to take the safer bet actually conflicts with some other of his higher-order attitudes. In order to test our intuitions about my four cases, however, let us assume that they all meet the conditions for authenticity set by the Frankfurtian account: all four agents wholeheartedly endorse the motives that move them to act – which entails that they are not self-deceived. Even if we make that assumption, it seems to me plausible to retain the charge of inauthenticity in these four cases. If so, this shows that the Frankfurtian account cannot adequately deal with cases of that sort.

Similarly to the Frankfurtian view, DeGrazia's account will only justify the charge of inauthenticity in relation to these self-creation projects if we assume that they fail to meet one of its two conditions for authenticity. Perhaps the agents in

question are being dishonest in some way. Even though we have already ruled out self-deception, it might still be suggested that they are misrepresenting who they really are to others. But *need* they be doing so? Ex-gay might, of course, lie to others about his past, but we need not assume that he does – on the contrary, he might be fully open about his personal struggle and transformation, using his own story when preaching to encourage other homosexuals to seek therapy. Similarly, I see no reason to think that Unconfident must necessarily be misrepresenting who he is to others. He has decided to take what he regards as the "reasonable" path as a musician – why should he lie to anyone about this? It is admittedly more difficult to imagine that Opportunist could avoid lying about his artistic convictions, if he did not wish to risk alienating S. as well as many potential listeners. In real life, the charge of inauthenticity in a case like his would therefore likely have to do with dishonesty, as DeGrazia might suggest. Yet suppose that Opportunist never actually lies about his artistic views, but simply keeps quiet about them (until the day when he is in a strong enough position to speak the truth openly). S. has fully guessed the true nature of his views, yet he judges that the quality of his compositions justifies turning a blind eye on this – Opportunist, he thinks, needs to be pushed on the right path even though he does not yet realize that that is where he is treading. Even in that scenario, where the charge of dishonesty no longer applies, many would retain the charge of inauthenticity against Opportunist.[19] Concerns about the authenticity of such self-creation projects cannot be reduced to concerns about honesty.

[19] Could one argue that by presenting his new creations to the musical world, Opportunist is implicitly giving them his stamp of approval, thereby misrepresenting the true nature of his artistic convictions to his audience even if he doesn't explicitly lie about them? I think that such a suggestion stretches too far the idea of misrepresentation.

Here DeGrazia could still make room for the inauthenticity charge by questioning the autonomy of these projects. As we have said, maybe Penitent and Ex-gay were never exposed to people who would challenge the intolerant attitudes towards homosexuals prevalent in their community, but if they were, would be led to examine these attitudes with a critical eye, and eventually disown them. On that assumption, their self-creation projects would count as inauthentic for DeGrazia, as they would violate condition (3) above: they wouldn't be autonomous. But suppose that in fact my agents would not regard the influences behind their choices as alienating were they to critically reflect on them, and that their self-creation projects also meet the honesty condition. Then on DeGrazia's view, we have no grounds for declaring these projects inauthentic. To be fair to DeGrazia, this does not mean that he would necessarily judge these self-creation projects to be beyond moral reproach. He could for instance say that Ex-gay's project makes him complicit with morally problematic social norms. He would just add that such a worry is distinct from the authenticity objection.

Here again, *pace* DeGrazia, I think that worries about authenticity can still be raised in cases like these even if his conditions are satisfied. And while the particular worries I have raised do presuppose a commitment to liberal values, let me stress again that there is no need to espouse my own particular values in order to be dissatisfied with DeGrazia's account of authenticity. As I have mentioned, those of a conservative bent are likely to maintain not only that Ex-gay is a model of authenticity (something that does not contradict DeGrazia's analysis), but also that it would have been inauthentic of him to embrace his homosexuality and reject his community's disapproving attitude. I suspect that they would stand by the latter claim

even on the assumption that Ex-gay's decision to accept himself were wholehearted, honest, and autonomous in DeGrazia's sense.

If so, we need an alternative to the analyses of authenticity that we have examined so far, and I believe this alternative will have to proceed from the self-discovery model. DeGrazia does consider various possible construals of the charge of inauthenticity based on a self-discovery approach (to be discussed in the remainder of this dissertation), but as he states them they are all unpersuasive, and he rightly rejects them as such. I will try and give reasons to think that we can do better.

2.5 More stringent conditions for autonomy?

Before we move on to a self-discovery approach to authenticity, however, we may want to explore further the possibility of vindicating the charge of inauthenticity in my four cases by appeal to the concept of autonomy (assuming autonomy is required for, or even equivalent to authenticity). One might argue that Frankfurt and DeGrazia's conditions for autonomy are too lax, and that these four self-creation projects are in fact not autonomous. What sort of account of autonomy might one use to support this line of argument? The contemporary philosophical literature on autonomy is vast, and I can only take a cursory look at it here. Accounts of autonomy like those of Frankfurt and DeGrazia are often called "content-neutral" or "procedural", meaning that they remain neutral as to the content of the higher-order attitudes that ground autonomous action (Dryden, 2010). Because of that neutrality, it seems unlikely that a procedural account could rule out the possibility that these four self-creation projects, as I have described them, might be autonomous.

In response to this, we might try and turn to "substantive" accounts of autonomy, which set greater constraints than procedural ones on the sorts of choices and actions that can count as autonomous. Authors like Bernard Berofsky and Paul Benson have thus stressed the importance for autonomy of a sufficient ability to appreciate the various reasons to act relevant to the situation in which one finds oneself (Benson, 1991; Berofsky, 1995). Similarly, Susan Wolf has been read by many as proposing the view that autonomy requires "normative competence", that is to say, an understanding of the Good in the moral, aesthetic and other realms (Wolf, 1990, pp.121ff). Benson, for instance, argues that oppressive forms of socialization can damage people's competence at critical reflection, thereby compromising their autonomy, at least within a certain domain of their lives. Furthermore, possession of the critical competence necessary for autonomy does not merely require, for Benson, being able to *seriously consider* a sufficiently wide range of relevant reasons for action. It also demands that we *respond* to the right reasons to a sufficient degree. Considering the example of women who have internalized the sexist norm according to which they should strive to look appealing to men, Benson suggests that *the very acceptance* of such a norm reveals damage to one's critical competence incompatible with autonomy (Benson, 1991, p.394). Following a similar line, one might argue that a person cannot be autonomous if e.g. he fails to recognize that the attitude of Penitent's community to homosexuals is unacceptably intolerant, or that the grounds for their belief about homosexuality are inadequate.

Such a strongly normative conception of autonomy might well rule out the possibility that any of my four agents might be acting autonomously. The problem is

that it appears to set excessive conditions for autonomy, making the notion too similar to those of practical wisdom and intellectual rigour. It seems to me we should allow that one can, for instance, be an autonomous opportunist, or autonomously hold poorly grounded or irrational beliefs, even though such a person clearly fails to respond to the epistemic or moral reasons she has. True, Benson does accept that "persons do not suffer reduced autonomy merely because they have false beliefs about what they should do, or because they act contrary to what there are the best reasons for them to do". For instance, he says, "one can exercise the requisite critical competence and still arrive at a mistaken assessment of what there is reason for one to do. Human competence can fail to yield perfect performance" (Benson, 1991, p.397). However, Benson does not tell us more exactly how we are to distinguish between a case of that kind and one in which the person's critical competence has been damaged in a way that undermines her autonomy. Why must e.g. a woman who, even after carefully reflecting on the social determinants behind her own attitude, persists in endorsing the norm according to which women ought to look aesthetically pleasing to men, necessarily fall under the latter description rather than the former? Perhaps Benson would reply that if such a woman did possess the requisite critical competence, but simply failed to appropriately exercise it in the circumstances I have described, this would eventually be manifested in her revising her attitude after further reflection and discussion with other critically competent interlocutors. Persistence in her initial attitude is a clear indication of damage to the relevant competence. If so, however, I would reiterate the criticism I made above: such a line of thought places undue constraints on who can count as an autonomous agent, by implying that someone who keeps holding on to an irrational attitude of the sort just mentioned (irrational because unresponsive to certain moral or other reasons), even

after critically reflecting on it and on its causal history, cannot be holding that attitude autonomously.

That said, Benson's approach does seem to be on to something, and it points to some limitations of the procedural accounts of autonomy we have discussed so far. Suppose Penitent had grown up in the religious community of which he is now part. All the people he respects have always represented homosexuality to him as seriously immoral, citing the Bible as supporting their view. While he occasionally gets exposed to contrary views through the media, he then discusses them with fellow community members who vehemently attack them, listing their supposed flaws, and rejecting as misguided any interpretation of the Bible that contradicts their view on such matters. Is Penitent's higher-order rejection of his homosexual desires really autonomous in such a scenario? We may doubt that it is, on the grounds that while he is aware of contrary viewpoints, his circumstances don't allow him to give them a fair hearing, and that he has been taught to defer to the interpretation of the Bible offered by his community leaders rather than thinking for himself on such issues. It is not clear that DeGrazia's understanding of "critical reflection", for instance, takes such concerns into account. Talking about a fictional workaholic who fully identifies with his desire to work long hours every day, DeGrazia considers what this person would think "after much reflection and discussion with people he trusts" (DeGrazia, 2005b, p.100). But if that is enough for someone to have successfully engaged in critical reflection, then there is no reason why Penitent, in the scenario just mentioned, should change his mind were he to engage in such reflection: it is unlikely that he would be impressed with the criticisms of his religion as intolerant and epistemically ill-founded if he were to solely discuss these criticisms with members of his own

congregation. On such a liberal understanding of critical reflection, Penitent's disapproval of homosexuality would indeed count as autonomous on DeGrazia's view, yet we have found reason to question such a verdict. While DeGrazia's approach to autonomy does seem to me on the right track, it may need to set stricter conditions on what is to count as critical reflection, conditions ruling out forms of oppressive socialization that prevent us from considering alternative viewpoints with a minimum degree of openness and impartiality.

Yet while such a revision would bring DeGrazia's account closer to the substantive approach to autonomy, it would still remain procedural in nature. I believe the procedural approach to be preferable to the substantive one in that it does allow to adequately distinguish autonomy from other notions like practical wisdom, but the counterpart of this is that, on the sort of view of autonomy I have just sketched, my four agents can in principle count as autonomous. As we have said, Penitent and Ex-gay may actually come from a liberal background that allowed them to develop a reasonable ability for critical thinking and gave them substantial exposure to tolerant views about homosexuality, yet they may still end up rejecting such views as flawed. If we also assume that their religious community didn't brainwash them and isolate them from the rest of society, it seems to me that their self-creation projects can be autonomous.

I conclude, therefore, that my two agents can in principle be acting autonomously while still being inauthentic, in which case authenticity in the sense that I am after cannot be reduced to autonomy. Nor is authenticity a necessary condition of autonomy in any plausible sense. It might seem more plausible to think

with DeGrazia that authenticity *requires* autonomy as a necessary (but not sufficient) condition, but I am inclined to deny even this. Consider the case, often cited in discussions of so-called "inverse akrasia", of Mark Twain's character Huckleberry Finn.[20] Huck has helped his friend Jim, a slave, to run away from his owner. As they travel down the Mississippi river on a raft, Huck comes to feel scruples for what he has done; by helping Jim escape, he has broken his society's norms and feels he has wronged Jim's owner. He resolves to turn Jim in. But when the opportunity arises to do so, Huck finds himself unable to betray his friend – and blames himself very much for this. In refraining from turning Jim in, Huck, we may assume, acts contrary to his considered judgement, and thereby fails to meet the conditions for autonomy set by accounts like those of Frankfurt or DeGrazia. Indeed, Huck does not endorse the first-order desire that leads him to refrain; on the contrary, he strongly disapproves of it, as it violates the norms he has been taught to respect. Yet should we conclude on that basis that Huck acts inauthentically when he finds himself unable to turn Jim in? We may think, instead, that Huck's behaviour is in fact authentic, and worthy of some degree of praise, insofar as it manifests genuine qualities of character that even his constricting education in a racist society wasn't able to extinguish, namely his sense of loyalty as well as fundamental humaneness. In short, even if Huck's choice not to turn Jim in isn't autonomous, I would argue that it is nevertheless authentic: it expresses a valuable aspect of Huck's true self.

[20] Inverse akrasia can be characterized as the mirror image of the phenomenon of akrasia, or weakness of the will: both involve an agent acting against her all-things-considered judgment about what she ought to do, yet in the former case, contrary to the latter, the course of action she chooses is actually superior to what her considered judgment would recommend. For discussions of Huck Finn's case as a potential example of inverse akrasia, see e.g. Bennett, 1974; Audi, 1990; and Arpaly, 2000. For a different take on the Huck Finn case, see Levy, 2011b.

2.6 "True self" accounts of authenticity

A more promising justification for the inauthenticity charge in my four scenarios, I would suggest, is thus provided by what I shall call "true self accounts" of authenticity. The essence of this family of accounts is not just that they include the popular characterization of authenticity as the quality of being true to oneself. DeGrazia, too, accepts that authenticity involves being true to oneself, but for him this is just another way of saying that it requires being honest with oneself (DeGrazia, 2005b, pp.108-9). The sense that true self accounts give to that idea is a different one: being true to oneself, on this analysis, entails being faithful to a "true" self that fits the description of the pre-given self offered above. This means that true self accounts fall under the self-discovery model of authenticity, whereas Frankfurt's and DeGrazia's views fall under the self-creation model. Indeed, we have said that on the Frankfurtian view, what counts as an authentic choice or action is entirely determined by whether or not one wholeheartedly identifies with the first-order preferences that guide it. No such preferences can have any weight as far as authenticity is concerned if we do not identify with them.[21] As for DeGrazia, we have seen that he rejects the idea of a pre-given self, since for him identification with any particular feature is a necessary condition of that feature being definitive of who we are. Accordingly, authenticity for him does not mean obeying the requirements made by such a hypothetical self.

[21] One might ask here whether *second-order* attitudes might not still constitute something like a pre-given self on the Frankfurtian account, assuming it does not require that we identify with *those* attitudes as well. But even if the answer is yes (which might be disputed), we can still see the difference between the Frankfurtian view and the self-discovery model by considering that authenticity on the former view is never a matter of conformity to *first-order* desires, or to any traits other than higher-order attitudes, if we do not endorse those features, whereas it can sometimes be so on the self-discovery model. Also, authenticity on that model can require respecting certain features even if we happen to *repudiate* them, whereas this can never be the case on the Frankfurtian view, since such repudiation is clearly incompatible with wholeheartedness.

Now what exactly does it mean to be faithful to one's true self? I believe the idea tends to be understood in two main ways, i.e. as entailing either:

a) Making the features constitutive of our true self, such as our personality traits, beliefs, feelings, or personal convictions, manifest in our behaviour when appropriate opportunities present themselves, rather than repressing them (and *sometimes*, as a result, misrepresenting who we really are, to others or to ourselves).

b) Declining to change such features when the opportunity to do so arises.

Failing to be faithful to one's true self in sense b) will sometimes entail that one is also doing so in sense a), namely when the trait one is changing counts as an "inner voice" of some kind. This relation of entailment, however, need not always hold, assuming the features constitutive of our true self can also include things that cannot plausibly described as an inner voice: our physical features, for instance. In cases where the relation of entailment does hold, the claim that we ought to be true to ourselves seems to imply that we ought to do so in both sense a) and b). That is to say, I should not just refrain from tinkering with my inner voice. I should also make sure I express it. (Preserving it while still repressing it would not be enough for authenticity.)

Sense a) covers DeGrazia's honesty condition, but goes beyond it, since it seems possible for instance to repress one's true self without thereby being dishonest:

we are assuming that this is exactly what Unconfident and Opportunist do. Penitent, too, fails to remain true to himself in sense a). Ex-gay is also likely to do so before he eventually manages to change his sexual orientation. After the change, even if from then on he always remains faithful to his (new) true self by living the life of a straight man, he will still count as having betrayed his true self in sense b), i.e. as having changed his original sexual orientation.[22]

We now want to ask what our "true self" is exactly. I will propose my own analysis of that notion in a subsequent section, but for the moment let us confine ourselves to what supporters of true self accounts typically associate with it. A common assumption seems to be that the true self involves a set of characteristics that play a key role in defining who a person is. Now this set of key features might be taken to be something like an essence, and true self accounts are sometimes referred to as "essentialist", to contrast them with views of a more existentialist spirit such as DeGrazia's (see e.g. Bublitz and Merkel, pp.360-1). For the reasons I have mentioned previously, I prefer not to use that term, which I find rather misleading. Elliott, for instance, accepts the idea of a true self as valid but does not understand it as an essence. I shall take a similar line myself (while acknowledging that *some* aspects of the true self might be essential in nature).

It is admittedly *possible* to take an essentialist view of the true self. This might be what the President's Council on Bioethics is doing in *Beyond Therapy*. Though the

[22] Could sense b) not be expressed using sense a) only? One might for instance argue that once Ex-gay has changed his original sexual orientation, he can no longer express it or live it out. Therefore Ex-gay is failing to be true to himself in sense a) while living his life as a straight man in scenario 2. But this suggestion sounds unpromising. Admittedly, Ex-gay is then not expressing his homosexuality, yet he is no longer gay! Demanding that he express his gayness in those circumstances seems to violate the "ought implies can" principle, since he can hardly express a feature he no longer has.

authors hardly use the term "authenticity" at all, we have seen that they do worry about a possible threat to our "identity" from enhancement technologies. And they sometimes appear to cash out this threat by suggesting that such technologies, at least if used in certain ways, might simply bring an end to our existence, as suggested by their talk of "turning into someone else" in the passage quoted above, or when they write that "[i]t is doubtful, to say the least, that biotechnical transformations of our bodies—or minds—will contribute to our realizing this goal [i.e. achieving excellence] *for ourselves*" (The President's Council on Bioethics (U.S.), 2003, p.149).[23] Such passages seem to fit with an essentialist view of the self: if the features targeted by enhancement technologies, such as personality traits or cognitive capacities, have the character of an essence, then modifying such features must mean putting an end to the existence of their possessor (and somehow replacing him with a different, though very similar person). The closest analogue to such an outcome that I can think of is Derek Parfit's "Simple Teletransportation" scenario, in which a certain person (call him S) has all of his molecular information recorded by a futuristic machine, which destroys him in the process and then recreates, out of new matter, an identical replica of him, S* (Parfit, 1984, pp.199-200). In Parfit's scenario, however, what makes it plausible (though not everyone will agree about this) to think that S has been destroyed and replaced by another person is the fact that there is no bodily continuity between S and S* – the former's body is destroyed, while the latter comes into existence from fresh material. The enhancement technologies I have been mentioning, by contrast, do not seem to pose any such threat to our bodily continuity.

[23] See also the footnote on the same page: "No sane person...would choose to be the fastest runner on two legs if it required becoming an ostrich. And few people would choose to acquire someone else's perfections of body or mind on condition of becoming that other person. Who, in the event of such self-transformative improvements, would we say now enjoyed them?". I agree with DeGrazia that the President's Council seems to fail to properly distinguish here between numerical and narrative identity: see DeGrazia, 2005, p.232.

And even if they did, it would seem somewhat misleading to then speak of someone *changing* his cognitive capacities, for instance. If the pre-enhancement person is not numerically the same as the post-enhancement one, the former has not so much changed his capacities as *destroyed* himself, cognitive capacities and all! True, one might perhaps argue that even destruction counts as a form of change. But even if we agreed to use the verb "change" in this unorthodox manner, it would still remain that the person could hardly be said to have *enhanced* herself, or her cognitive capacities. The very idea of a person enhancing some of his traits presupposes that he remains numerically the same throughout the enhancement procedure. For that reason, the possibility of enhancing *essential* traits seems a priori ruled out. Of course, it *could* in principle have been the case, for instance, that a different sperm might have fertilized the particular egg from which we originated, possibly resulting in an individual with greater cognitive capacities than ours. And having come from a certain sperm and egg is arguably an essential trait of ours. Yet had this happened, it would not have *enhanced* or even *changed* us compared to whom we are now. Rather, we would simply not have come into existence at all – a different individual would have instead.

A true self account of authenticity that relies on an essentialist view of the self thus does not offer a promising basis for an objection to self-creation projects, given that few such projects would seem to pose a threat to our numerical identity, even when they involve radical self-transformation. Mind uploading, defined as "the...process of transferring the mental structure and consciousness of a person to an external carrier, like a computer"[24] might perhaps be one such case, but I shall not consider such radical procedures here. It is true that even in more mundane cases of

[24] Definition given by transhumanist Anders Sandberg on his personal website: URL= http://www.aleph.se/Trans/Global/Uploading/#INTRO [accessed 2/3/2013].

self-transformation, people sometimes say that an acquaintance "is no longer the same person" they used to know, for instance after a religious conversion has very much changed her. Yet such a way of speaking is best understood as referring to narrative, not numerical identity. We do not usually assume that a religious conversion consists in someone's being literally destroyed and replaced by a clone with a new set of beliefs, no matter how different the new person might be from the pre-conversion one in terms of his personality and outlook on things. The same holds for procedures like personality modification, mood manipulation, or cosmetic surgery. And surely worries about numerical identity are misplaced in the four cases I have described at the start of this section. I will therefore leave essentialism aside in the coming discussion.

Once we have abandoned an essentialist view of the true self, what alternatives are we left with? In *The Ethics*, as we have seen, Taylor talks about an inner voice or an inner nature that tells us what is our own unique way of realizing our humanity. While sometimes citing Taylor, Elliott, as we have seen, also refers to the idea of a person's "core", suggesting that the true self might be understood as a set of traits which together constitute that core (this is the conception of the true self that DeGrazia relies on in his critique of the authentity objection). Elsewhere Elliott also describes the true self as "a relatively stable set of mental and physical attributes" encompassing not essential traits, but things such as someone's ethical and religious ideals, and personality or character traits (Elliott, 2003, pp.50-1). This approach strikes me as much more promising than the essentialist line.

I have suggested that the true self approach to authenticity can, contrary to Frankfurt and DeGrazia's views, support the charge of inauthenticity in my four cases. It is also the approach that authors sympathetic to the authenticity objection to enhancement, such as the members of the President's Council on Bioethics, tend to favour. Yet it is also possible to endorse this approach while rejecting the authenticity objection in the sweeping form in which I have characterized it – this is actually the position I shall defend. To determine what the true self approach implies in relation to the cases we have considered prior to those four, we would need an account of authenticity more substantially developed than those that have so far been offered by supporters of that approach. I shall attempt to develop precisely such an account in the next part of this dissertation. For the moment, let me note that abandoning an essentialist understanding of the true self allows us to also abandon the implausible idea that any of my four agents is putting his numerical identity at risk. How, then, are we to spell out what is wrong with their self-creation projects from the point of view of authenticity? The following three explanations might be offered, using both senses a) and b) of the idea of faithfulness to the true self (while the essentialist line focused on sense b)):

1. *The inauthentic agent turns away from her "inner voice" or nature*. This proposal follows the Taylorian line already sketched. The view of authenticity he presents in *The Ethics* is naturally interpreted as an instance of a true self account, one that might support the charge of inauthenticity in my four examples.[25] My agents, it might be argued, are not listening to the "inner voice" which tells them, in the cases

[25] Though as I have already noted, this interpretation may not reflect his actual view on the matter. Rather than trying to ascertain how exactly Taylor himself meant his remarks to be read, I will concern myself here with what his view entails *if* interpreted as a true self account.

of Unconfident and Opportunist, to realize their own audacious ideas in their compositions, even if S. disapproves of them; and in the cases of Penitent and Ex-gay, to "come out" and live their life as gay men. These four agents all fail to listen to the inner voice telling them what a good life means for each of them, a voice they ought to have taken as their guide. One might also phrase it slightly differently and say, particularly in relation to Penitent and Ex-gay, that they are ignoring or turning away from their true "nature".

2. *Changing "core" traits that should not be tinkered with.* This solution chiefly applies to Ex-gay's case (since the composers don't *change* their creative powers, but simply fail to make real use of them in their music; and Penitent is changing his behaviour rather than his sexual orientation itself). The present suggestion, contrary to the previous one, focuses solely on the idea of *changing* the true self, leaving out the idea of repression, and no longer identifies the true self with an inner voice or nature but, more broadly, with a set of "core traits" (which need not be psychological ones). The idea, as DeGrazia states it in *Human Identity and Bioethics*, is that while such core traits are not part of some individual essence, they are nevertheless fundamental aspects of people's narrative identity, so fundamental that they are inviolable – in other words, it is wrong to tinker with them even in the pursuit of important life goals. Remember Elliott's suggestion, quoted early in this dissertation, that it might be morally problematic to change "the very core of what the person is". DeGrazia also cites the President's Council's allusion to "a *human* "givenness," or a given humanness, that is also good and worth respecting" (The President's Council on

Bioethics (U.S.), 2003, p.289, quoted in DeGrazia, 2005, p.233).[26] I shall discuss DeGrazia's critique of this proposal in the next chapter of this dissertation, as well as several other variants that he does not consider (including one stressing the supposed value of "natural" traits).

3. *Producing new traits that are not really the agent's "own"*. As we have seen above, besides the idea that cosmetic neurology and other forms of enhancement threaten our "identity" in some sense, Elliott also suggests another way of construing the authenticity objection when he writes that the new personality Prozac would give me would not really be "my" personality (Elliott, 1998, p.182). This line might also be applied to the case of Ex-gay. It could of course be understood as a concern about numerical identity: the person with the new personality or sexual orientation would not really be *me*, but a numerically distinct individual. Yet we have already noted the implausibility of interpreting the charge of inauthenticity along those lines. In response to this, an alternative way of developing Elliott's suggestion would be to say that even though someone like Ex-gay remains numerically the same person after changing himself, his new, deliberately induced sexual orientation is not his "real" one, but is somehow "fake". This would justify regarding his self-creation project as inauthentic. A third possibility would be to maintain that while the new feature the agent is acquiring might not be "fake", it nevertheless isn't *uniquely his*. The personality I might acquire on Prozac, it might be said, would be a standardized "Prozac" personality, in contrast to the one I originally had, which uniquely identified me as an individual. Whatever the plausibility of this third suggestion in the Prozac case, however (an issue I shall consider later on), it doesn't seem to apply to Ex-gay's

[26] I will assume that the idea expressed in this passage is not just a re-statement of the worry about numerical identity that we have now left aside.

case. Indeed, there is no sense in which his original homosexual orientation was "uniquely his own" – even though it was a minority orientation, he still used to share it with millions of other men worldwide.[27]

The composers' self-creation projects, on the other hand, are about a choice of artistic direction, not about the acquisition of new traits. Yet we could plausibly describe Opportunist's new musical orientation as "fake": perhaps he is deliberately misleading others regarding his true artistic preferences (in which case DeGrazia's honesty condition would provide an adequate justification for the fakeness charge), or at least these preferences are simply not being expressed as they should be in his compositions. The latter construal, however, would seem to lead us back to the idea that Opportunist is acting inauthentically because he is failing to listen to his inner voice. The same would apply to the (plausible) suggestion that the two composers are turning away from what is "uniquely theirs", namely their own unique ideas about music, becoming instead yet another avatar of S.'s musical outlook.

Regarding specifically the charge of producing fake traits, let me stress that it need not necessarily be grounded in a true self analysis of authenticity. True, the accounts of authenticity we have been looking at so far were mostly concerned with our *choices* and *actions*, and by extension, with the authenticity of our lives – to the extent that an authentic life involves making authentic choices and actions with enough consistency, perhaps especially on key occasions, such as when choosing to embark on a particular career. They were not primarily meant to offer criteria for

[27] Of course, there might be unique ways of *living out* a particular sexual orientation. Yet there is no reason why Ex-gay's new orientation should necessarily be less uniquely his own than his initial one in this sense either.

assessing the authenticity of our various *features*, including psychological ones like sexual orientation. However, these accounts could easily be extended to cover these things as well. The Frankfurtian line, for instance, would entail that a psychological feature is authentic if the agent wholeheartedly identifies with it, and inauthentic if not; and, it might be added, inauthentic features are necessarily "fake". DeGrazia's analysis also offers a natural criterion for the "phoniness" of certain traits: namely, the agent might merely be pretending to have the relevant traits, and be misrepresenting who he really is. Nevertheless, when applied to our psychological features in this way, the Frankfurtian analysis looks rather implausible. Let us take the emotion of fear as an example. Suppose I experience fear in a situation where I would rather not experience it (say, a job interview). Accordingly, I do not identify with my feelings. Surely this mere fact is not enough to call my feelings of fear "fake". Furthermore, neither the Frankfurtian line nor DeGrazia's analysis can provide a criterion of phoniness that might be used to criticize a self-creation project like Ex-gay's, if we assume that he wholeheartedly endorses his new state post-therapy, and is not misrepresenting who he is, either to others or himself. That said, still other criteria of phoniness might be put forward that do not commit us to a true self analysis, and I shall look at these in a subsequent section.

As we shall see, the construal of the inauthenticity charge according to which someone like Ex-gay develops a new feature that isn't really "his own" can avoid some of the objections that might be levelled at the two other analyses (number 1 and 2 above) we have previously considered. However, I shall argue that all three variants encounter problems. Let us now take a closer look at each of these.

2.7 Difficulties with the "true self" approach

2.7.1 The idea of repressing one's "inner voice" or nature

Let us begin with the idea that my four agents fail to listen to the "inner voice" indicating to them how they ought to live. This suggestion does seem to me to be on the right track, but it needs to be adequately spelt out. It doesn't seem satisfactory to hold that *the mere fact that one is failing to listen to one's inner voice or nature* is enough to render a self-creation project ethically problematic, because inauthentic. Indeed, nothing in existing true self accounts appears to rule out the possibility that this inner voice might ask us to perform evil, or otherwise despicable actions. Think of a sadistic killer, who enjoys torturing and murdering innocent people without experiencing any remorse as a result – someone like Dennis Rader, the infamous "BTK killer" (for "Bind, Torture, Kill", which sums up the treatment he inflicted on his victims). Let us suppose that this person could take a safe drug that would allow him to develop a minimal sense of empathy for others, leading him to refrain from hurting innocent people and animals for fun and removing his disposition to enjoy such activities, or at least giving him the motivation to work on his sadistic urges with the help of a therapist. Also, we can further assume that the drug would not turn him into a meek dove and weed out all aggressive inclinations in him. It would simply make him a minimally "civilized" individual, no longer posing a threat to others. Should this person, in the name of authenticity, listen to the inner voice telling him to keep hurting others and forfeit the possibility of moral improvement? To claim that he

should sounds very implausible, yet this is exactly what we will have to do if we develop this line of thought in the way just suggested.[28]

One might, however, try to mitigate this implausibility by adding that authenticity merely gives the sadist a *reason* not to get rid of his murderous disposition by taking the drug, but that this reason is (given the nature of the case) very weak, and largely outweighed by the other-regarding reasons he has to take the drug. Yet even this more modest claim fails to persuade me. The mere fact that taking the drug would mean failing to listen to his sadistic inner voice would arguably not be enough to provide the sadist with as much as a reason not to take it. I concede that he might have a *prudential* reason – which could be understood as authenticity-based – not to take the drug, yet this will only be true on *some* ways of construing his case: suppose that as a result of taking it, he could expect his long-term subjective well-being to be compromised, perhaps because he would feel unfulfilled in the absence of his original sadistic urges and of the pleasure he took in satisfying them. He might also start feeling the unpleasant emotion of guilt. But if taking the drug could on the contrary be expected to promote his future happiness, the sadist will no longer have any reason not to take it.

What if one argued that the sadist actually has a broadly *moral* (or alternatively, quasi-aesthetic) reason to remain as he is – though of course one that is outweighed by other moral reasons? This seemingly paradoxical claim might be justified by arguing that even someone like the sadist has a duty of self-respect.

[28] To be clear, I am not claiming that those committed to a true self approach to authenticity would usually defend, on grounds of authenticity, the sadist who refuses to take the morality pill. As I shall explain later, they are more likely to be implicitly assuming that the inner voice we ought to listen to is a *good* one. All I am trying to do here is to explicitly articulate the normative constraints that a plausible ethic of authenticity should incorporate.

Imagine the following scenario: the sadist gets arrested for his activies and is incarcerated. The tyrannical governor of the prison where he has been confined wants to force him to take the "morality pill", not because he cares in the least about the potential harm that this man might cause if he escapes (and about the threat he poses to guards and other inmates), but simply because he enjoys asserting his authority. In this scenario, the sadist may well have a moral (self-regarding) reason to refuse to take the pill, even though all things considered we can agree that he ought to take it: i.e. he would thereby express his refusal to obey the arbitrary whims of the prison governor (and let us assume he could not do so while, unbeknownst to others, taking the pill). I agree with this point, and in the next part of this dissertation I shall actually offer an analysis of my four initial cases that gives a central place to the concept of self-respect. However, the point only shows the sadist could have a moral reason to refuse to take the pill. It does not show that he would have an ultimate moral reason to *listen to his cruel impulses*. Arguably, the sadist would have no such reason. On the contrary, moral reasons in such a case would require him to do everything he could not to act on these impulses and, ideally, to get rid of them somehow (if not by pharmacological means, then e.g. through therapy). Secondly, the "tyrannical prison governor" is only one possible variant of the sadist case among others, and if the governor were actually a fully reasonable person who only wanted to make the sadist take the pill on the grounds that he posed too much of a threat to other people inside the prison, then arguably the sadist would no longer have a moral reason to resist him.

An alternative way of defending the claim that we always have at least a *pro tanto* reason to listen to our inner voice would be to argue that my imaginary sadist is suffering from a personality disorder (perhaps from what has sometimes been called

sadistic personality disorder), and that a person's "inner voice" can never be identical with a mental disorder. On the contrary, when mental disorder strikes, it always stifles the person's inner voice. Just as we typically regard a depressive episode as overshadowing the person's true self, not as expressing it, it might be argued that we should treat the sadist's impulses as the symptoms of a disorder and therefore as hiding rather than reflecting his inner voice. If so, the sadist's case does not constitute a counterexample to the analysis of authenticity we are considering. But such a reply would be unconvincing. Indeed, we may reasonably dispute the claim that a trait cannot reflect a person's authentic self if it is part of a psychological disorder. Various people who have struggled with mental disorder, whether it be ADHD, anorexia, or even depression, have taken the view that their disorder formed part of who they really were. As a result, they saw medication as suppressing an important part of themselves, sometimes quite an interesting and creative part.[29] While it is true that *some* cases of depressed mood seem best construed as involving a stifling of the person's true self, this interpretation is not as plausible in other cases of mental disorder, including cases of depression, that are structurally different. The view that a depressed mood stifles rather than reflects the person's true self appears especially plausible if that person has consistently shown a higher mood level throughout her life, before experiencing a sudden drop. The view seems less convincing when applied to cases where someone's baseline mood has always lied below the level defined as "normal" by the medical profession, and where we cannot point to some independently identified physiological dysfunction as the cause of the person's low mood. Similarly, even if the sadist's urges are indeed the reflection of a personality disorder, it seems that they can still plausibly count as representing his "inner voice".

[29] For reports of such a view in relation to depression, see Karp, 2006, pp.112-16; in relation to anorexia, see Hope et al., 2011, pp.24-5; and in relation to ADHD, see Bolt and Schermer, 2009, p.105.

We should therefore, I believe, accept that people do not always have so much as a *pro tanto* reason to listen to their "inner voice". As I said, I agree with the suggestion that my four agents are in some sense failing to listen to their inner voice, and that this is a problematic aspect of their self-creation projects, yet we cannot explain why it is problematic simply by stating that we always have at least a reason to listen to our inner voice. One way of responding to this would be to continue to define authenticity as the quality of living in accordance with our inner voice, no matter what this voice might say, but add that while some forms of inauthenticity are problematic, as illustrated by my four scenarios, not all of them are – some, like the inauthenticity the sadist would manifest were he to resist the promptings of his sadistic inner voice, are actually good, even praiseworthy. Call this a *descriptive* form of the true self analysis of authenticity. On the descriptive approach, someone like the sadist can count as authentic if he listens to his sadistic inner voice, even though he has no reason to do so, and even though his action's being authentic isn't even something to be said in its favour, something that makes it respectable or admirable in at least one respect. Contrast this with what I shall call the *normative* approach, which posits that an action's being authentic is at least something good and praiseworthy about it.[30] According to the normative approach to authenticity, while the sadist might be listening to his inner voice in refusing to take the pill, he nevertheless doesn't count as acting authentically, because he deserves blame, not praise of any kind, for doing so. Simply defining authenticity as the quality of listening to our inner voice appears implausible if we adopt the normative approach, but not if we opt for the descriptive one.

[30] Mikko Salmela also distinguishes between a descriptive and a normative sense of authenticity in Salmela, 2005.

Which approach to authenticity should we prefer? The descriptive one is certainly coherent, and I see no knockdown argument against it. Ultimately, which approach we choose to follow seems to be largely a matter of semantic stipulation. No consensus exists about the exact meaning of "authenticity" by appeal to which we could reject either choice as illegitimate. However, we may remember that someone like Taylor characterizes authenticity as a moral ideal, which he defines in turn as "a picture of what a better or higher mode of life would be" (Taylor, 1991, p.16). I am not sure that authenticity always constitutes a specifically *moral* ideal,[31] yet it does seem that many people use the concept so as to imply that a certain life is good, even admirable, in some way. Elliott concurs, identifying an authentic life with "a good life, a meaningful life, a life properly oriented toward the good" (Elliott, 1998, p.182). When used in this way, the concept clearly indicates *praise* – as it does e.g. if we say that Ex-gay would be living a really authentic life if he refused to change his sexual orientation. The descriptive approach to authenticity, however, fails to capture the intuitions behind this common use of the concept. The problem isn't necessarily its implication that there are authentic actions we have no reason whatever to perform. It might be that we should accept that implication. To take an analogy, suppose it is possible to show courage while fighting for a bad cause in a war. If so, the soldier who shows such courage may not have had any reason to perform the courageous actions he is now performing. Maybe all his reasons for action spoke in favour of becoming a conscientious objector. Even if authenticity is like courage in this respect, however, there remains force to the intuition that an action (or a person, or her life)'s

[31] Except perhaps if we understand "moral" in an unusually broad sense, such as the one described by John Mackie in his book *Ethics: Inventing Right and Wrong*: "A morality in the broad sense would be a general, all-inclusive theory of conduct: the morality to which someone subscribed would be whatever body of principles he allowed ultimately to guide or determine his choices of action" (Mackie, 1977, p.106).

being authentic at least makes it *prima facie* good or admirable, just like the soldier's courageous acts are arguably morally admirable in one respect, even if he has no reason to perform them. The normative approach to authenticity, unlike the descriptive one, does incorporate this intuition. We may take this to be grounds for preferring the former.

If we are to develop a more plausible variant of the normative approach than the candidates we have considered so far, it seems that our definition will need to place certain constraints on the aspects of our true self we ought to be faithful to. One possible suggestion would be that authenticity consists in listening to our inner voice if it is *healthy*, not disordered. This view resembles one we have previously discussed, but is different in that it accepts that someone's inner voice can be disordered; it just maintains that if it is, listening to it does not count as authentic. However, this suggestion still shares a defect with the previous one: namely, the fact that some feature of ours might be regarded as the expression of a mental disorder does not seem enough to make it irrelevant from the perspective of authenticity, even on the normative understanding of the notion. Think for instance of people with so-called "gender identity disorder", who feel that they were born in the wrong body (they might e.g. feel that they are a woman trapped in a man's body) and often request sex reassignment surgery in order to align their biological sex with the gender (male or female) they identify with. As Neil Levy notes, many of us today are sympathetic to the idea that transitioning to the opposite sex is the authentic thing to do for such people, a necessary step towards a good life in their case. It may allow them to "become who they really are", even though the procedure means complying with the demands of a supposed disorder (Levy, 2011a, pp.7-9). Similarly, as we have seen,

some people on medication for ADHD choose, in the name of authenticity, to get off it when they can, and it is not clear that their view is indefensible.

What about stipulating instead that authenticity only requires us to listen to our inner voice if it is not evil, i.e. if it does not demand that we inflict significant harm on innocent people? The inner voice of my four agents, it might be argued, is not evil, which is why there is at least something good about listening to it. While this suggestion appears on the right track, and does allow to avoid the sadist problem, it still faces another, more serious difficulty. How, we want to ask, are we to identify the person's "inner voice" in each of those cases, and on what grounds can we rule out competing candidates for what that voice might be that would lead to a different verdict about the authenticity of these self-creation projects? After all, Opportunist is guided by a strong desire for professional success; Unconfident, by the desire not to take too many risks from a purely aesthetic perspective; and Penitent and Ex-gay want to live a life that is "pure" and good in the light of their religious convictions. What prevents us from regarding each of these desires as part of the agent's true self? These desires cannot plausibly be described as evil, and neither can Penitent and Ex-gay's religious convictions – their congregation, I am assuming, doesn't advocate the eradication or mistreament of homosexuals, but simply asserts that they have a moral obligation to "cure" themselves. (We can further assume that no gay member of the congregation is ever coerced into getting therapy, though there are social pressures to do so.) Why shouldn't we say, then, that these four agents are actually being in touch with their inner voice, and are thus acting fully authentically when they act on such desires? This point is forcefully made by Neil Levy in a recent paper, on the basis of which he further argues that the self-discovery model does not support the

authenticity objection any more than the self-creation one does (Levy, 2011a). While disagreeing with some aspects of Levy's position, I believe he is right that such cases involve competing candidates for what is to count as the person's "inner voice", and that existing true self accounts do not tell us how to decide between those candidates. The normative constraints we have considered so far did not prove helpful in this regard. We therefore need to look for a different one.

As we will see in the next part of this dissertation, recent work in experimental philosophy suggests that one common way in which people tend to resolve apparent conflicts within the true self is by ruling out one of the competing candidates on the basis of their own value judgments. That is to say, they seem to attribute the features they value more to the person's true self, and to exclude from it the features that conflict with those. Despite its appeal in some types of cases, I shall suggest that such a way of resolving those conflicts is problematic, yet this will require an extended discussion that deserves a section of its own.

Some people might respond to Levy's challenge by countering that a process of soul-searching, if undertaken carefully enough, will always eventually reveal to us that only of one the features in conflict within us reflects our true self, the other one being "fake", e.g. a belief or desire implanted into us through social conditioning that distorts our true nature. In a piece for the *Huffington Post*, New Age guru Deepak Chopra thus critically comments on the research in experimental philosophy just alluded to, on the concept of the true self. While granting that discovering one's true self is likely to be a lengthy and challenging endeavour, Chopra still maintains that

once one has completed the required introspective process, no doubt will be left concerning which features actually reflect our true self:

> "Know thyself" doesn't mean taking a 30-minute quiz. It means going through a lifelong process of self-reflection, contemplation, and questioning. The point is that when this journey is taken seriously, the opposites within ourselves are resolved. The war between reason and unreason exists at many levels of the self, but it doesn't exist at the level of the true self... For a conflicted gay Christian, such a journey seems far more promising that batting him around between various opinions, right and left. (Chopra, 2011)

While Chopra's view might be shared by many, the assumptions on which it relies seem questionable. There is no doubt that the sort of inner journey Chopra recommends can bring significant benefits. Some people might well find, by looking deep into themselves, that their self-image had been distorted by the expectations of those around them, and that they had allowed their lives to be guided by social pressures rather than by their own deepest desires and values. They might also get clearer about what their fundamental values and priorities in life really are, and achieve Frankfurtian wholeheartedness as a result. Yet this does not mean that embarking on such an inner quest will *always* resolve the inner conflict between different core features of the person. It certainly seems possible that a conflicted gay Christian might, even after carefully looking inward, still retain both his homosexual orientation *and* his religious beliefs condemning it, without deceiving himself about any of these features. Chopra appears to rule out such a possibility, but he gives us no reason to think we should – he simply asserts that no such conflict can then persist. Perhaps he is assuming, in keeping with the self-creation model of authenticity, that if I find, after a lengthy introspective process, that I do not identify with some feature of mine, then that feature is not part of who I really am. But such an assumption is precisely what the self-discovery approach is denying.

Another possible reply to the challenge raised above would be to point out that the agents we are talking about both decide to comply with the social norms that they are being exposed to. Such compliance, it might be said, is incompatible with a real contact with our inner voice. Each of the composers represses his unique ideas and becomes just another representative of S.'s school of composition. Penitent and Ex-gay fight off their homosexual orientation, which would otherwise have put them on a non-mainstram life path, in order to fit the mold imposed by their religious community. Taylor's originality principle might be appealed to here: each of these agents' "inner voice", it might be argued, must be the one that tells him to take the original path, not the one telling him to go with the flow.

It is certainly true that we often associate authenticity with originality, particularly in the case of artists, which is why this appeal to the idea of originality as deciding criterion appears quite plausible in the case of the composers. However, it seems less persuasive when applied to cases involving other kinds of dilemmas than those of a creative artist. Consider the following:

Juan's case. Juan is a farmer in Colombia. He is descended from a long line of farmers, yet his parents were keen to ensure that he would have reasonably broad career prospects when reaching adulthood. Juan thus spent more years in education than they did, and his good marks would have allowed him to move to the city and get training for a variety of different jobs there. However, after careful reflection, he finally chose to stick to the family tradition and persuaded his parents to let him come and help them on the farm instead. He has now taken over from them, and has never regretted his choice: on the contrary, being a farmer feels like a vocation to him, and he senses that he would not have enjoyed life in the city

nearly as much. His existence involves a repetitive daily routine that does not leave much space for creativity, but Juan is fully happy with his life and has no desire to change it.

Suppose that Juan is in fact right to suppose that life in the city would not have suited him as much as the simple, traditional life of a farmer. Under that assumption, it seems plausible to think that he is indeed following his "inner voice" and living an authentic life, even if we assume he is not fully realizing all the originality he is capable of. *Pace* Taylor, a person's inner voice need not always say something original. For some people, a rather traditional and ordinary existence might be what makes for an authentic life, the one in which they would find the most meaning and fulfillment. If so, originality will not always provide an adequate criterion for identifying a person's inner voice in cases where competing candidates are found.

One could perhaps counter here that while the sort of life chosen by Juan might not be particularly novel, it is still original to the extent that, we may reasonably assume, it is very different from the lifestyle of most people from his generation. Alternatively, it could be argued that Juan's inner voice is original insofar as it entails diverging from his parents' initial expectations, or from the choice most people would have made in a situation like his – isn't it likely that most would have preferred the city life? But none of these suggestions appears convincing. Imagine that in the world that Juan inhabits, many young Colombians who are given the same opportunity as him also choose to decline it and prefer to become farmers, because they find such an occupation more fulfilling than any other. Juan's preference, and the life to which it leads, are then no longer original even by the liberal standards just suggested for originality, yet is it enough to deny that they can reflect his "inner

voice" or calling? Surely not. It does not seem that the question whether following that preference is authentic or not for Juan hinges on the particular number of people from his generation who happen to have embraced a similar lifestyle, or who would have done so had it been an option for them. And while Juan certainly manifests a laudable sense of independence in sticking to what he sees as his vocation, despite his parents' expectations (which I am assuming are not oppressive), independence is not the same thing as originality. A conservative person, keen to defend existing traditions and institutions, can show a strong independence of mind by opposing the reforms and innovations supported by most of her contemporaries.

It might also be objected that I have misinterpreted Taylor's principle of originality. True, "original" can be synonymous with "inventive" or "novel", but it can also simply mean "not copied". Yet to this it must be replied that Taylor's principle of originality is not an injunction against modelling our way of life after that of others (though such an injunction is indeed its counterpart). Rather, it is supposed to *describe* what our own "inner voice" has to say. Also, Taylor's phrasing makes it clear that the principle of originality comes *on top* of the demand to be in touch with this inner voice rather than conforming to external standards. It is not identical with such a demand. Taylor thus writes that the moral ideal of authenticity "greatly increases the importance of this self-contact by introducing the principle of originality: each of our voices has something of its own to say" (Taylor, 1991, p.29).

Could appealing to the idea of a person's "nature", rather than to that of an inner voice, prove more helpful to resolve cases of apparent conflict within the true self? Many people may be drawn to the view that Penitent and Ex-gay's nature

involves sexual attraction to men, whereas their religious beliefs condemning such attraction are entirely the product of social influences, working against their nature (and unlike the sadist, their nature can express itself through perfectly harmless actions). Some might disagree here, presumably on religious grounds, with the idea that homosexual behaviour need not harm anyone. If, for instance, such behavior were contrary to a divinely ordained "natural law" (as taught by the Roman Catholic Church), and thereby compromised the chances of salvation of those who engaged in it, it would indeed count as necessarily harmful. As I believe that views of that kind are based on indefensible metaphysical and moral presuppositions, I will leave them aside. Even if we grant the presumption of non-harmfulness, however, we still need an argument for the view that homosexuality is part of Penitent and Ex-gay's nature, whereas their religious beliefs are not. Clearly, appealing to the familiar concept of "human nature" will not help: insofar as the concept refers to the essence of what it is to be a human being, neither being gay nor being straight can be part of human nature, since neither is a precondition of being human. It is *their own individual nature* that Penitent and Ex-gay are allegedly violating according to the line we are considering, not human nature in general. But how do we determine which features belong to a person's individual nature, and which do not?

It would seem that a common way of understanding this idea of an individual nature involves teleological presuppositions. Many people will thus say that they were "born" for some purpose – to possess certain features, exercise a certain profession, or accomplish certain things. Tennis champion Pete Sampras has thus stated multiple times that he felt he was "born to win Wimbledon". Such teleological thinking often goes along with religious beliefs, the assumption being that God has got a specific

plan for each of us. Living in accordance with our nature then means following that divine plan, and expressing and cultivating the various traits and talents He intended us to have. Yet the idea that it is our "destiny", or that we are "meant" to become a certain kind of person, need not be wedded to religion. Elliott actually stresses that the contemporary ethic of authenticity precisely leaves aside this assumption of a divine plan determining what counts as a good life for us (Elliott, 1998, p.181). What I believe Elliott does not emphasize enough, however, is the fact that a number of those who are committed to such an ethic seem drawn to some kind of natural teleology, according to which nature, if not God, somehow "designed" us to have certain characteristics. However, once we have put religious creeds to the side, it does not seem we can identify, for each particular individual, some sort of *telos* or end for which nature designed him/her. Neither ancient Greek teleology nor evolutionary biology, the two main sources to which we might turn for help, seem able to provide a solution here. To the extent that they can be said to attribute a "function" to us (e.g. the promotion of our evolutionary fitness), this function is not an individual one, but one common to all human beings. And if it were suggested that a person's individual nature actually consists in the set of traits that would best allow her to fulfil the said function given her particular circumstances, the proposal would still fail to yield the specific sort of features normally associated with the idea of an individual nature. For instance, on such a proposal, merely changing an individual's environment could entail changing her nature, provided that a different set of traits would best promote her evolutionary fitness in the new environment. But it is not usually assumed that we can change our own individual nature *through the sheer act* of changing our external circumstances (though some will accept that this nature can eventually change *as a result* of exposure to the new circumstances).

That is not to say that all talk of "destiny", or of being "born" to become a certain kind of person, should be dismissed as misleading. Such ideas can certainly be useful, as they likely have been for someone like Pete Sampras. What matters is to understand them figuratively: that is to say, not to base them on questionable teleological assumptions, but rather to interpret them as pointing to the desirability of playing to one's strengths, and following (when possible) one's passions. There is no need to assume that nature somehow "destined" us to have certain talents or preferences to acknowledge, first, that we do have such features, and secondly, that being aware of them and shaping our life in accordance with such knowledge is likely to help us live a good, flourishing life.

Some might want to identify the person's nature with his genome, assuming e.g. that it contains "instructions" to the effect that Ex-gay is meant to be a homosexual and that his course of therapy somehow prevents those instructions from being properly carried out in his phenotype – which would mean that he is still going against his own nature even after he has successfully changed his sexual desires (since he didn't tinker with his own genome). But as we have already suggested, such a view is naive and misguided: genes are not the sole determinant of features like the one Ex-gay is changing (and of most other human features for that matter), even in therapy-free circumstances. Taken on their own, independently of their interaction with particular environments, they do not tell us that we "ought" to have any specific set of physical or psychological features. It is true that the variation between different individuals is explained to a greater degree by genetic differences with regard to some traits (say, eye colour) than others (e.g. which particular language one happens to

speak). If it were suggested, however, that a feature will be part of our individual nature if variations between individuals with respect to it are mostly due to genetic factors, it is unclear that homosexuality would meet that requirement: though there isn't yet any real consensus as to the respective contributions of "nature" vs. "nurture" to the development of that trait, a recent study of twins in Sweden found that about 35 % of differences in same-sex behaviour were accounted for by genetics, which means a smaller proportion than environmental factors (Langstrom et al., 2010). Of course, we could also lower the bar for including a feature into our nature, or we could suggest that when two features lead to a practical conflict, authenticity requires us to give priority to the one that, to put it simply, owes more to "nature" than the other (i.e. the one for which variance among individuals is explained by genetic factors to a greater extent than the other trait). Yet as we do not currently have a precise idea of the respective contributions of genes vs. environment when it comes to differences in traits like sexual orientation and religious beliefs, we cannot rule out the possibility that both traits might meet the more modest threshold we might set for inclusion into a person's nature. Also, this means that the claim that a homosexual orientation owes more to nature than a person's particular religious beliefs is, at this point in time, little more than a conjecture. And finally, even if the correctness of that claim could be empirically established, it would remain unclear why it should carry any normative weight. Suppose I have a natural (but harmless) disposition to laziness, which I nevertheless manage to override out of a desire to be hard-working and successful. Suppose, too, that research has shown differences among individuals in the former trait to owe more to genes than differences in the latter. Would this really entail that I have a reason, grounded in authenticity, to abandon myself to laziness, and to forget about self-discipline? And that the reverse would have been true had the findings

turned out the other way round? This sounds absurd.

The most plausible way of spelling out this idea of an individual nature seems, on reflection, to simply identify it with the concept of the true self. If so, for all we know, both Ex-gay's homosexual orientation and his religious beliefs could in principle be part of his nature. And once we have abandoned the project of identifying the features belonging to a person's nature by appeal to her genotype, there seems to be no good reason to deny that Ex-gay is really *changing* one aspect of his nature through his course of therapy, rather than just repressing it. Could one then argue that such a change is still problematic from the perspective of authenticity? This leads us to the second possible construal of the inauthenticity charge we have mentioned in section 2.6, the charge of modifying core traits.

2.7.2 Modifying "core" traits

The idea that self-creation projects like that of Ex-gay involve the modification of a "core" part of the agent's identity (in the narrative sense) that should not be tinkered with is discussed in some detail by DeGrazia, who ends up rejecting it. He begins by asking "[w]hat is wrong with changing someone's narrative identity, or self-conception, assuming she autonomously consents to the change". For reasons I shall expound in a later section, I believe that the phrasing of DeGrazia's question is inadequate, because what worries the proponents of the authenticity objection need not be the fact that Ex-gay is changing his *self-conception* – defined by DeGrazia as a person's "most central values, implicit autobiography, and identifications with particular people, activities, and roles" (DeGrazia, 2005a, p.266).

Their primary concern is that Ex-gay is changing a significant aspect of his *identity*,[32] and contrary to what DeGrazia suggests, narrative identity cannot simply be equated with a person's self-conception in his sense, even though it is plausible to think that it will typically *include* it. Insofar as the notion of narrative identity is meant to provide an answer to what Schechtman calls the characterization question (which characteristics fundamentally define a certain person?), it can plausibly be said to encompass things beyond those listed by DeGrazia. For instance, one might argue that Ex-gay's homosexuality, at least before the therapy, is part of his narrative identity, even if, far from identifying with it, he always disowns it as an illness or sin from which he wishes to purge himself, and adamantly denies that it plays any role in defining who he really is.

We need not dwell on this issue, however. In the rest of his critique, DeGrazia does consider the possibility of understanding narrative identity in terms of the idea of a true self encompassing things beyond one's self-conception. Even then, he thinks we cannot justify the charge of inauthenticity in cases where someone modifies some aspect of her identity in an honest and autonomous manner. He rightly rejects the claim that *drastic* self-transformation is inherently inauthentic (DeGrazia, 2005b, p.113). If a convicted criminal had a moral epiphany while serving his sentence, so that on his release from prison he decided to completely change his lifestyle and social circle, eventually working full-time for a charity group, there would be no

[32] It might be more accurate to say that what worries them is that Ex-gay is *choosing* to change a key part of his identity. Most people accept that it is legitimate to change some aspects of our identity, such as certain religious or political convictions we used to have, provided that we sincerely believe they have been discredited. However, it matters that we do not *deliberately decide* to change our minds on such issues. Rather, such a change happens spontaneously once we have weighed the relevant considerations. What we *can* choose is to honestly acknowledge that we have changed our mind, and confront the implications of this, instead of deceiving ourselves. By contrast, trying to deliberately change our convictions while sincerely thinking they were correct would usually be seen as inauthentic.

reason to regard such a radical personal transformation as inauthentic. Another view DeGrazia discusses, which I shall call the "core traits" version of the inauthenticity charge (CTA), can be stated as follows:

Core traits analysis (CTA): Ex-gay's self-creation project is inauthentic insofar as he fails to remain true to himself by deliberately changing a "core" feature of his (narrative) identity, and it is *always* morally wrong to do so.

However, such a radical view is clearly too extreme to be plausible. As DeGrazia has no trouble showing, it certainly seems permissible to deliberately modify some of our core traits in certain circumstances, including personality traits, as in the case of shyness, or of someone like the sadist considered above (DeGrazia, 2005b, pp.235-41). DeGrazia, however, doesn't consider the case of sexual orientation, and it might be suggested that such a feature constitutes a better candidate for an "inviolable" characteristic. But the claim of inviolability is implausible even in relation to such a feature. It seems permissible to deliberately change even one's sexual orientation in certain circumstances – suppose, for instance, that Ex-gay could expect to lose his life at the hands of homophobic thugs if he chose to remain gay.

But rejecting CTA does not yet commit us to concluding that the charge of inauthenticity is unpersuasive in cases like that of Ex-gay. Indeed, more nuanced versions of CTA, not considered by DeGrazia, could also be suggested. One possibility would be to abandon the idea that core traits are properly regarded as inviolable, and to maintain, more modestly, that the central role they play for defining who we are still gives us a *reason* not to modify them, only one that can sometimes be outweighed:

CTA+: Ex-gay's self-creation project is inauthentic insofar as he is deliberately changing a core feature of his identity. We always have a strong authenticity-based reason not to modify a core trait of ours, but this reason can sometimes be outweighed by others, grounded e.g. in the agent's own interests.

CTA+, however, does not necessarily imply that Ex-gay's self-creation project should be morally *condemned* for being inauthentic. To get to that conclusion, one would have to add that his authenticity-based reason to remain true to himself is not outweighed by competing ones, and also that the costs he would suffer were he to stick to his homosexual orientation do not warrant excusing him for acting inauthentically. That CTA+ should not necessarily entail blaming Ex-gay seems to me to count in its favour, given that the social and psychological costs that authenticity might impose on him certainly seem relevant to an ethical assessment of his project – even if we think that authenticity is worth such costs.

That said, CTA+ still appears flawed, due to its assumption that if some trait is a central constituent of our identity, this in itself gives us a reason not to change it. The sadist's case again provides a counterexample here. His sadistic propensities do seem to count as a "core" trait of his. If so, according to CTA+, this person has an authenticity-based reason not to take the drug (though one that might be, and presumably is outweighed by others). Some might be willing to bite that bullet, but I find such a consequence implausible. Arguably, the fact that his cruel propensities constitute a core trait does not in itself give the sadist *any* reason to decline to take the drug. As we have said, he might have a prudential reason not to take the drug if this would promote his subjective well-being in the long run, but were the drug not to

have such a positive effect, this reason would disappear. If so, the sadist's case constitutes a counterexample to CTA+.

Can't we add a further proviso to CTA+ that allows it to avoid that problem? Following a suggestion previously made, one might argue that this analysis is meant to apply, not to *any* core feature whatsoever, but only to features that are not, by their very nature, harmful to others. This would rule out the undesirable implication that someone like the sadist has an authenticity-based reason to refrain from taking the pill. However, it would still imply for instance that an overly shy person has an authenticity-based reason not to work to overcome her shyness. I agree that such a consequence seems more palatable – some people may for instance find shyness charming, even in degrees significant enough to be handicapping. However, while granting this, most of us would still say that the authenticity-based reason to retain one's shyness will often be outweighed, for instance, by prudential considerations, given that this trait (especially, perhaps, among men) can be severely limiting when it comes to life opportunities. But if the proponent of CTA+ wants to concede this, it seems difficult to see how she could avoid the conclusion that authenticity-based reasons are equally outweighed in Ex-gay's case, whose original sexual orientation is a serious disadvantage in the community to which he belongs and might also prove to be such in other parts of society with a bias against homosexuals. And while it may well be appropriate to excuse Ex-gay for the choice he makes, in light of his particular circumstances, I nevertheless do not believe that considerations of authenticity are *outweighed* by prudential ones even in his case. Indeed, I have suggested that it would still be *better* if he had chosen to remain true to himself, and that his self-creation project is morally problematic, even if excusable.

The amended version of CTA+ just suggested also faces another problem. Indeed, it entails that we always have at least an authenticity-based reason not to improve ourselves in various ways: for instance, by working to become more courageous, disciplined, patient, or to cultivate other virtues, since none of these qualities is, by its very nature, harmful to others – quite the contrary. Such a consequence seems implausible. It may perhaps be true, as some authors have suggested, that cultivating moral virtue *beyond a certain degree* might sometimes mean forfeiting certain types of nonmoral value, in a way we might find regrettable: think of someone with a scathing wit who, in an effort to become kinder and more considerate, were to lose some of his interesting edge.[33] However, even if this claim were to prove correct in *some* cases, it surely doesn't *always* apply. Take someone who, after years of being unable to publicly stand up for his convictions due to crippling feelings of anxiety, eventually manages to conquer his fears thanks to regular practice at debating and public speaking, and repeated exposure to other intimidating situations. Why must such a form of moral self-improvement necessarily involve any loss in pre-existing nonmoral value? I do not see that it must.

A similar problem would affect a different way of restricting the core features to which CTA+ is meant to apply, which would stipulate that the relevant features, besides being non-harmful, must also be *natural* ones – not according to any of the genetic criteria previously discussed, but rather in the sense that they did not result from deliberate self-manipulation. On this view, authentic features are "natural" ones; intentionally contrived ones, by contrast, are inauthentic. (In the context of the

[33] See e.g. Susan Wolf's classic discussion of this issue in 'Moral Saints' (Wolf, 1982).

enhancement debate, the supposed "unnaturalness" of the relevant technologies is frequently cited as an objection to their use.) The assumption may be that natural features have a special kind of value that deserves protection, or that deliberately changing them reveals a morally questionable character trait,[34] or both. This suggestion, too, seems unpersuasive: many desirable qualities, including at least many virtues, are "unnatural" insofar as they need to be deliberately cultivated by those who wish to acquire them (we don't just happen to develop them without striving to). It isn't clear that we have as much as a (defeasible) reason not to develop such qualities and to stick to our original traits. And far from displaying an objectionable, hubristic drive to tinker with a valuable "given", a self-transformation project that involved, say, the cultivation of virtues like courage or honesty would seem to warrant praise.

It is true that, all else being equal, we prefer e.g. the spontaneous expression of a "natural" or "raw" personality to a contrived, artificial personality style, no matter how respectable the intentions of the self-maker might be. This criticism may sometimes apply to those who take the ideal of "turning one's life into a work of art" too seriously. Yet even in such cases, it isn't clear that it is *the trait's not having been deliberately shaped* which we value, rather than just its being *free of affectation*.[35] It seems to me more plausible to value the latter, yet it doesn't entail the former. Some features that we deliberately choose to acquire can become "second nature" to us, in which case they will not seem contrived or artificial. Many of those who saw violinist Yehudi Menuhin perform in his younger years were struck by his "natural" way of

[34] Michael Sandel makes a point of that sort in specific relation to enhancement technologies, the use of which, he writes, reveals "a Promethean aspiration to remake nature, including human nature, to serve our purposes and satisfy our desires". According to him, such a practice is incompatible with "an appreciation of the gifted character of human powers and achievements" (Sandel, 2007, pp.26-7).

[35] As I shall explain in section 3.4 when discussing some ideas by Jerry Cohen, there are also other reasons why we may value our existing features and believe they should be preserved, yet it isn't clear to what extent the "naturalness" of those features is among those reasons.

playing, yet Menuhin's skill clearly wasn't innate, nor had it been acquired by accident, or even solely because of external influences. Depending on the effectiveness of his course of therapy, there is no reason in principle why Ex-gay's behaviour after his transformation shouldn't spontaneously flow from his new sexual orientation, without the need for affectation of any sort.

Still, the feeling may persist that at least *some* features have special value *in virtue of being natural*. Perhaps the strongest example that might be adduced in its support is parental love. Imagine a parent who, after giving birth to a long-awaited child, discovers that she does not feel the affection she had expected for her baby. Despite her hope that motherly love will eventually kick in, as months go by, she still feels estranged from her child – to her great dismay. As a last resort, she decides to start taking a new pill that has been proven to foster feelings of love for one's child where these were initially lacking. The pill works, and the mother is soon submerged with feelings of parental love. Surely, she is still missing out on something extremely valuable – the ability to love her child "naturally", spontaneously and without the need for crutches of any kind?

While acknowledging the force of this example, I am nevertheless not convinced that it shows spontaneous and immediate parental love to be preferable to contrived love because of its *naturalness*. First, I suspect that the tendency to see love induced by a pill as inferior to natural love depends on an implicit assumption that the former is not the real article. True parental love, we may think, presupposes among other things being responsive to the loveable qualities of one's child. And we may doubt that a pill might be able to induce such responsiveness where it is lacking. Instead, we

may assume that all it can do is produce "warm and fuzzy" feelings independent of any such insight, which will therefore not count as genuine feelings of parental love despite their resemblance to them. I shall examine the concern about producing "fake" traits in the next section. For the moment, since our question is not whether true parental love can be artificially induced, but whether love so induced, even if fully real, can have the same value as natural love, it is important to rule out any potential assumption of "fakeness" in the examples we are considering.

Imagine then that rather than taking a pill, the mother in our example resorted to more traditional means of fostering love for her child: for instance, if the child's birth had been unplanned and, as a result, disruptive of her life plans, by reminding herself that the child was not responsible for her arrival into the world and its consequences; or by taking more time out of her daily schedule to spend with the child, to allow feelings of affection for her to grow (Liao, 2011). If such methods were successful, there would be less reason to doubt the psychological reality of the resulting feelings of love, but also, I think, less plausibility to the claim that parental love that had been learnt in this way was necessarily less valuable than immediate, natural love. Furthermore, even if one does want to stand by that claim, a more plausible justification for it seems available than an appeal to the value of naturalness: namely, one might think that loving one's child is an appropriate affective response to her existence, and therefore that learnt love, to the extent that it only comes with a *delay* compared to natural love which is immediate, temporarily involves an inappropriate emotional response, which might make it less desirable both in itself and in view of its potential negative psychological impact on the mother and the child. Yet even if we accept this, we are not yet committed to the claim that naturalness has special

intrinsic value. Imagine that the mother in our example, having herself been neglected by an unloving mother and concerned that she might end up feeling and behaving the same way towards her own children, takes several preparatory steps to increase the probability that she will be able to love them when they are born: she works in a nursery for some time to learn how to bond with young children more easily, watches movies featuring cute babies, etc. As a result, when her first child is born, she spontaneously experiences the affectionate feelings her own mother could never conjure up for her, and which she herself, we may assume, would not have experienced without such prior training. Her feelings of love are then still "unnatural" and inauthentic by the criteria given above, yet I do not see that they are, on that account, any less valuable than similar feelings that would have arisen naturally. (Indeed, in this last example, the relevant feelings do not come with a delay.)

As I shall explain in the next part of this dissertation, I believe that rather than focusing on the "natural" character of our core traits, we should instead take the view that authenticity only gives us a reason to preserve our core features when their appropriate expression has, or would have, *intrinsic value* (for reasons that may be independent of their naturalness).[36] Before we move on to such a solution, however, let us explore one more way of explaining the inauthenticity of self-creation projects like Ex-gay's without introducing any direct reference to intrinsic value into our explanation. Indeed, the use of a more neutral criterion might be considered desirable, as the question whether the expression of some particular trait has intrinsic value will typically be a controversial matter. On the other hand, even though a proponent of

[36] By "intrinsic value", I mean the value a thing has for its own sake (what Christine Korsgaard has referred to as "final value"). I do not wish to imply that it has such value solely in virtue of its intrinsic properties.

CTA+ might well be assuming that core traits (or their expression) are valuable, at least – it might be said – there is less room for controversy regarding whether some trait is a "core" one or not (or whether it is natural or not), and therefore, regarding whether the deliberate modification of that trait should be considered authentic or not on the basis of CTA+. By contrast, there will be disagreement as to whether the solution I am proposing declares any particular instance of self-transformation authentic or not. We might feel that it is an advantage if our analysis of the inauthenticity worry avoids such potential controversies (though I shall undertake to show that trying to avoid them eventually leads us into a dead end).

2.7.3 Producing traits that are not really our own

This third construal of the inauthenticity charge, like the previous one, is mostly relevant to Ex-gay's case. It seems to give voice to a worry that a number of people share about enhancement technologies, and if it could be cashed out in a plausible manner, it would not be vulnerable to the objections we have raised against CTA and CTA+. For instance, even though the sadist we described above may not have a reason to retain his barbaric cruelty simply in virtue of the fact that it counts as a "core" trait of his, it would still seem fair to see a problem with the more humane personality he would acquire thanks to pharmacology, if this personality were not really "his" but were somehow fake. Perhaps, one might say, it would nevertheless be better to give the sadist such a fake personality if it made him behave in a more civilized way, than to allow him to remain the way he was and thereby pose a threat to others. Yet, the argument would go, the fact that this new personality was fake would still count as undesirable. And to anticipate the discussion of the impact of

enhancement technologies on our authenticity, if psychological features produced through the use of pharmacological enhancers must necessarily be fake, then this will at least partly vindicate the authenticity objection to enhancement, insofar as many of the traits that people are looking to change using enhancers, such as shyness or low mood, are not harmful to others in the way that the sadist's personality is. The badness of the supposed fakeness of the new traits will then be less likely to be outweighed by the benefits of the change.

The key challenge for this proposal is to provide a justification for calling e.g. Ex-gay's new sexual orientation, or the sadist's more civilized personality on the drug, "fake". A straightforward approach would be to simply state that traits that involve stifling one's inner voice, or that have replaced core traits, are necessarily fake. However, besides the fact that such an approach would seem to reduce this third analysis of the inauthenticity charge to one of the two previous ones, it would be unpersuasive in Ex-gay's case. Indeed, as we have said, there is no plausible sense in which, once he has undergone reparative therapy, Ex-gay is still stifling his inner voice, nature, or true self. Rather, he has actually *changed* it. And to call his new sexual orientation "fake" on the grounds that it has replaced a core trait of his represents, quite simply, a misuse of that term. If you have changed quite significantly after undergoing a religious conversion, or serving in the army during the Iraq war, this surely isn't ground enough for calling your present self "fake".

As I have mentioned before, the idea that some trait is somehow fake need not be tied to a true self analysis of authenticity. What are the available alternatives? The criterion most commonly employed to declare some particular psychological feature

fake is by pointing out that it is not psychologically real.[37] In Ex-gay's case, this would seem to be a non-starter: his new sexual orientation is, I am assuming, psychologically fully real – he is neither deceiving others nor himself about it. Is this an unacceptable assumption? I have previously conceded that if his case were a real-life one, Ex-gay would indeed be unlikely to really have changed, given what we know about the effectiveness of procedures like "reparative therapy". Yet it still makes sense to assume that he has actually changed, if our ultimate aim is to discuss enhancement technologies – technologies that have the potential to genuinely re-shape who we are, in ways that haven't been available to us so far. Some authors have argued that emotions that have been deliberately managed by the person experiencing them are never authentic, the underlying assumption being that only an emotion that arises spontaneously can be the "real thing",[38] and this claim could be extended to cover other psychological features like sexual orientation. Yet even if our focus is on traditional methods of psychological self-management, such a claim doesn't always seem correct (though it certainly is in *some* cases). Suppose that, after training myself to focus my attention better, I start to find comedy shows more amusing than I did initially. Is my enhanced amusement necessarily "fake"? I fail to see why, and I see no more reason to assume that sexual preferences cannot *in principle* be psychologically real if they are the result of deliberate self-management. If procedures like reparative therapy fail to effect a real change, this can only be a contingent fact, one that might be falsified by the development of new techniques.

An alternative criterion of fakeness that might be put forward would be the

[37] This criterion is often proposed in discussions of emotions, yet it could plausibly be extended to other aspects of a person's psychology, such as sexual orientation. See e.g. Kraemer, 2011.
[38] See e.g. Mulligan, 2009.

insufficient stability of the person's new feature: any feature, it might be argued, needs to persist for a sufficient amount of time to *really* be attributable to us. Yet since we can assume Ex-gay's new heterosexual orientation to persist for the rest of his life, this criterion may not be able to declare it fake. Some might argue that such a criterion can more plausibly be used to declare the sadist's new personality to be fake, at least if we assume that he needs to keep taking the "morality pill" in order to retain that more civilized personality. In addition, another criterion of fakeness that might be offered in relation to the sadist's case would be that the pill fails to solicit his rational capacities, when these ought to be solicited (a suggestion similar to one we have already encountered when discussing parental love pills). The pill, it might be argued, acts directly on the sadist's brain to make him more empathetic and less cruel. It changes his attitudes and desires by purely causal means without providing him with genuine moral insight, when such insight ought to be at the root of the change. It does not allow him to truly realize that the way he had been treating others so far was wrong, and that he should repent of what he did and mend his ways. Because his transformation is not based on genuine insight, it is fake, and so less than ideal, even though it might still be better than the status quo. Given that the "morality pill" might be regarded as an enhancement technology, I will postpone the discussion of these two suggestions regarding the supposed fakeness of the sadist's new personality (and of pharmacologically-induced parental love) to the final part of this dissertation. For the moment, let me note that the "bypassing rational capacities" criterion of fakeness will not apply, either, to Ex-gay. Indeed, it does not seem plausible to assume that our sexual orientation is something that crucially depends on the exercise of our rational capacities. Rather, it seems to be a "brute" preference that is causally but not rationally determined, like a number of other individual differences, from

temperament to eye color.

I thus believe it is better to abandon the "fake traits" charge as a way of explaining why Ex-gay is acting inauthentically. As for the suggestion that Ex-gay's new heterosexual orientation is not *uniquely his own*, we have already noted its unhelpfulness – since his original homosexual orientation isn't either.

In the next part of this dissertation, I shall argue that there is a plausible sense in which Ex-gay's new sexual orientation can be called inauthentic, but that this sense isn't equivalent to "fake". Instead, his new orientation is inauthentic simply in the sense that it isn't his *original* one, by which I mean, the one that he just happened to develop prior to engaging in reparative therapy.[39] And the reason why inauthenticity in that sense can be ethically problematic isn't that it is always problematic to tinker with our original, "authentic" traits. Rather, the reason is that appropriately living out his original homosexual orientation would, in Ex-gay's case, be *intrinsically valuable*. I will also try and offer an analysis of the inauthenticity of my remaining three cases that improves upon the one we have considered in section 2.7.1. On that basis, I will propose a "true self" analysis of authenticity revolving around two fundamental definitions. Before I undertake this, however, I need to examine and defend the legitimacy of the concept of the true self, which lies at the centre of the view I propose to defend.

3 A NEW ACCOUNT OF AUTHENTICITY

[39] I am therefore not using "original" here the way Taylor does when putting forward his principle of originality.

3.1 The concept of the "true self"

3.1.1 Introduction: some problems with the concept

The concept of the true self, as we have seen, can be understood as referring to a set of relatively stable attributes that contribute to defining who we are, and that are relevant to our (narrative) identity regardless of whether we have made them part of our self-conception. Before I proceed to argue for the respectability of the concept, I want to acknowledge straightaway that certain assumptions about the true self commonly found among non-philosophers, and occasionally encountered in the philosophical literature too, do appear to raise serious problems. One such assumption says that the true self has the nature of an "essence", which as we have seen would prevent us from treating features like personality traits, mood propensities, or personal convictions and commitments as constituents of the true self (thereby undermining the basis of authenticity concerns about enhancement). Another problematic assumption we have already alluded to is the idea that simply looking inward and getting in touch with our true self, without taking a stand on the value of its constituent traits or of their appropriate expression, is all we need to figure out how to live authentically. This assumption must also be rejected if, as seems plausible, people's true selves can sometimes include features in conflict with one another.

I will now mention two further problematic assumptions that appear to be commonly made by those who believe in a "true" self. The second of those assumptions will lead me to consider some recent work on the concept in experimental philosophy.

3.1.2 The true self as static

Closely related to the essentialist understanding of the true self (though it does not necessarily presuppose essentialism) is the assumption that the true self is a static entity that remains the same throughout our lives: we cannot really change it even if we try to. The most we can do is hide it under a "fake" exterior – bringing us back to the idea of fake traits previously discussed – and thereby manage to deceive others, and even ourselves, about who we really are. This approach can either rely on an essentialist view of the true self, or it can accept the possibility that our true self might change through time, yet deny that this can happen through our own efforts. As DeGrazia puts it:

> One possible view envisions the self as completely "given," although to discover its shape and true colors one may have to dig (with reflection, therapy, or the like). One can find the self but not change it; any change is due to forces outside one's agency. One version of this view takes a person's "inner core," the values that define the individual, to be entirely constructed by society. (DeGrazia, 2000, p.36)

DeGrazia is right that this view is implausible. As already suggested in section 2.7.3, it certainly seems possible *for us* to change some fundamental aspects of ourselves. Take someone who manages to finally get over the shyness that had been limiting his life for years, with the help of a course of cognitive-behavioral therapy, requiring him to engage in various exercises aimed at behaviour change. There is good evidence that such programs can genuinely reduce shyness on a long-term basis (see e.g. Zimbardo, 1977), and even though the person embarking on such a program does to some extent depend on the guidance and encouragement of his therapist, he clearly needs to play an active part in the self-transformation process – he cannot expect the therapist to continually be present by his side and push him to take the initiative in social

situations. The claim that we cannot deliberately change who we really are simply does not stand up to scrutiny, and the rise of enhancement technologies further undermines its plausibility.

3.1.3 The moralization of the true self

To the extent that the concept of the true self designates a set of features summing up who a person "really" or fundamentally is, it seems we should not assume that these features must be morally good, or valuable in some other way. Yet recent work in experimental philosophy suggests that this is exactly what people tend to assume. Joshua Knobe and his colleagues thus conducted a series of studies destined to test people's intuitions about the concept of the true self. In one of these studies, they presented the participants with a series of imaginary vignettes depicting agents engaging in behaviours that were sometimes morally good (e.g. a honest CEO), sometimes bad (e.g. a dishonest CEO), and sometimes neutral (e.g. choosing red wine over white wine). They were then asked to what extent they agreed with the claim that "there was something deep within the person calling him or her to behave differently" (Newman et al., Under review, p.11). It turned out that the participants were significantly more likely to assume that the agent's true self called him to behave differently when his behaviour was morally bad than when it was good or neutral, even in the absence of any evidence suggesting the presence of a good true self lying behind the surface. This appears to support the idea that people tend to posit the existence in everyone of a morally good true self, quite independently of the available evidence (ibid., p.41). In another study, Knobe and his colleagues presented a group of subjects, some of whom identified themselves as conservatives and others

as liberals, with a series of imaginary vignettes. Some of these vignettes depicted people who changed some key part of their lives in a sense that conservatively-minded people would approve of: e.g. a homosexual who stopped having sex with men and entered into a heterosexual marriage. Other vignettes depicted a change that liberals would favour, e.g. a sexist person who came to embrace the ideal of gender equality. The participants were then asked how much they agreed that the change in question constituted a manifestation of the person's true self, which so far had remain hidden from view. The results, Knobe reports, were that "conservative participants were more likely to agree that the behavioral change resulted from the emergence of the person's true self for the conservative items...than for the liberal items... Conversely, liberal participants were more likely to agree that the behavioral change resulted from the emergence of the person's true self for the liberal items...than for the conservative items" (Newman et al., Under review, p.21). Though Knobe concedes that further research is needed on this issue, he takes these results to provide some support for the hypothesis that people tend to regard the traits they value in someone as part of that person's true self.

If that hypothesis is correct, it means that the way a person's true self is commonly individuated is problematic, at least if our goal is to accurately identify who that person "really" is. True, "moralizing" the true self in this way has the advantage of offering in principle a way to decide, in cases of inner conflict such as those I have described, which of the conflicting traits reflect the person's true self – namely the more valuable ones. I don't deny that this solution seems to have some appeal in some cases. For instance, the assumption that the composers' bolder ideas reflect their "inner voice" is especially plausible if we assume that they are valuable.

If we assume that their development of S.'s style is in fact aesthetically far superior, the latter becomes a more plausible candidate for the status of "inner voice" – maybe their artistic vocation is to be continuators rather than innovators. Yet the advantage of such a solution comes at a significant cost. Besides the fact that people often attribute certain traits to others in the absence of any corroborating evidence, there is, as we have said, no reason to assume that who a person "really" is must only encompass valuable traits. Being e.g. manipulative or utterly lacking in empathy might well be a fundamental attribute of certain individuals, like the sadist in our previous example. It might perhaps be retorted here that such traits only develop as a result of pernicious forms of social conditioning, and that they can never arise in people who grow up in "healthy" environments, because, as Rousseau thought, human nature is fundamentally good and all evil implies distortions from that nature. Such a Romantic view, however, sounds rather naïve today, and the idea that traits like psychopathy are to be explained solely in terms of environmental influences does not fit with the current state of our knowledge (see e.g. Skeem et al., 2011). Furthermore, it is not clear that such a reasoning is at work when people attribute traits they value to a person's true self. The (presumably unconscious) criterion of attribution seems to be the perceived value of the trait, rather than its supposed causal origin. Knobe and colleagues actually suggest that this tendency to "moralize" the true self might be related to the assumption of a static self: if the features constituting the true self are taken to be immutable, people may find it difficult to accept that some immoral dispositions are not amenable to change (Newman et al., Under review, p.37).

A possible objection to my critique would be that people simply tend to use the notion of the true self in a different way from mine. I am assuming that the notion should be interpreted as a purely *descriptive* one, but Knobe's research suggests that people do not agree with that assumption: they use it as a *normative* concept, involving certain value judgments about people's features. On what grounds can I claim that they are wrong to do this? Again, I am willing to concede that the issue is ultimately a semantic one, and that nothing strictly forbids us to treat the true self as a normative concept if we so wish. Yet as I have already indicated, the very phrase "true self" suggests that we are referring to whom the person *really* is, to the features that play a fundamental role in defining her. If so, it seems misleading to include into the concept the assumption that those features must be valuable ones. Doing so prevents us from saying, for instance, that being a sadistic sociopath largely defined who Ted Bundy "really" was. Rather, we will have to say that while this was indeed a key aspect of his personality, it nevertheless wasn't part of his true self, because people's true self can only include valuable qualities. Such a way of speaking strikes me as rather confusing and unhelpful. Similarly, I don't find it plausible to understand such a general notion as that of an "inner voice" in a fundamentally normative manner, even though we may sometimes be tempted to do so. Surely it is appropriate to say that when paranoid schizophrenic killer Herbert Mullin murdered 13 people in California in the 1970's, he was listening to an "inner voice" (supposedly telling him that this was the only way to prevent an earthquake)?

That said, I should stress, first, that "moralizing" the true self in this way might still yield the same ethical verdicts as those I have initially offered. If we assume for instance that being gay is a good trait to have, more valuable than the religious belief

that homosexuality is immoral, then on this line of thought this belief isn't part of Ex-gay's true self, in which case we should conclude, as I would, that the authentic thing to do for him is to accept himself as he is and reject the belief. I would simply reach this conclusion in a different way. On my assessment of the case, the said belief is indeed part of Ex-gay's true self, yet it is less valuable to be faithful to it than to his homosexual orientation, which is why the latter is the one that counts from the perspective of authenticity. Secondly, even though this tendency to moralize the true self might be undesirable from the point of view of sheer epistemic accuracy (determining the exact nature of a person's true self), it may still carry some *practical* benefits that might count against eradicating it, and might give us a reason to think about the true self differently in other contexts than that of detached philosophical reflection. These benefits may include a positive impact on people's subjective well-being and a disposition to facilitate positive self-change. I will come back to this shortly.

3.1.4 Rescuing the concept: the true self and narrative identity

Given the problematic nature of the common assumptions about the true self I have just described, should we simply dispense with the concept, as the critics of the self-discovery approach to authenticity would suggest? I believe not. There does seem to be something to the view that people have a set of fundamental, stable traits that explain much of their behaviour, and contribute to defining their identity, yet that these traits need not be ones that their possessor herself endorses. I would suggest that we can secure some respectability for the notion if we understand it on the basis of a prior account of narrative identity – the true self being on my analysis a subset of

narrative identity, the part of it that is most fundamental to whom the person is. On the view that I shall propose, which we might call the "objective account of narrative identity", a feature will count as part of our identity if it meets the following four conditions:

1) First, the relevant feature must be *real* or *genuine*. It must be properly attributable to the person, not merely counterfeited, or part of the deluded view he may hold of himself. If someone were e.g. very good at cheating people by pretending to be an honest and well-meaning person, this would not mean that honesty and benevolence were part of his identity – though the disposition to fakery might indeed be.

   Clearly, this first condition entails *realism* about the features constituting our narrative identity. If such features were mere fictions constructed by us,[40] possibly useful for some purposes yet not referring to any existing dispositions or other properties of ours, then my genuineness requirement would be absurd. Now how do determine whether a particular feature is real, or properly attributable to a person? As far as other people are concerned, the answer seems to be: by observing the person's behaviour, and by relying both on her self-reports, and on the observations made by others who know her. Of course, even once we have that information, we will still need to *interpret* it in order to attribute some feature to the person; and several different interpretations of that information will usually be possible, even though some will be more plausible than others. The person herself, by contrast, has a reduced ability to

[40] In the next section, I shall criticize such a suggestion as applied specifically to the concept of the true self.

directly observe her own behaviour, yet still has access to feedback from others (and from technological devices, in the form of videotapes, etc.), as well as to an additional source of information unavailable to others, namely introspective access to her own beliefs, desires, and feelings.

When the person and others around her agree in their attribution of some feature to her, this will usually constitute good evidence that the feature in question is indeed attributable to her. But what if they disagree? The answer will depend on the case. Sometimes, a person will be right against others, because she has introspective access to some aspects of her identity that are concealed from others: consider for instance an undercover agent who manages to persuade the members of the milieu she infiltrates that she holds certain religious or political beliefs which in fact she does not hold. In other cases, others will be right, rather than the person. Indeed, a person can be self-deceived, or simply mistaken about some aspect of herself, and in such cases the truth might be more readily accessible to others around her (though it need not always be). Even the special introspective access we enjoy with respect to some of our own mental states isn't infallible. For instance, the influence of social expectations can sometimes lead us to sincerely believe that we have certain feelings we don't in fact have; say, that we love the mean-spirited family pony, because that is the way we are supposed to feel (Wilson, 2002, pp.118-9). And in yet other cases, we simply won't be able to tell who is right. The person and those around her might for example offer different, yet equally plausible interpretations of her behaviour in some circumstances, or they may all feel uncertain about what the correct interpretation is. There may be cases

where we simply are not in an epistemic position to establish some particular truth about who a person is – even when that person is ourselves.

2) The second condition is *stability*, implying that the features constituting our identity must endure for some minimum period of time – though not necessarily throughout the life course, as we should acknowledge that most people's narrative identity changes through time. This minimum should not be understood as a sharp cut-off point, yet it seems that in order to become identity-constituting, a feature will typically need to be properly attributable to the person for something like a few years at least. To take an example, if someone were to convert to Roman Catholicism and to live according to its teachings for a mere few weeks before forfeiting the faith and returning to her previous lifestyle, we would generally not regard her as having shown a commitment to the Catholic faith sustained enough to make it part of her identity. However, were she to abandon the faith, say, a decade or two later, after much inner struggle and many years spent living as a devout Catholic, it would no longer be appropriate to deny that being a Catholic really had been part of that person's identity throughout those years.

3) The third condition is *distinctiveness*: that is, the relevant feature needs to help *individuate* the person to whom it belongs. This does not imply that it must necessarily be *unique* to her, but it does mean that it normally won't be a feature she shares with everyone else. Being a Canadian, or a conservative, or an extrovert are thus plausible candidates for being part of a person's identity, because they help pick her out from others, even though they are not unique to

her (since there are many Canadians, etc. in the world). By contrast, even though having lungs, or a skeleton, are features that also meet conditions 1) and 2), they would not usually be mentioned when characterizing who a person is. The reason for this seems to be their lack of distinctiveness: they are common to all humans.

4) The fourth condition is that the relevant feature needs to have a significant impact on our life course, i.e. on at least some of the following: the way others view and treat us, our life opportunities (e.g. in terms of educational or economic attainment), our motivational set and mental life and, as a consequence, our behaviour, including the way we treat others and the choices we make about which environments to seek out, etc. By impact on our life course, I wish to refer to a causal contribution to the particular shape our life happens to take in those various dimensions: i.e. our life must take such a shape, to a significant extent, because of our possession of the relevant feature. Personality traits typically meet that particular condition. Other features that will often do so include our self-conception and inner story, ethical and religious convictions, cognitive capacities, and individual preferences (including our sexual orientation). I believe that the extent of its impact on our life course is an important determinant of how *central* we take a feature to be in defining a person – and we normally recognize as part of a person's identity only those features that play such a central role. This helps explain why, for instance, even if the property of having specific number *n* of hairs on my head were to meet conditions 1) to 3), it will still not be counted as part of my identity. Indeed, given that neither I nor anyone else is presumably aware,

whether consciously or unconsciously, of the fact that I have exactly *n* hairs on my head, my having this particular property cannot affect my mental life and behaviour, or the way others treat me, in any way.[41]

When exactly will a feature count as having a "significant" impact on our life course? There does not seem to be any way of precisely measuring such a thing, and I acknowledge that a subjective element is probably ineliminable in such judgments. Yet this doesn't mean that we cannot still reach a substantial degree of consensus about such matters. For instance, if someone's religious beliefs were to lead her to become a nun and to spend the whole of her adult years in prayer and contemplation in a convent, it seems uncontroversial that this feature of hers would count as having a significant impact on her life. By contrast, the opposite verdict would clearly be warranted if the only difference these beliefs made to her life concerned which box she ticked when filling in a census form, and pretty much everything else would have been the same had she been an atheist. There will, of course, be more ambiguous cases, and when dealing with these I believe we ought to make room for the possibility of disagreement between reasonable people as to the correct verdict to reach.

That said, a proviso still needs to be added to this fourth condition. Indeed, it might legitimately be objected that either *undiscovered* traits, such as untapped talents or capacities to enjoy certain things (e.g. opera), or *repressed* ones (e.g. certain sexual preferences), can also be part of someone's identity

[41] Of course, it would be unlikely to affect these things even if my having that property *were* to become known to me and to others, given that none of us presumably attaches any importance to the specific number of hairs on anyone's head.

(and of her true self), even though they may not have any impact on that person's life course as long as they remain in that latent state. This argument might apply more to undiscovered than to repressed traits, since the latter might still be able to influence the person's life course, e.g. by resulting in neuroses. Nevertheless, the general argument seems correct. To deal with it, we should qualify condition 4) by stating that in order for a feature to be identity-constituting, it ought to be the case *either* that it actually has a significant impact on the person's life course, *or* that it *would* have such an impact *were it to be expressed, in the world as it is*.[42]

Now, a trait that meets these first four conditions will count as part of our *true self*, and not just of our identity, if it also satisfies the following three, additional ones:

5) To count as part of our true self, a trait must be more than a mere *appearance*. Appearances can meet condition 1) – *appearing* to have a certain quality can be a genuine property of yours – and they can be part of a person's identity: think of someone who, simply because of the nature of his facial features, looks like an aggressive person, even though he is in fact very kind and mild-mannered. Looking aggressive would on my view be part of his identity, if this quality met conditions 1) to 4). But we would not usually say that it is part of his true self. Rather, we would contrast his gentle true self with his deceptive, rough exterior.

[42] I emphasize that it is the *actual* world that is relevant here, because for each of us there is a possible world in which discovering e.g. that we have *n* hairs on our head does make a significant difference to our life course. (People in those possible worlds deeply care about such numbers.) If all such worlds were to be treated as relevant, condition 4) would become useless, as no feature of ours could then fail to meet it.

6) The traits constituting the true self must be grounded solely in *intrinsic* properties of the person, and not e.g. in certain relations she bears to others.[43] As we have seen before, features like social roles, while they may often contribute to defining a person's identity, are normally not considered part of someone's true self. Yet such features can in principle meet conditions 1) to 4) stated so far. They do not, however, meet condition 6). The property possessed by Louis XVI of being the King of France in the late 18th-Century depended on a general social consensus stating that he owned such a title and should be treated accordingly. Once that consensus had been disrupted when the constitutional monarchy was abolished in 1792, Louis XVI lost that property. His intrinsic properties, e.g. his view of himself, were irrelevant in this regard. By contrast, being e.g. prone to experiencing negative affect is a feature one could in principle possess simply in virtue of the particular state of one's brain. It does not fundamentally require that we stand in certain relations to others.[44] This dependence on intrinsic properties of the person is, I believe, what disposes us to see the constituents of someone's true self as even more central to whom that person is than the features that are merely part of her identity, beyond the implications of condition 4) in this regard.

Let me add that in this dissertation, I shall focus more specifically on *psychological* features as constituents of the true self. The main reason for this

[43] Though social relationships can certainly influence the development of various aspects of our true self. Condition 6) only denies that the traits in question ever *directly* depend for their existence on the presence of such relationships, but not the possibility of relations of *indirect* dependence – it allows e.g. that having found oneself in such a relationship in the past may sometimes be a necessary causal antecedent of the acquisition of the relevant trait.

[44] Of course, if we were never exposed to stimuli triggering negative affect, our affective propensity would never be actualized. Yet this propensity could still exist even in such a situation, provided that it were true that *if* we were subjected to relevant stimuli, we *would* experience negative affect.

is that the traits the technological modification of which I shall discuss in part IV, focused on "cosmetic neurology", all fall into that category. Furthermore, confining the true self to psychological traits helps reach conclusions we may find appealing, for instance, in the case of transsexual people. Indeed, if we assume that the "real" person is always defined by her inner psychological features, we should conclude in the case of transsexualism that the person's true self is reflected in her gender identity and not in her external biological characteristics – something the liberal-minded among us will generally accept. Also, the first sense of "being faithful to one's true self" I have distinguished above concerns the expression or repression of its constituent traits, yet it sounds somewhat strange to talk about expressing or repressing *physical* features. We would rather speak of *showing*, hiding, or misrepresenting them. Excluding those features from the true self avoids this awkward consequence.

That said, other cases seem to support the view that a person's physical traits can be part of her true self, too. Many would for instance argue that, say, an African-American woman who underwent cosmetic surgery, and bleached her skin, in order to look more Caucasian-like, would be betraying her true self (besides exposing her body to quite a significant risk of harm). Unfortunately, I do not have the space to address this issue thoroughly enough here, and will therefore remain agnostic about the legitimacy of extending the concept of the true self to include physical features. This shouldn't matter too much for our purposes, given that it is chiefly psychological traits we will be discussing in what follows. And even if we were to conclude that physical traits should in fact never be counted as part of the true self, it might still be possible to

express the worry about authenticity often raised about the technological alteration of such traits by appealing to the concept of *identity* instead, while otherwise retaining the features of the analysis of inauthenticity I shall propose in future sections.

7) Finally, it seems that we ought to impose additional conditions of "depth", or resilience, on the features that can form part of our true self. I would suggest that for any such feature, it ought not to be the case that the person would forfeit it were she able to critically assess its origins, in the robust sense of critical reflection I have described in section 2.5. In this, the true self seems to differ from the more general notion of narrative identity: arguably, it can be part of the identity of someone like Nora, in Ibsen's play *A Doll's House*, to want to live the life of a dutiful and docile wife, as condoned by her husband, father and the society of her time, yet it is not part of her true self, as she realizes at the end of the play. Accordingly, she leaves her husband, and the stifling environment he presides over, in order to discover who she really is, and figure out what her true desires and beliefs are.

It might be asked here whether this last condition hasn't in fact got to do with *autonomy* rather than with the "depth" of those features. Though the two issues do seem connected (to the extent that features acquired in an autonomy-undermining environment will often not be deep enough), I still believe they should be distinguished, which is why I am saying that the person ought not to *forfeit* the relevant feature on critical reflection, and not that she ought to still *identify* with it. Consider for instance someone who, after much reflection,

> decides that his belief in God has resulted from a constricting education and stops identifying with it. He finds, however, that he still cannot shake off that belief: he simply cannot help believing in God. Should we deny that this person's belief is part of his true self? It is not clear to me that we should. In keeping with the spirit of the self-discovery model of authenticity, I want to at least leave open the possibility that recalcitrant beliefs or preferences, even if formed under conditions incompatible with autonomy, might still form part of someone's true self, because of how deeply rooted and persistent they are.

On may note that I haven't included originality as an identifying criterion for the true self. The reason for this is that such a criterion only seems plausible in individual cases, and not in any systematic way. As we have seen, it appears convincing in the case where an artist has bold, valuable new ideas. We will then say that they represent the "real him" better than any more traditional ideas he may have, on the grounds that they owe more to his own creativity than the latter. But if we assume that his bold ideas are worthless and his more traditional ones much better, many will then find it more plausible to view the latter as an expression of his true self. Also, the case of Juan provides further evidence that the originality criterion doesn't apply in all cases. I believe it is preferable to try and capture the importance of originality for authenticity by means of the general claim, which I shall defend later, that the *value* of expressing the features constituting our true self – a value which often, though not always, depends on these features themselves having intrinsic value – is key to the question of authentic action.

Also, given what I have said before about clinical depression being usually viewed as overshadowing the true self, it might be thought that we should include something like a condition about "normal functioning" here. The reason why I haven't done so is that such a notion is very tricky to spell out in a plausible manner, and that our thinking on this issue appears far from systematic. Some disorders, e.g. ADHD, seem better candidates than others for inclusion in the true self. People's intuitions will likely diverge here. Also, we sometimes want to say that the oppressive circumstances in which a person finds herself are thwarting the expression of her true self. But we also accept that some people cope with such circumstances better than others. Isn't this coping capacity part of their true self? Yet it is precisely expressed in such difficult circumstances. I prefer to avoid these difficulties by not adding any further conditions to my account, even if as a result it may appear over-inclusive to some.

3.1.5 Possible objections to my analysis

Let me now consider some issues that might be raised about my seven conditions. With regard to the second one about stability, it might be asked what my analysis implies for people whose identity involves aspects that shift on a regular basis. Think of people with bipolar disorder, who regularly move from periods of depression to periods of elated mood. Such people will change a lot from one period to the other, which can sometimes succeed each other at a fast pace. Does my view imply that these people's condition is irrelevant to their identity, and their true self, since neither their depressive nor manic episodes seem stable enough? This implication sounds counterintuitive. However, my account does not commit us to it.

Instead, we could for instance say that even in such cases, there is still a relevant feature that remains stable through time – namely, the person's disposition to experience serious mood swings, which the concept of bipolar disorder refers to. Accordingly, this feature can be part of the person's identity and true self, provided that it meets my seven conditions. A more intriguing question is, in cases where each phase in the disorder (manic vs. depressive) is actually long enough to meet my stability condition,[45] whether we should distinguish *several* true selves succeeding each other within the person (without necessarily causing a break in her numerical identity), given that each phase involves a discrete set of affective dispositions, attitudes and beliefs. Even if there are cases that warrant positing multiple true selves, however, they do not seem to be the norm. The question of what one ought to do to live authentically in cases of that kind has, I think, a solution similar to the same question asked about cases of conflict within a single true self. I shall offer such a solution in section 3.5.

Regarding my sixth assumption, it is fair to ask whether it is universally shared across cultures, or whether it might rather reflect a particularly Western way of thinking. For instance, anthropologist Clifford Geertz, in a study of Balinese culture, argues that the Balinese do not draw the sort of distinction I have sketched between a person's true self and her social role(s). On the contrary, he writes, they would say that "their role is of the essence of their true selves" (Geertz, 1973, p.386, quoted in Rorty and Wong, 1990, p.28). Because it seems to me that assumption 6) is common to many, at least in the West, who accept the concept of the true self, I will retain it in

[45] It might also be suggested that if we put together all the episodes associated with a particular phase, and ignore the interruptions from the other phase, the two phases are actually stable enough to meet my second condition.

what follows, yet it is important to note that some degree of cultural disagreement might exist on that matter. It also raises the interesting possibility, in relation to the authenticity objection to enhancement, that certain technological interventions will be regarded as posing a threat to our true self on a Western understanding of the notion, but not from the perspective of some non-Western cultures. Perhaps some of them would view e.g. mood enhancers as authenticity-promoting, if they could help someone better fulfill the requirements of her social role? Even if there are such cultural divergences about the true self, however, my first four conditions, which bear on narrative identity more generally, will hopefully not raise such disagreements. Condition 4), it might be noted, allows that the content of people's identity might show systematic differences from one culture to another. This is not because my account is relativized to Western culture, but rather because different cultures might treat different features as significant for determining who a person is and how she should be treated. That is to say, different features might satisfy my initial four conditions in different cultures.

Finally, condition 7) seems to have the implication that addictions, for instance, can be part of someone's true self, even in the case of an unwilling addict, provided that the condition is stable enough. But surely, it might be objected, this is implausible. Addiction, at least when it is not desired by the person, is the paradigm of a feature that stifles and stands in opposition to the true self. Think of the example of Akratic above, usually treated as a typical case of inauthentic action. Also, there is evidence that recovering substance users tend to understand their true self as separate from their "addicted self", a way of thinking that helps them recover and retain a positive self-image (see e.g. Kilty, 2011). Tony Hope and colleagues have made a

similar point about patients with anorexia nervosa (Hope et al., 2011, p.23). I do not believe, however, that this discredits my condition 7). To think that it does is to beg the question against the key intuitions behind the self-discovery approach to authenticity. Imagine someone who has been unsuccessfully battling alcoholism for years. This person's behaviour, psychology and life trajectory have all been significantly shaped by his addiction. I think it is plausible to say that alcoholism has become part of that person's true self, even if he still doesn't identify with and keeps struggling against it. That said, the objection does touch on an important point, which I have already alluded to earlier on, namely the practical usefulness of the concept of the true self. Thinking of one's true self as good, and excluding from it any undesirable characteristics one is trying to change, such as an addiction, does seem to carry benefits in terms of psychological well-being and recovery success. This makes all the more sense if we remember that people tend to view the true self as "static", something that cannot be deliberately changed. If undesirable traits are seen as part of an unchangeable core, this is likely to be a depressing thought for the agent, possibly undermining his motivation to change. From a *third-person*, philosophical perspective on the question of who someone fundamentally is, I believe that my analysis of the true self provides the most plausible answer. Yet from the *first-person* perspective of a person struggling with an addiction or mental disorder (or from the third-person point of view of a therapist trying to encourage her patient in his efforts to change), moralizing the true self might, sometimes at least, be a more fruitful approach.

I thus do not claim that my analysis of the true self incorporates all the intuitions that people may have about it. It is not meant to. We have seen that several assumptions commonly made about the concept are problematic, at least if our aim is

to propose a way of answering Schechtman's characterization question – leaving aside the potential practical benefits of those assumptions. My analysis aims to preserve, as much as possible, what is sound in the concept, while avoiding those problematic assumptions. It does not gratuitously assume that an agent's true self must be morally good. Indeed, the conception I am proposing allows for the possibility that someone's true self might be downright evil – again, think of Ted Bundy or Dennis Rader. Nor does it commit us to viewing the true self as having the nature of an essence, or as in any way static. In his *Hastings Center Report* piece on authenticity and enhancement, DeGrazia criticizes what he calls the "misleading image of the self as 'given,' static, something there to be discovered" (DeGrazia, 2000, p.35), thereby suggesting that the ideas of givenness and self-discovery must commit us to the implausible notion of a static self. But such a suggestion is itself misleading. Personality traits, for instance, are certainly "given" to us at least early on in our lives, insofar as they initially develop out of an interaction between our genetic endowment and early environmental influences, which are hardly a matter of choice for us.

Could it be retorted that this notion of givenness still becomes inapplicable at least once we have reached adulthood, the life stage at which most potential users of enhancement technologies will likely find themselves? After all, children are already active, from quite an early age, in deliberately shaping themselves. While conceding this, I deny that this shows the concept of the "given" to be irrelevant to adult features. First, children do not all engage in self-creation to the same degree: e.g. some naturally quiet children strive very early on to become more outgoing, whereas others do not. Secondly, as DeGrazia himself acknowledges, we do not have an

unlimited capacity for shaping our personality as we wish (otherwise there would be no interest in new technologies that promise to give us greater control in this regard). Thirdly, some facets of our personality are more amenable to change than others: there is, for instance, much more empirical evidence for the possibility of overcoming one's shyness, than there is for the possibility of turning an introvert into an extrovert. And finally, we should be careful to distinguish between *personality traits* and *behavior*. While the concept of personality does include behavioral dispositions, and perhaps also certain patterns of actual behavior, it includes *affective* dispositions too, dispositions that are not always in line with our behavior. And we clearly have more control over our behavior than over our feelings and desires. An introverted child, for instance, may make an effort to be more outgoing with her peers, as a result of which she may be judged reasonably extroverted by them. But this would not yet mean that she had actually turned herself into an extrovert. Indeed, no matter how often she may push herself to act extroverted, she will in all likelihood still retain the preferences characteristic of introverts, such as needing alone time rather than social time to restore her energy levels, or feeling more comfortable in small groups than large ones. The fact that at least some aspects of our personality, and most of our individual preferences, are unchosen means that there is plenty of room for us to discover who we are, even in adulthood. We may for instance try out different courses and sports at University in order to figure out which ones we enjoy the most, before committing ourselves to these. But this process would make no sense if the self were something to be created entirely, rather than discovered. Nevertheless, we can certainly acknowledge all of this and still recognize that our personality and preferences often change to some degree in the course of our lives, sometimes through our own deliberate efforts at self-creation.

It might be helpful here to clarify how exactly I understand the concept of personality traits. I will chiefly rely on the so-called "Five-Factor model" of personality, in light of the wide acceptance it currently enjoys among psychologists (several authors relevant to our debate, such as Peter Kramer, also rely on something like it when talking about personality). The Five-Factor model, as its name indicates, states that there are five main, stable dimensions (the "Big Five") along which people differ in terms of their behavioral, cognitive and affective styles: Extroversion, Neuroticism, Conscientiousness, Agreeableness and Openness to Experience (McCrae and John, 1992). These dimensions refer, first, to certain stable patterns of behaviour, thought and feeling. Some authors have identified personality traits with such patterns, thereby denying that they can properly be used to *explain* human behaviour (see e.g. Buss and Craik, 1983). I shall follow the main proponents of the Five-Factor model, such as Costa and McCrae, in assuming that personality traits are not mere behavioral (and other) regularities, but are actually part of the *causal* history, and thus of the explanation of people's behaviour (McCrae and Costa Jr, 1995). Accordingly, I will understand personality traits as a set of cognitive, affective and behavioral dispositions that play a role in producing those regularities, dispositions that are physically realized in the person's particular neurobiological constitution. This makes room for the possibility that people might fail to express one of their personality traits on some (even many) occasions, e.g. in compliance with social pressures. Obviously, this possibility would be ruled out if personality traits were simply summaries of people's behaviour patterns. I think my approach is supported by the evidence we currently have on this issue, including the fact that personality seems to be substantially influenced by genetic factors (Bouchard, 2004). It should be noted that

other approaches to personality traits than the Five-Factor model have been proposed, such as Hans Eysenck's model, which posits only three fundamental traits (Extroversion, Neuroticism and Psychoticism), or Mischel and Shoda's "Cognitive-Affective System" (Mischel and Shoda, 1995), which can roughly be understood as positing a greater number of narrower traits, more context-dependent than the Big Five. These alternative models, however, also admit that the relevant traits are grounded in certain dispositions, the roots of which are partly biological, and it should be possible to adapt the remarks I shall make about personality traits to fit any such model were it to prove superior (though things might then look messier, and the vocabulary less familiar, depending on the model).

Though I hope to have successfully refuted DeGrazia's criticism that the true self is an implausible concept because it refers to an imaginary static entity, he would still take issue with the notion, as well as with my account of narrative identity, because they both entail that certain features can come to define who we are independently of the question whether we *identify* with them or not. He thus writes for instance that "whether certain personality traits are definitive of someone depends on whether she identifies with them", and that "who we are has everything to do with what we value" (DeGrazia, 2000, pp.37-8). It should be clear by now that I disagree with those premises. If someone is gay, or a transsexual, or an introvert, these features will at least often meet my seven conditions, and this seems enough for them to be definitive of who that person is, regardless of whether he identifies with them or not. *Pace* DeGrazia, it seems that our refusal to identify with some such feature can actually make it a *more* important part of who we are, if this refusal results in a prolonged psychological conflict with important consequences for us. Think of someone like

Ted Haggard, the American evangelical pastor who had been preaching against homosexuality and opposing same-sex marriage, before admitting that he had used the services of a male escort – following which he had to resign from the various positions he held within the church. At the time at least Haggard clearly did not identify with his homosexual inclinations, yet it seems conceivable that had he found a way to live at peace with them, without hiding them from others, they might have ended up forming a less salient part of his identity than they actually did as a result of his unsuccessful struggle against them (though of course he would then not have been able to become the leader of the National Association of Evangelicals).

Whereas authors like DeGrazia privilege the subject's own viewpoint about who she is in their account of narrative identity, my analysis incorporates Hilde Lindemann's insight that "[i]dentities are not simply a matter of how we experience our own lives, but also of how others see us" (Lindemann, 2001, p.81). It is, however, worth stressing how exactly other people's perception of us impacts our identity on my account. Being gay is e.g. part of Penitent's identity on my view, and this is partly because this property of his significantly impacts how others see and treat him – which in turn impacts how he sees himself, and his motivational set. Yet even if most of the people in his life believe that homosexuality is a serious moral flaw, being morally corrupt will not thereby become part of Penitent's identity, because contrary to what his community may think, being gay does not mean that one is morally corrupt. Others cannot simply make it true that we are whatever they take us to be. By attaching special importance to certain features we *do* possess, however, they can cause them to form part of our identity.

DeGrazia might still insist here that my view is in a sense paternalistic or oppressive, on the grounds that it imposes certain features on people as constitutive of their identity, and denies their fundamental freedom to be the authors of their own life stories. Considering the imaginary example of a woman called Marina, who like some of Kramer's patients takes Prozac to become more outgoing, confident, and less obsessional, he says that "it is ultimately up to Marina to determine what counts as Marina and as not-Marina; the story is hers to writes (within the constraints set by the various factors beyond her control)" (DeGrazia, 2000, p.37). My view, however, is fully compatible with that claim. Where I differ from DeGrazia is simply on the idea that Marina can determine what counts as Marina simply by refusing to identify with the features she does not like about herself, if those features meet my seven conditions. In order to shape her own life story in accordance with her preferences and values, Marina will need to *actually change* those features, e.g. by taking Prozac on a long-term basis. Here again, DeGrazia again seems to be illegitimately assuming that the self-discovery model of authenticity must commit us to the assumption of a static self, one that the person cannot deliberately change, and that the only way of avoiding that assumption is to embrace the self-creation model.

In order to further alleviate the worry that my account of narrative identity paternalistically imposes certain features on people as constitutive of who they are, even if that is not the way they see themselves, I should also stress the following. My account does *not* imply that if some trait is part of someone's identity, it is necessarily crucial that this person ought to *think of herself* in this way. Consider for instance an introverted high school student with a passion for science, who spends much of her free time reading books and conducting small experiments at home. Her personality

and interests are quite salient to her peers and affect how they perceive and behave towards her: they find her rather strange, mysterious and, in fact, not much fun to be around. Consequently, she doesn't get invited to many parties (which she doesn't mind, since she typically has something else, science-related, planned on most evenings). Some of her relatives and close friends tell her she should make an effort to be more sociable and talkative, but she is happy with the way her life is going and sees no need to change her behaviour. Though others view her as very introverted, she does not regard that feature as an important part of who she is. Rather, she mainly sees herself as someone with a vocation for science – keen to understand the world around her, and desirous to emulate the great scientists she admires. Though my view does imply that being introverted is part of that person's identity (and of her true self), it does not necessarily entail that she is at fault for paying little attention to that aspect of herself. Self-knowledge is usually a good thing, but it seems that there can also be value in the sort of spontaneity and lack of self-consciousness that this person displays when it comes to some aspects of her personality. Some might argue that her possession of such qualities actually allows for a purer, more authentic expression of who she is as a person than if she were more reflectively aware of her particular personality style.

Another objection often raised against the idea of a true self starts from the observation, made for instance by William James, that we present very different "faces" to others in different contexts. We might for instance be outgoing with our close friends, but much more shy or reserved when interacting with authority figures, or strangers. From this, the objection concludes that the notion of a true, deeper self

transcending the various masks that we put on and discard according to the demands of our social life is simply an illusion. As Elliott puts it:

> Many theorists, under the influence of Erving Goffman and Judith Butler, argue that social life is *all* performance; that masks are, in essence, our true selves – a notion that dissolves completely the notion of a true self. To think that there is a true, core self apart from the social roles we play, they suggest, is at best wishful thinking and worst pure delusion. Psychiatrist Arnold Ludwig writes, "all the distinctions between true and false selves...make little sense, and have no basis in reality". (Elliott, 2003, p.47)

This objection (which Elliott rightly rejects) flies in the face of the evidence showing that people do have a number of stable core features. Many people show consistency in their adherence to certain commitments, e.g. moral or political. Of course, chameleons and opportunists do exist, but this is not enough to conclude that the world contains nothing but such people. Also, it seems implausible to deny the existence of features like cognitive capacities and personality traits, which, in addition, display significant stability through most of adulthood (sadly, the final years of life are of course typically marked by cognitive decline). Interpreting the fact that people behave differently in different social contexts as refuting the existence of stable personality traits seems to involve a misunderstanding of that concept. The underlying assumption appears to be that if we had such traits, they would *systematically* be manifested in our behaviour. But this is not how personality traits are understood by contemporary psychologists. They do not claim for instance that a high neuroticism scorer will *always* feel depressed or anxious and behave accordingly, including when surrounded by friends at their birthday party. They merely claim that such a person will *tend* to experience and manifest negative affect more than a low scorer in relevant situations, e.g. stressful ones, and that this tendency is not simply a function of the influence of situations on the person. The

objection we are considering does not offer any evidence contradicting the existence of such patterns (and of the underlying dispositions), which is fully compatible with the observation that social context influences our behaviour. (This does not entail that it *entirely determines* it.) As previously alluded to, we may often experience feelings or impulses consistent with a certain personality trait, yet refrain from acting on them, e.g. because doing so would violate the requirements of etiquette or politeness. Therefore, the objection does not warrant the conclusion that we lack a stable, true self. Furthermore, it is not clear that its use of the notion of a social mask makes much sense. This notion is precisely appealed to in cases where there exists a discrepancy between our actual feelings and our behaviour. We hide for instance our nervousness when the time comes to give a presentation at a conference, or our frustration at not having been picked by the session's chair to ask a question after someone else's presentation. But as Elliott points out, the very fact that we are then said to be wearing a "mask" presupposes that we are thereby *concealing* something else, namely our real feelings – an aspect of our true self (Elliott, 2003, p.50). If what the objectors want to say is something different, e.g. that we never in fact experience such a discrepancy and that our feelings and behaviour are entirely shaped by what we take to be the demands of our social environment (hence the reference to "social masks"), then their claim becomes wildly implausible.

Yet doesn't my view of the true self still run into one of the same problems I have pointed out in relation to the "folk" conception? Namely, it doesn't seem able, either, to help us determine what the authentic thing to do is in cases like that of the person dissatisfied with her personality who wants to change it with the help of Prozac, since it acknowledges that someone's true self can encompass such

conflicting features. However, rather than seeing this as an objection to my view, I rather take it to show that we should abandon the assumption, inherent in the popular ethic of authenticity, that getting in touch with our true self is enough on its own to provide guidance as to how to live our lives. As I have previously hinted at, in order to obtain such guidance, we cannot avoid taking a stand on the *value* of the traits that constitute our true self, and of their appropriate expression – an issue I shall elaborate on.

A final objection to my view might be that it is a misnomer to call it an "objective account of narrative identity", because it does not give enough importance to the idea that our identities must fundamentally involve some kind of *narrative*. Authors like Schechtman and DeGrazia, for instance, have proposed accounts of identity that grant a central place to the idea of a *self-narrative*. Though their respective views involve subtle differences, they both entail that the answer to the question "who am I?" is given by my self-told inner story – provided that the story is realistic enough (Schechtman, 1996, pp.93-6; and DeGrazia, 2005b, pp.83-8).[46] By contrast, while I concede that people's inner stories are typically relevant to their identity, my view only treats them as *one* aspect of identity among others.[47] What is more, it seems to me an open question whether that aspect is even a *necessary* one. Consider for instance the few individuals who have experienced severe retrograde amnesia following an accident, and are unable to recall any of their past experiences.

[46] In DeGrazia's case, this looks like a departure from the criterion he presents in his *Hastings Center Report* piece, according to which it is necessary that a person *identifies* with some particular feature for it to be part of her identity. Indeed, it is clearly possible to (realistically) incorporate into our inner story features with which we do not identify: think e.g. of someone who sees himself as a recovering alcoholic.

[47] To be clear: unlike Schechtman and DeGrazia, I do not hold that what makes some feature part of your identity is that it figures in your (realistic) inner story. Rather, I am saying that *your having an inner story* with a certain content (realistic or not) is one aspect of your identity.

Interestingly, despite their lack of episodic memory, these people can still retain knowledge of some general facts about themselves: e.g. where they come from, or that they have worked in an office in the past. It is not clear, however, that such a capacity is enough to provide them with a coherent self-narrative of the sort required by the views of the authors just mentioned (Craver, 2012). Schechtman herself suggests that such people do not actually have a self-narrative (Schechtman, 1996, p.147). Yet even if they don't, is that nevertheless enough to maintain that they also lack an identity altogether? I am not convinced that it is. I would be more inclined to say that they do have an identity (and a true self), though one that has been in many ways curtailed compared to who they were before the onset of amnesia. Indeed, even in such extreme cases, there is an answer to the characterization question in relation to the individual: this answer will mention her personality and preferences, and some key parts of her biography, including e.g. the fact that she had a serious accident a few years ago, and that she has become amnesic as a result.

It might be retorted here that such a criterion for the appropriateness of attributing an identity to someone is far too lax: after all, we can also ask a "characterization question", and provide an answer to it, with regard to non-human animals (who also have personalities, various physical characteristics that can affect the shape of their lives, etc.) and even inanimate objects. Yet we wouldn't normally speak of the (narrative) identity of a dog, or of a paperclip. In reply to this, I agree that talk of narrative identity seems more appropriate when referring to what we take to be the central characteristics of a *person*, a category that rules out dogs and paperclips. The question, then, is whether individuals with severe retrograde amnesia should be considered persons or not. On Schechtman's view, they should not, because they lack

a self-narrative, the possession of which is a precondition of personhood. But here again, I find the proposed criterion unpersuasive. After all, such people still seem to enjoy self-awareness, and the capacity to reason. Isn't that enough for personhood? I will not try to resolve these difficult questions here, but only wished to explain why I am not sure that a self-narrative is even a necessary component of a person's narrative identity, no matter how paradoxical this may sound. Even if there are cases where someone has a narrative identity without a self-narrative, however, my analysis of the person's identity will then still introduce a narrative element, insofar as it will refer to facts about her past, requiring us to tell at least short bits of her life story when describing who she is. Of course, in the case of an amnesiac, this minimal narrative will be a third-person, not a first-person one.

Unfortunately, I do not have the space to discuss the views of Schechtman and DeGrazia in detail here, but I will briefly describe my main source of dissatisfaction with their approach. I have conceded that people's self-narratives are relevant to the definition of their identity: there is no doubt that they influence their affective and behavioral responses to the world. Still, I believe that focusing *exclusively* on the content of such narratives will, in some cases at least, only provide us with an incomplete account of a person's identity. If I am, say, a 32-year old Philosophy student who lived in Switzerland for the first 26 years of his life, this fact is arguably part of my identity, even if it does not figure in my own self-narrative because I have lost all my memories of the first 26 years of my life due to an episode of amnesia. And even in less extreme cases, we normally accept that people can *discover* important aspects of themselves that they were ignorant of. Someone might for instance discover that she enjoys playing chess and is very good at it, so that as a

result the game becomes an important part of her life. She may have had that talent for years before discovering it, and it seems to me that during that time it was already part of her identity (it should figure in an adequate answer to the question: "who is she?"), even though it did not figure in her self-conception. DeGrazia and Schechtman's approach, however, prevents us from saying so.[48]

Schechtman, it is true, distinguishes between a person's *explicit* and *implicit* self-narratives, and defines the latter as "the psychological organization from which his experience and actions are actually flowing" (Schechtman, 1996, p.115). On the face of it, an implicit narrative in that sense is something that can be attributed even to amnesiacs. If so, however, it is not clear to me that it is appropriate to call such a psychological organization a self-*narrative*. Recognizing that such an objection might be raised, Schechtman herself ends up saying that "[i]t does not matter much whether we say that identity is determined by a person's self-narrative or by his psychological organization" (ibid., p.117). I see no reason, then, not to call my own analysis an account of narrative identity.

Having set out that analysis and defended (I hope successfully) the concept of the true self, I will now suggest a possible development of the true self approach to authenticity that justifies criticizing the self-creation projects of my four agents (and those relevantly similar to them) as inauthentic, while avoiding the problems we have

[48] Couldn't they respond to this objection by appealing to the *realism* condition they impose on identity-constituting self-narratives, and argue that this person's inner story would fail to meet that condition, to the extent that it did not mention a key quality she possessed? No. As Schechtman and DeGrazia construe it, the realism condition roughly implies that a person's self-narrative should cohere with *other people's* conception of her (see e.g. Schechtman, 1996, p.95). It does not entail that her self-narrative should mention all the facts relevant to her identity. (If it did, it would already presuppose an independent criterion for the relevance of those facts, and the appeal to the person's self-narrative would become superfluous).

encountered so far, even if we assume that these projects meet Frankfurt and DeGrazia's conditions for authenticity.

3.2 Opportunist's case: authenticity and integrity

Let us begin with Opportunist's case. The suggestion we have considered previously, according to which his self-creation project is inauthentic because he is failing to listen to the "inner voice" telling him how to live authentically, seems to me on the right track. However, it needs to be refined in order to avoid the problems we have already mentioned about it, the main one being that while Opportunist may be betraying his true self insofar as he chooses to produce music that contradicts his own artistic convictions, he is also being true to his desire to further his career, a desire which in his view outweighs competing considerations. In addition, and more importantly for our purposes, Opportunist might also be said to be listening to another "inner voice", the one that suggests to him ways of developing S.'s musical approach. I believe we should concede that Opportunist (and, as we shall see, Penitent and Ex-gay too) is indeed being faithful to his true self in one respect or more, while betraying it in another. This is not absurd if, as I have suggested, a person's true self can in principle involve features that conflict with one another. Yet doesn't this concession undermine the charge of inauthenticity, by yielding the contradictory conclusion that Opportunist is being both authentic and inauthentic? No. What we should say is that Opportunist is failing to listen to the *right* inner voice, or that he is not being faithful to the right aspect of his true self, the one relevant to authentic action. Let me now try to defend that claim.

I would justify the charge of inauthenticity in relation to Opportunist by noting first that – as his name suggests – he is guilty of a lack of artistic integrity. He fails to stick to his artistic convictions when this would have been the right thing to do, and complies with someone else's judgment even though he considers it inferior to his own. On the present suggestion, then, integrity can be understood as a species of authenticity. More specifically, integrity entails having a reliable disposition to act in accordance with certain views, principles or commitments one accepts, and to do so for the right reasons.

What exact sort of views, principles and commitments? The term "integrity" is most often used in a way suggesting that these are of a specifically *moral* nature. However, as indicated by the use of the phrase "moral integrity", other kinds of integrity are also distinguished, such as artistic integrity, the one relevant to Opportunist's case, or intellectual integrity. In those cases, integrity will consist in being faithful to the sort of commitments relevant to the particular sphere of activity in question. Tying these all these varieties of integrity together, some may speak of "personal integrity" to refer to a person's adherence to the commitments she deems most important among all those she has (though the phrase "moral integrity" may also be used in that sense). Some commitments, by contrast, do not seem relevant to any of the types of integrity traditionally distinguished. For instance, someone may have made a resolution to stick to a strict diet, and may take this commitment very seriously. Yet if on some occasion this person yields to the temptation to have some chocolate cake to end his lunch, against the requirements of his diet, it would seem a stretch to call this a failure of integrity. Rather, this person has shown a lack of self-

control.[49] The reason why we may take this view is that such a person is unlikely to understand dieting as a moral ideal or even as the condition *sine qua non* of a "higher" mode of life, or to treat the failure to stick to his diet as meaning that he had done something "beneath him". Rather, it seems more reasonable to assume that he simply has a strong desire to lose weight, that his resolution to diet is a means for him of achieving that goal, and that the failure to stick to it simply means the failure to achieve his goal. Admittedly, it is not *inconceivable* that he should regard dieting, and being slim, in the way I have described as relevant to integrity. Such a way of thinking, however, would be rather unusual, which is why tend to assume that e.g. resisting the appeal of the chocolate cake doesn't count as a manifestation of integrity.

It also seems that we ought to place some further constraints on the sort of views, etc., faithfulness to which can count as manifesting integrity. How substantive these constraints should be, however, is debatable. At the very least, the views in question ought not to be downright *evil*, and the person of integrity cannot appear to us as a complete moral alien. Indeed, attributions of integrity do seem to entail some measure of *praise*. Because of this, few of us would want call Adolf Eichmann a man of integrity on the basis of his unfailing devotion to the cause of National-Socialism. Some have argued in favour of more substantive constraints. Damian Cox and colleagues thus suggest that "attributions of integrity involve the judgment that an agent acts from a moral point of view those attributing integrity find intelligible and defensible (though not necessarily right)" (Cox et al., 2012). I am not sure that such further constraints are required. For instance, I personally do not regard as defensible the Catholic Church's view of homosexual acts as an "intrinsic moral evil"

[49] I accept that showing self-control by sticking to one's resolutions in a praiseworthy way represents a form of authentic action. Nevertheless, it is important to distinguish self-control from integrity.

(Congregation for the Doctrine of the Faith, 1986), or the more basic moral premises on which it is grounded, such as the claim that only sexual acts that can lead to procreation are morally permissible. Nevertheless, I accept that the Pope will show integrity if he refuses to recant his views on homosexuality in the face of public pressures, or if he tries to promote them by non-violent means. There does seem to be something respectable about such a steadfast commitment to what he sincerely takes to be the moral truth.[50] I would, however, no longer feel any respect for his sense of commitment if he were to advocate the extermination of unrepentant homosexuals and of those who support them, even if he sincerely believed this to be required by traditional Catholic doctrine. We may also note that when inviting homosexuals to strive to re-shape their sexual preferences, the Pope is presumably recommending something he sincerely believes to be in their own interest (e.g. increasing their chances of salvation), no matter how misguided his understanding of their interests may be. There is, by contrast, no plausible sense in which someone like Eichmann could be said to have cared about the interests of the Jews when he worked to implement the Final Solution.

I acknowledge that the question of the appropriate constraints to impose on the commitments, etc. relevant to integrity is a tricky one, and I do not have strong views on this issue. Accordingly, I will stick in what follows to the minimal constraint I have mentioned above, while leaving it open whether more substantive constraints, of the sort suggested by Cox and colleagues, should be introduced.

[50] To be clear, what I would find respectable is *not* the Pope's refusal to change his mind on this issue (on the contrary, I believe he ought to change his mind), but rather his refusal to *recant* views that he sincerely thought were correct.

My approach to integrity has some features in common with that of Bernard Williams, who characterized it as consisting in the maintenance of certain core commitments that can be regarded as "identity-conferring" – what Williams also refers to as "ground projects" (Williams, 1981, pp.12ff).[51] These commitments or projects are those that people are most deeply involved in, those that, to a significant degree, give meaning to their lives. Contrary to Williams, however, I believe it is plausible to regard integrity as a *virtue*. Williams argued that it couldn't be, because, he writes,

> while it is an admirable human property, it is not related to motivation as the virtues are. It is not a disposition which itself yields motivations, as generosity and benevolence do; nor is it a virtue of that type, sometimes called 'executive' virtues, which do not themselves yield a characteristic motive, but are necessary for that relation to oneself and the world which enables one to act from desirable motives in desirable ways – the type that includes courage and self-control. (Ibid., p.49)

I find Williams's arguments unpersuasive. Even if we accept his conditions for what can count as a virtue, it is not clear that integrity must fail to meet them, as Cox and colleagues have shown (Cox et al., 2003, pp.70-1). A person of integrity will for instance be better able than someone lacking integrity to withstand temptations to act against her deepest values. In this regard, integrity doesn't seem radically different from self-control, which Williams himself cites as a virtue, though as we have seen, integrity bears on a narrower range of commitments than self-control. Also, self-control doesn't seem to entail the sort of consistent sense of commitment that integrity requires: a person guided by whims could in principle show self-control if acting on a particular whim required him to overcome contrary inclinations like fear or laziness.

[51] Williams, however, does not impose any normative constraints on the content of those ground projects.

Artistic integrity clearly demands that Opportunist stick to his beliefs about what constitutes good music, provided that these beliefs are at least not morally obnoxious. By selling out as he does, he is therefore violating the requirements of this type of integrity. By contrast, even though he is thereby being true to his desire to advance his career, this does not plausibly count as showing integrity of any sort. Opportunists are typically the very opposite of persons of integrity. They are not guided by anything like a moral or personal ideal, but simply by self-interest, and they are not striving to live a "higher" kind of life, or to avoid doing things that they regard as beneath them.[52] Similarly, realizing certain ideas you have even though they do not meet your criteria for sufficient quality isn't a manifestation of integrity, but rather its opposite. If considerations of integrity were the only ones determining the authentic course of action in Opportunist's case, it is therefore clear that refusing to sell out would be the authentic thing to do.

I do believe that the fact that sticking to his own ideas about music would manifest artistic integrity entails that it is at least *prima facie* authentic, yet I am not sure that the question of authenticity can be reduced to one about integrity, even in Opportunist's case. (This will become even clearer when I discuss the cases of Penitent and Ex-gay, who are both plausibly viewed as showing integrity, yet are, if I am right, still acting inauthentically.) Indeed, as I have mentioned earlier, the question of which "inner voice" Opportunist ought to listen to appears to partly depend on the *value* of his own bold ideas, as opposed to his more traditional developments of S.'s

[52] Even if we assumed, rather implausibly, that Opportunist *did* regard the relentless pursuit of career advancement as a higher mode of life, and any departure from it as "beneath him" (perhaps he is a committed ethical egoist), I would still maintain that the "integrity" he would display in his commitment to such a pursuit would be of much lesser value than artistic integrity, and therefore irrelevant to the question of authenticity in his case. (See the following paragraphs.)

ideas. If the latter were significantly superior to the former, it might plausibly be suggested, perhaps not that authenticity would paradoxically call for Opportunist to sell out, but rather that it would require him to revise his opinion of his own ideas, and to follow S.'s advice, as only then would he truly realize his artistic potential. And realizing that potential, one might think, would be intrinsically even more valuable than the integrity he will display by sticking to his own current approach – assuming the values in competition in such a case are not incommensurable. To conclude that the authentic thing to do, full stop (rather than just *prima facie*), for Opportunist, is to stand by his own bold ideas, we therefore need to assume that these are good enough to make it the case that faithfulness to his current views should have greater intrinsic value in his case than any rival way of being faithful to his true self. The intrinsic value in question will therefore include not only the moral value inherent in the virtue of integrity, but also the values of artistic accomplishment and originality. Once this assumption is made, we can also conclude that selling out means doing the opposite of what authenticity requires for Opportunist, and is on that account inauthentic.

Integrity, I have said, involves being true to one's views or commitments for the right reasons. What would that mean in Opportunist's case? It would involve doing so out of adherence to a principle stating that one ought to stick to one's artistic vision, even when doing so means forfeiting opportunities for career advancement (perhaps together with the belief that selling out would be "beneath oneself"); or simply on the basis of a stable disposition to follow one's own artistic ideas *because they are one's ideas*, and one judges them to be good. It might be that people we regard as models of integrity tend to be those who do so as a matter of principle, but it

seems to me that a motive of the second sort just mentioned would also warrant attributing integrity to Opportunist.

Besides artistic integrity, does Opportunist lack *moral* integrity as well? The answer to that question will depend on the further particulars of his case. If Opportunist were the sort of person who does not care for morality at all, being solely committed to advancing his career without much concern for moral considerations of any sort, then he could indeed be said to lack moral integrity. People to whom we are willing to attribute moral integrity must take seriously at least *some* moral rules, even if we disagree with the content of those rules. This entails that even though acting with moral integrity always means acting authentically (in one respect at least), acting in a way that demonstrates a lack of such integrity does not necessarily mean acting inauthentically. A lack of moral integrity will only count as a form of *inauthenticity* when the agent is acting against some moral principles or commitments he accepts, thereby betraying himself. But the person who lacks moral integrity simply because he does not care for morality is not inauthentic on that account. Rather, he is just a scoundrel.

That said, the fact that Opportunist's self-creation project goes against the requirements of artistic integrity does not necessarily imply that he must also lack moral integrity. To some extent, people can display integrity within one domain of their lives but not in others, and we could imagine that Opportunist consistently adhered to certain non-artistic commitments and principles, e.g. prescribing fairness, or loyalty to his friends, but was still prepared to sell out as a composer if he thought his career would significantly benefit from it. However, this is probably an unlikely

scenario, given the place that art occupies in Opportunist's life. If he is prepared to sacrifice convictions as important as his views about what counts as good music for the sake of career advancement, it seems natural to suspect that Opportunist is more generally not a man of conviction.

As I have construed Opportunist's example, he chooses to sell out for purely pragmatic motives, not because he can expect to face intolerable costs if he were to stick to his own ideas. (Making a living out of his music will then require more effort, but it is certainly feasible.) Accordingly, he is open to moral blame. Yet we could also construe his example differently, assuming that refusing to sell out would impose very serious costs on him. Under such an assumption, it would be more plausible to say that Opportunist could have moral integrity, though he had chosen to sacrifice artistic integrity. We can even conceive of situations in which moral integrity actually *required* him to sell out. Suppose he chose to please S. because he knew this to be the only way to earn financial security for his family (whom he cared very much about), who would otherwise face the spectre of poverty. In such a case, Opportunist should certainly be described as having moral integrity, partly *because* he was willing to sell out given the circumstances, even though doing so violated the requirements of artistic integrity. This tragic scenario shows that the requirements of different types of integrity can sometimes come into conflict with one another. The same seems to apply to Williams's fictionalized Gauguin, who leaves his family to struggle in poverty while he goes to Tahiti to realize his artistic potential (assuming such a move is indeed necessary for Gauguin to realize his potential, which might of course be disputed). Indeed, Williams makes it clear that his Gauguin is not indifferent to his

moral obligations to his family, yet chooses to privilege his art (Williams, 1981, pp.22ff).

It might be asked here whether my analysis hasn't just run back into the problem we were trying to overcome regarding the charge of inauthenticity as applied to Opportunist. Namely, I have just acknowledged that there are e.g. cases in which being true to one's moral commitments, and thereby displaying moral integrity, will entail failing to be true to one's artistic commitments, and thereby violating the requirements of artistic integrity. Since I have described integrity as a species of authenticity, am I not committed, after all, to the problematic conclusion that one and the same action can be both authentic and inauthentic? And if so, is there any way it can settle the issue of what would be the authentic thing to do for someone caught in the "family vs. art" dilemma just sketched?

To take the second question first, we need to clarify what it is exactly we are asking. If what we want to know is what it would mean for Opportunist to act with "overall integrity" (rather than any specific type of integrity) in such a dilemma situation, then the correct answer will have to follow a Frankfurtian line of reasoning, according to which we should turn to the agent's hierarchy of values to settle such matters. That is to say, if Opportunist, in the "family vs. art" dilemma, regards artistic integrity as carrying even more weight than his obligations to his family, showing overall or "personal" integrity in his case will mean sticking to his artistic convictions.[53] By contrast, if he judges his moral duties to be paramount, then acting

[53] Which of course doesn't mean that this is *what he ought to do*, all things considered. We might think that in such a situation, Opportunist ought in fact to put his family before his art and sell out, even though doing so wouldn't count as acting with integrity relative to his current set of priorities.

with personal integrity will mean selling out. Indeed, personal integrity is a "higher-order" form of integrity that takes into account *all* the views, commitments, or principles accepted by someone, and concerns the order of priority that this person assigns to all those various views, etc. taken together. However, the question of authentic action in a case like Opportunist's cannot simply be equated with that of personal integrity, no more than with that of artistic integrity. The value of expressing one's best creative powers, already alluded to, is relevant as well, and can sometimes trump considerations of personal integrity from the perspective of authenticity. Consider a case blending Williams's Gauguin and Penitent. In this variant, Gauguin has become a member of the same religious congregation as Penitent some time after his stay in Tahiti. When he presents his paintings of Tahitian women, several of them erotically charged, to other members of the congregation, they are shocked and denounce them as "immoral". Gauguin's works clearly bear the mark of the devil, they tell him. He needs to look to his newfound faith to help him turn away from his dangerous propensity to sensuality. Their words do make an impression on him, and he ends up destroying all of his paintings containing the slightest hint of eroticism. The art he goes on to produce after his conversion mostly consists in *images d'Epinal* depicting the life of the saved after Armageddon, in a sentimental style inspired by his congregation's various publications. It seems to me plausible to say in such a scenario that Gauguin's choice to repress his own artistic originality in compliance with the congregation's prudish ethical code is inauthentic, even though it accurately reflects his highest-priority values. The reason for this is that there would be more value in his realizing his great artistic potential than there is in following the repressive moral code of his congregation.

To reply now to the first question previously asked, my view only allows for the possibility that the same course of action might in principle be both authentic and inauthentic *prima facie* – but not all things considered. In a case like the "religious Gauguin" just described, prioritizing his moral values over his artistic ideas is *prima facie* authentic to the extent that he thereby displays moral (and personal) integrity, yet all things considered, it is inauthentic, because it goes against the realization of his artistic potential, which would be more valuable in those circumstances. If there were cases in which we found the different species of authenticity in conflict (e.g. integrity vs. artistic self-expression) equally valuable, it may be that we ought to declare the respective courses of action they prescribed equally authentic. Or we could say that there won't be such a thing as *the* authentic course of action in such circumstances: rather, each option will be authentic from one particular perspective. However that may be, none of the alternatives will count as *inauthentic* all things considered, because, on my view, this entails that the relevant way of being true to oneself be *less* valuable than the one in competition with it, and we are now assuming that it isn't so. In such hypothetical cases, the concept of authenticity, as I am proposing to understand it, would thus appear unable to provide guidance as to which of the two alternatives we should opt for. (That is not to say that no considerations whatever could decisively count in favour of one of those alternatives even in such cases. Other considerations than authenticity, e.g. fairness, could in principle do so.)

The cases of Opportunist and Williams's Gauguin raise further interesting questions, such as the relation between artistic integrity and so-called "moral luck". Unfortunately, I cannot investigate such questions here. Let me now say a few related words about Ex-gay's case, which is of greater relevance to the coming discussion of

authenticity and enhancement. Isn't Ex-gay displaying (moral) integrity on my view? After all, he is acting in keeping with sincerely held moral convictions, which he inherited from his religious community. My account thus does allow that Ex-gay, and Penitent as well, might be showing moral integrity by choosing to undergo therapy to change their sexual orientation. However, given what we have just said in relation to the religious Gauguin case, that does not mean the view I want to defend is committed to the contradictory implication that their self-creation projects are both authentic and inauthentic. They may well be *prima facie* authentic, to the extent that they manifest moral integrity. Yet all things considered, they are both inauthentic on my view, because both agents are failing to be true to themselves in a way that carries more normative weight than moral integrity does in their case. I will now consider how exactly this failing is to be cashed out in Penitent's case. Similar conclusions will apply to Unconfident.

3.3 Unconfident and Penitent's cases: authenticity as valuable self-expression

Both Unconfident and Penitent can, again, be said to be true to themselves in one respect. Unconfident is acting on his desire not to take any creative risks. Penitent is following his religious convictions. These features can in principle meet my seven conditions for inclusion into the true self. Yet both agents are also failing to be true to themselves, in ways that carry more normative weight. I have already touched on this issue when discussing the case of the religious Gauguin; let us now look at it in more detail.

I believe Unconfident's decision is criticizable as inauthentic provided we assume that bringing his more audacious ideas to fruition would allow him to make his *own* personal contribution to contemporary music, and that this contribution would be a valuable one. He fails to show the courage and determination to "find his own way" as an artist, and as a result, fails to express or realize a valuable part of himself, namely his unique artistic gifts. By contrast, had he shown such courage and determination, he would deserve to be praised as an authentic artist. In a case like Unconfident's, the originality criterion we have previously considered does seem pertinent. However, here again it also seems crucial to assume that Unconfident's bolder ideas, besides being original, are aesthetically sound – unless we understand originality as implying not just that something is novel or unusual, but that it is so in a *good* way. If these ideas were not valuable, and Unconfident's developments of S.'s style were much superior, it would become more plausible to agree with S. that such developments represented the "inner voice" Unconfident ought to follow. We tend to praise those whom we see as realizing their "nature" or their "vocation" in their lives, which can include all sorts of different qualities including being gay, a poet, or an adept of "simple living", yet we only do so on the condition that we see the expression of such qualities as conducive to the "higher mode of life" mentioned by Taylor.

Bernard Williams himself emphasized this sense of the concept of authenticity when describing his own work in an interview: "If there's one theme in all my work it's about authenticity and self-expression... It's the idea that some things are in some real sense really you, or express what you [are] and others aren't" (Jeffries, 2002). We have seen, however, that this distinction between what is "really" you and what is not

isn't always drawn in a normatively neutral manner. Unconfident's case illustrates this, at least if I am right that we will be more inclined to see his more audacious ideas as expressing the "real" him if we assume that they are valuable. Strictly speaking, it may be more accurate to view him as the seat of a conflict between different "inner voices". One such voice tells him not to take creative risks. The reason why this voice isn't relevant to authentic action is that it is simply an expression of pusillanimity, a character defect. Following that voice thus has no intrinsic value, contrary to his new, audacious ideas.

Should Unconfident also be regarded as lacking artistic integrity? I believe not, for contrary to Opportunist, he is more tentative in his assessment of the aesthetic merit of his own ideas. He is not choosing to disregard firmly held artistic convictions. Rather than being opportunistic, he is timid and lacks belief in his own creative powers. Given this, whereas Opportunist does seem to deserve some degree of moral blame, such blame may be less appropriate in the case of Unconfident. Rather, we may simply find his choice of artistic direction highly regrettable.

Penitent's decision can be declared inauthentic along similar lines. The difference is that, as I have noted, Penitent can properly be said to be acting with integrity, since he is being true to his religious convictions. Nevertheless, he is also repressing a part of his identity, his sexual orientation, the appropriate expression of which (after changing his negative view of it) would be more valuable than the sort of integrity he is showing. Penitent, of course, wouldn't agree with this value judgment. On the contrary, he would regard embracing his homosexuality as a serious sin, and by contrast sees great value in his efforts to repress it in order to live in accordance

with the tenets of his religion. But Penitent is mistaken on this issue. He doesn't recognize that his homosexual orientation could, if appropriately expressed, bring him certain intrinsic goods – such as valuable relationships of a specific kind – of greater value than those entailed by his obedience to the repressive teachings of his community on homosexuality.

I will not commit myself to the more controversial claim that gayness is *itself* intrinsically valuable. After all, it seems that homosexuality, and sexual orientation in general, is a feature that can be expressed in both good and bad ways. (Sadly, rapists do exist.) True, we may know some gay people whom we like, among other things, for their particular mannerisms and personal style (attributes we therefore see as having intrinsic value), and we may view these as an expression of their gayness. Yet this may not in fact be an accurate view to take. Besides the fact that not all gays share the same mannerisms and personal style, such features, when present in a particular individual, do not strictly speaking constitute the expression of a *sexual orientation*, but rather of a unique personality. It may be that the latter has partly been shaped by the former (though this is no easy thing to establish), but even if it has, the two nevertheless remain distinct. Furthermore, someone like Penitent is highly unlikely to display a personal style that people would (rightly or wrongly) view as "quintessentially gay", given his strong disapproval of homosexuality and desire to eradicate all traces of gayness in himself. Together with the fact that his homosexual orientation is the cause for him of an inner struggle, involving feelings of guilt and "impurity", this suggests that Penitent is hardly making anything valuable out of his gayness – quite the contrary. I will therefore confine myself to the assumption that Penitent's gayness is the *potential basis* of certain intrinsic goods. This nevertheless

seems to imply more than sheer *instrumental* value: my point isn't that Penitent's homosexual orientation would allow him to achieve distinct intrinsic goods, e.g. fame, that could in principle be achieved through other means. Rather, the intrinsic goods it makes possible would themselves represent particular *instantiations* or expressions of that orientation.

The idea that appropriately living out his homosexual orientation would be intrinsically valuable can also help us understand why Ex-gay's self-creation project might be seen as problematic from the perspective of authenticity, as we shall see in the next section.

3.4 Ex-gay's case: self-respect and the "authentic" self

In section 2.7.2, we have seen that the justification for the inauthenticity charge provided by CTA and CTA+ faced serious problems, even if we added the proviso that the relevant core traits must not be harmful by their very nature. But a more promising way to go is to say that Ex-gay is changing a core trait of his, i.e. an aspect of his true self, the appropriate expression of which would be intrinsically valuable. I shall henceforth refer to such features as "valuable core traits" – a technical phrase by which I mean that the features in question are either intrinsically valuable themselves, *or at least that their appropriate expression can be so valuable*. We would then get an account of the following sort:

Valuable core traits analysis (VTA): Ex-gay's self-creation project is inauthentic insofar as he chooses to change a valuable core trait. We always have a strong (but defeasible)

authenticity-based reason to preserve such traits.

VTA avoids the problematic implications of CTA+, by emphasizing that Ex-gay is changing a valuable core trait. This fact suggests an explanation of why Ex-gay's authenticity-based reason against changing his sexual orientation is not outweighed by his prudential reasons to do so, whereas the charge of inauthenticity does not seem appropriate in a case where someone changes an overly shy personality, or in the sadist's case. Indeed, such personality traits do not count as valuable core traits in my sense. A personality like the sadist's is obviously unpleasant to others, but for *good* reasons, given that it renders him harmful to them and completely oblivious to their fundamental rights. An excessively shy person will often be unable to do what needs to be done if he cannot assert himself at all in his interactions with others. Such traits might therefore be described as character flaws, the expression of which cannot be valuable for its own sake. Yet this description doesn't seem to apply to the fact that someone is gay. This trait only puts Ex-gay (before his transformation) at a disadvantage because of the intolerant attitudes of his community. There is nothing wrong with Ex-gay as he originally is; the wrong lies in those social attitudes, which are unfair and discriminatory.

The case of shyness demands further clarifications. I have conceded before that at least some forms of shyness might endow their possessors with some intrinsically valuable qualities, such as a unique charm. To the extent that those qualities are an *expression* of the person's shyness, and cannot be present without it (i.e. if the non-shy, while they can certainly be charming, can never be so in quite the same way as a shy person), then it seems that the particular instance of shyness in question will be a

source of intrinsic goods and therefore a valuable core trait in my sense,[54] in which case the person will have an authenticity-based reason to preserve it according to VTA. However, if there also are excessive, paralyzing forms of shyness that lack those merits, we will arguably have no authenticity-based reason to preserve them. Contrary to CTA+, VTA does not imply that we have such a reason.

Some might perhaps retort, however, that the drawbacks resulting from shyness, even in its most extreme forms, are in fact dependent on the current Western social ethos, which emphasizes individualism, competitiveness and assertiveness and lacks tolerance for those who do not meet those norms. On that basis, it might be argued that the "excessive" forms of shyness I have described are in fact a socially construed disadvantage. They *could* contribute to a valuable mode of life, but our society's intolerant norms prevent them from doing so. And perhaps they have at least some of the same merits as more moderate forms of shyness, in which case they also count as a valuable core trait. If so, VTA will imply that self-creation projects aimed at overcoming even those forms of shyness raise legitimate concerns about authenticity, no matter how widely accepted these might be in our society. After all, if Ex-gay found himself in a social environment so dominated by homophobia that no valuable way of expressing his sexual orientation were available to him, the proper conclusion to draw would be that action should be taken to change that environment, or that Ex-gay should, if possible, leave it for a morally healthier one – not that his gayness was bad, or a disorder, and that he ought to get therapy for it. Arguably, we shouldn't judge whether some core trait is valuable in my sense by asking whether it can lead to

[54] Those who do not understand shyness simply as an affective/behavioral disposition, but use the term to cover its *manifestations* as well, should conclude that in such cases, shyness itself has some intrinsic value.

valuable modes of life in severely constricting environments marked by intolerant attitudes towards it. The notion of "appropriate expression" of a trait is meant to rule out such inhospitable environments.

The idea that traits like shyness tend to be negatively perceived today chiefly because of an intolerant social ethos is an interesting one. Let me note, however, that even if severe forms of shyness also have the merits claimed above, this may still not warrant the conclusion that we have an authenticity-based reason to preserve them. It will only be the case if we make the further assumption that the valuable qualities or modes of life made possible by such forms of shyness cannot be retained through the process of reducing one's shyness. This assumption seems implausible to me. I don't see that it is impossible for a very shy person to remain e.g. just as charming after successfully working to become less shy (as long as he doesn't go *too far* in the other direction). This point actually reveals a weakness in VTA. Indeed, it seems to run into one of the problems already encountered by CTA+, insofar as it also implies that we have an authenticity-based reason not to improve ourselves in various ways that seem highly desirable, e.g. by working to become less shy in the way just described, or more virtuous. This is implausible. This problem might, however, be corrected quite easily, by adding the proviso that we always have an authenticity-based reason to preserve our valuable core features *unless we can change them without threatening the intrinsic goods they make possible*.

However, the main objection likely to be levelled at VTA is that Ex-gay actually changes his sexual orientation. He becomes straight, which allows him to enjoy a fulfilling heterosexual relationship. Now surely such relationships can be just as

valuable as homosexual ones. But then doesn't this imply that Ex-gay's new sexual orientation makes possible just as valuable a mode of life as the one he could enjoy were he to embrace being gay? If so, why think that he has any authenticity-based reason not to change himself?

We can begin to reply to this challenge by appealing to the proviso we have added above to VTA, and by emphasizing that Ex-gay's homosexual orientation is his *original*, existing one, temporally preceding his new one post-therapy. By changing his original orientation, one might argue, Ex-gay is clearly threatening the intrinsic goods made possible by it: once he has become straight, it will no longer be possible for him to enjoy e.g. a valuable homosexual relationship. True, he will then become able to enjoy a different kind of valuable relationship, one only accessible to heterosexuals. Yet one might suggest that the sort of value peculiar to a gay way of life is of a different *kind* than the value of a heterosexual way of life, and that the former cannot simply be replaced by the latter. To the extent that they are valuable in my sense, existing features, on this view, enjoy a kind of normative priority, and we have a reason (authenticity-based on the present suggestion) to preserve them. In an unpublished manuscript, philosopher Jerry Cohen has defended a view of that kind (though he doesn't present it as an analysis of authenticity), arguing that we have a (defeasible) reason to refrain from producing even a greater sum of value than currently exists if doing so means destroying *already embodied* value, the value towards which it is right to be biased (Cohen, unpublished).[55]

Cohen's defense of the normative significance of existing features in some cases

[55] A different version of this article is available in Cohen and Otsuka, 2012.

is more thoughtful than that found in the writings of most critics of enhancement technologies. Most of us would no doubt be horrified if our friends and relatives suddenly started radically transforming their appearance, personality, or sexual orientation, perhaps on the grounds that they just wanted to "try something new", and the reply that their new identity was just as valuable as their original one would be unlikely to appease us. As Cohen argues, we value others *for the particular persons they are*, and not just because they embody a certain total amount of intrinsic value that can in principle be realized in many different ways. Cohen's view may help explain why we are sometimes tempted to think that a trait's being "natural", in the sense of not having been deliberately shaped by us, gives us a reason to preserve it: when a trait that is "natural" in this sense instantiates a particular kind of value, we sense that if its possessor were to deliberately modify it, even to replace it with an equally valuable one, the specific kind of value embodied in the original trait would be lost. This leads us to conclude, rightly, that the person has a reason to preserve the original trait, but also, and possibly mistakenly, that this reason is grounded in the trait's being "natural". Naturalness may or may not sometimes have such normative force, yet Cohen's view has the advantage of not committing us to the controversial assumption that it does. Furthermore, it can explain the normative priority of existing traits even when they are not "natural" ones. Imagine that Influenceable, in our example above, had through prolonged effort cultivated qualities like self-discipline and industriousness to the point that they had become "second nature" to her. She moves to a new town to start a University education there. Her new social circle now consists largely of carefree pleasure-seekers who introduce her to smoking, binge drinking, and other extra-curricular activities of the same kind. She realizes that if she keeps spending time with those people, she will soon become like them: her studies

will take a back seat, and the habits of mind she had spent years developing will be a thing of the past. It doesn't seem to me that concerns about authenticity, or remaining "herself", i.e. disciplined and hard-working, would be inappropriate in Influenceable's case, on the mere grounds that those traits weren't "natural" ones.

While I believe that something like Cohen's intuition underlies many of the worries about authenticity in the context of the enhancement debate (as we shall see in part 4), appealing to it in its initial form might not be most convincing in Ex-gay's case. It would work best if one assumed – as I prefer not to – that his homosexual orientation itself has intrinsic value, even though he doesn't currently express it in valuable ways. Indeed, Cohen talks about the need to preserve already embodied value, not just already embodied potential sources of intrinsic value. We could, however, amend his view to cover these as well.

The fact that Ex-gay is forfeiting a valuable existing aspect of his true self does seem to make his self-creation project regrettable. Yet I believe that what makes it morally disturbing – rather than just regrettable – is a distinct, though related fact: namely, he is failing to show a proper appreciation of his initial (i.e. "authentic"), gay self. Put differently, Ex-gay shows a lack of *self-respect*. As Chris Gowans puts it, "to have self-respect implies having a proper appreciation of one's value or worth, and living one's life in a manner that is consonant with this appreciation" (Gowans, 2006, p.104). I should emphasize that what Ex-gay is failing to properly value is one particular aspect of his own *individual* identity, rather than the properties that make him a *person*, which he shares with all other persons. Properly valuing the latter is what self-respect is generally taken to be about in the contemporary philosophical

literature. It is, for instance, what Robin Dillon refers to when she speaks of "recognition self-respect" (Dillon, 1992, p.133). However, it is not clear that either Ex-gay or his community must necessarily fail to properly value his personhood. They could of course have a moral code declaring that homosexuals were somehow inferior beings whose interests did not count as much as those of heterosexuals, the way racists think of other ethnicities than theirs, thereby disregarding what is owed to Ex-gay in virtue of his being a person. Yet they need not do so. It is certainly possible that they might recognize Ex-gay's status as an equal person among persons, and treat him accordingly, while still holding that he had the misfortune of being tainted with "impure" sexual preferences, and that he had a duty to fight against these. Think of the way our society treats pedophiles. Though we deem it inappropriate to blame them for the desires they have, to the extent that these desires are not the product of their choice, we do consider it fit to pressure them to work on themselves with the help of a health professional, so that they may learn to manage their urges and ensure they don't act on them. Ex-gay's religious community might take a similar attitude towards homosexuals. What they and Ex-gay fail to appreciate is thus not necessarily his intrinsic worth *qua* person, but rather the fact that homosexuality, contrary to pedophilia, is not a morally problematic sexual preference that is best eradicated. On the contrary, as I have said, homosexuality can provide the basis for valuable mode of existence. Were Ex-gay able to appreciate this, and to accept himself, despite the unpropitious environment in which he finds himself, he would deserve to be praised for his sense of self-respect.

On that basis, we can reply to the above objection by saying that even though a heterosexual way of life need not be intrinsically inferior to a homosexual one, and

even though Ex-gay may certainly enjoy a highly valuable relationship with a person of the other sex after his transformation, the choice to change himself in keeping with his religious convictions, and the life that results from it, are *not* in fact as valuable as the path of self-acceptance would be. They are *less* valuable, even taking the value of Ex-gay's moral integrity into account. Indeed, if the analysis I have just presented is correct, Ex-gay manifests a significant (though perhaps excusable) moral failing in making the choice he makes, as he shows a lack of self-respect. And we can expect that the rest of his life will be shaped by this morally problematic choice.

Such a line of argument provides grounds for finding Ex-gay's self-creation project morally problematic. Yet why claim, more specifically, that it is *inauthentic*? Why not simply say that it betrays a lack of self-respect, and that it involves forfeiting a valuable trait, explanations that might be considered more illuminating? The reason why I think that talk of inauthenticity is appropriate (though not necessarily mandatory) in Ex-gay's case is that, as I have previously suggested, it involves the loss of a valuable aspect of his "authentic" self, i.e. his original one, temporally prior to the new one he develops after going through therapy.

This analysis might be expressed through the following revised version of VTA:

VTA+: Ex-gay's self-creation project is inauthentic insofar as he chooses to change a valuable core trait; and it is morally problematic to the extent that it demonstrates a lack of self-respect, in the sense of a failure to properly appreciate his original, authentic self. It is always wrong (though it can be excusable) to act in such a way.

It seems to me that VTA+ captures the key intuitions underlying the charge of inauthenticity in cases like Ex-gay's while avoiding the problems encountered by the previous variants. VTA+ does not imply that is always wrong to modify some of our core features, even to a significant degree – on the contrary, it is compatible with the view that it is sometimes (perhaps even often) morally acceptable, and authentic, to do so. It only claims that it is always wrong to change our *valuable* core features, *when doing so demonstrates a lack of self-respect* – a claim I take to be much more reasonable. (Can it ever be permissible to disrespect ourselves?) Neither does VTA+ imply that the sadist described above has any authenticity-based reason to remain the way he is, given that just like VTA, it is only concerned with valuable core features in my sense. Finally, unlike VTA, VTA+ can satisfactorily explain why we may find Ex-gay's self-transformation morally problematic even though a heterosexual way of life need not *in principle* (leaving aside the details of Ex-gay's case) be inferior to a homosexual one.

The appeal to ideas like self-respect, and proper appreciation of one's present self, actually allows us to bring my four cases together. Penitent shows a lack of self-respect just as Ex-gay does. Unconfident also fails to properly value an aspect of his true self, namely his own creative powers and the ideas he has as a result of their possession. It is true, though, that we are likely to say he lacks *self-belief*, rather than self-respect. It seems that we regard a different sort of language as appropriate to features like talents and aptitudes than to traits like sexual orientation, as the latter are not valued for their ability to produce admirable artefacts or performances. This difference, however, is a minor one. Finally, Opportunist can also properly be said to lack self-respect. The relevant sense of self-respect in this version of his case,

however, is a somewhat different one, which has more to do with meeting certain *standards of behaviour* than with properly valuing one's present identity. Elizabeth Telfer calls this "conative" self-respect: she characterizes it as "a desire not to behave in a manner unworthy of oneself, or a disposition which prevents one from behaving in a manner unworthy of oneself" (Telfer, 1968, p.115). She adds that "[s]ome things are thought to be unworthy of the man *qua* man, i.e., unworthy of anyone; other things are unworthy of him in view of some role or status or special quality" (ibid.). The relevant role in Opportunist's case is, obviously, that of artist: there are things that are unworthy of an artist to do, and these include betraying your own artistic views for the sake of career advancement. Had Opportunist chosen, for the right reasons, not to sell out, he would have provided evidence of possessing the disposition not to behave in a manner unworthy of an artist, i.e. of having conative self-respect. By contrast, in my original example, Opportunist lacks such a disposition, and consequently conative self-respect.

That said, the connections I have drawn in this section between my four cases actually conceal the presence of two *distinct* senses of authenticity, each corresponding to a different sense of the idea of being faithful to one's true self. I shall now undertake to tease these two senses apart.

3.5 What, then, is authenticity?

To sum up my analysis so far, Unconfident, Opportunist, Penitent and Ex-gay (before his transformation) are all acting inauthentically in a similar sense, which also explains why we may call inauthentic the self-creation projects of the first three

agents. This sense is captured by the following specific definition of authenticity, which centers on the first of the two senses of "being faithful to one's true self" that I have distinguished earlier in this dissertation:

Specific definition of authenticity, no.1: the intrinsically valuable expression[56] of one's true self. Failing to express one's true self when doing so would be intrinsically valuable is inauthentic. In cases where expressing some aspect of one's true self would conflict with another way of being faithful it, and both options are intrinsically valuable, both are *prima facie* authentic. If one is more valuable than the other, it is the authentic thing to do, full stop. If both are equally valuable, they are both authentic, full stop.

My four agents are all failing, at some point at least, to express a particular aspect of their true self when doing so would be intrinsically valuable. And part of the reason why it would be so valuable is that the feature they are repressing is itself valuable, in the sense given above. The proviso contained in this first definition allows us to deal with cases where someone's true self is conflicted: authenticity then requires us to be faithful to the aspect of our true self the expression of which is more valuable (or to any of the aspects in conflict, if expressing either is equally valuable). An analogous solution applies to cases (assuming there are any) where someone has more than one true self, as suggested in relation to bipolar disorder. Authenticity will then require the person to be true to herself in the way that is most valuable. This could mean privileging one specific true self among those in conflict, or expressing any of these (if the alternatives are of equal value), or both in succession, or acting on her desire to be free of the disorder and get treatment – depending on how we assess

[56] "Expression" here is meant to stand for the disjunction "expression or realization", as appropriate to the case under consideration.

the intrinsic value of these various options. (Again, whatever the exact answer, it does not yet mean that considerations of authenticity must necessarily be decisive in such cases: it may well be that other considerations, e.g. relating to the person's best interests, carry even more weight.)

The reason why this first definition cannot justify calling Ex-gay's self-creation project inauthentic is that, contrary to the other three agents, he is actually *changing* himself. While he is indeed failing to live out his sexual orientation before undergoing therapy, this is no longer so afterwards. To capture his case, we thus need a different definition of authenticity, focused on the second sense of faithfulness to the true self distinguished previously:

Specific definition of authenticity, no.2: the preservation of one's true, authentic self, when the traits being preserved are valuable, and such preservation warrants praise (for instance because it demonstrates a strong sense of self-respect). Failing to preserve one's authentic self when such conditions are met is inauthentic. In cases where preserving some aspect of one's authentic self would conflict with another way of being faithful it, and both options are intrinsically valuable, both are *prima facie* authentic. Etc. (same proviso as in definition 1).

The repetition in this definition of the proviso contained in the first one is necessary to avoid potential clashes between the two, which would lead to the conclusion that a particular course of action can be both authentic and inauthentic (full stop, rather than just *prima facie*). As we have seen, Ex-gay's self-creation project is arguably authentic in my sense 1, but it is inauthentic in sense 2. And if I am right that the latter is more valuable than the former, the proviso entails that the charge of inauthenticity does apply in Ex-gay's case; by displaying integrity, he is

only being *prima facie* authentic. Conversely, we can imagine a scenario in which Penitent has become convinced that his community is wrong about the morality of homosexuality. He decides that he wants to live his life as a gay man. However, he knows he will not be able to do so if he stays in the community: their attitudes are highly unlikely to change, and he cannot expect to ever be able to accept himself as long as he remains among them. The only solution for him is to leave the community and join a support group for gay Christians. Yet he also fears, let us assume with good reason, that if he takes that path, he will likely lose some of the valuable qualities that his membership of the community had fostered in him: for instance, he can expect to become more individualistic and less focused on helping those in need, having observed such an evolution in others who left the community before him. In such a scenario, the authentic thing to do for Penitent might still be to leave the community and learn to accept himself as a gay man. Still, choosing to stay in order, among other things, to preserve the qualities he is afraid of losing might nevertheless count as *prima facie* authentic, to the extent that these qualities are valuable ones. But my second definition will only declare a deliberate modification to one's true self inauthentic, full stop, if the value of preserving the relevant feature (which will depend of the particular value of that feature itself, and/or on the virtues that such preservation might manifest) would be greater than the value of listening to any "inner voice" condoning the change.

It might be asked whether we should, according to my two definitions, regard authenticity as a *virtue*. After all, I have suggested that authentic actions warrant praise, and have also stated that integrity, which I acknowledged was a virtue, could be understood as a species of authenticity. While granting that there are virtuous

forms of authenticity, integrity being one example, the two definitions I have offered arguably do not entail that authenticity is fundamentally a virtue. Indeed, they allow that somewhat isolated choices and actions might in principle count as authentic. We may also say that if a person performs enough authentic actions (in the senses given by my two definitions) throughout her life, especially on key occasions, she will count as having lived an authentic life. But none of this yet implies that the person who performs those actions possesses a *reliable* disposition to act in such ways, for the right reasons – a disposition which, as noted for instance by Julia Annas, is a pre-requisite for the possession of a genuine virtue (Annas, 2011, p.9). Indeed, the person in question may never in fact encounter demanding circumstances in which authentic action is called for. Perhaps she was just lucky to be able to live authentically, but wouldn't have managed to do so had she been really put to the test. If so, this person won't plausibly count as possessing the *virtue* of authenticity, even though she may have led an authentic life and performed many authentic actions. My two definitions thus seem to treat authenticity as an "admirable human quality", to borrow Williams's description of integrity, one that *can* sometimes amount to a virtue, but need not do so. We could, however, define authenticity in a way that does turn it into a full-fledged virtue: we could for instance offer two specific definitions of authenticity as a virtue corresponding to the two I have offered, according to which it consists in a reliable disposition to act authentically in my senses 1 and 2. Or we could have a more general definition that included both senses:

General definition of authenticity. Authenticity as a virtue: authenticity is the reliable disposition to be faithful to one's true self, in a way that is praiseworthy, and intrinsically

more valuable (or at least no less valuable) than any rival way of being true to oneself in the circumstances.[57]

My two specific definitions of authenticity, I believe, follow common uses of that term, and for that reason are not as demanding as this general one. Like the definitions offered by authors like Frankfurt and DeGrazia, they are focused on characterizing authentic *choices* and *actions*, and they do not presuppose that these should necessarily manifest the *virtue* of authenticity, as expressed in my general definition. They allow that authentic actions might simply represent what Rosalind Hursthouse calls "everyday virtuous action", that is, action that conforms to the virtue that gives it its name, but isn't performed by an agent actually possessing that virtue (Hursthouse, 2010, p.317). It might be that when we describe a *person* as authentic, rather than her choices, actions, or even life as a whole, we are indeed presupposing an understanding of authenticity as a virtue. Even so, however, I take such a general definition to be less helpful than the familiar concepts we already use to refer to virtuous forms of authenticity of a more specific kind, such as integrity or, as we shall see in section 3.7, sincerity (though such virtues only entail *prima facie* authenticity, and not necessarily authenticity full stop, as my general definition does). Moreover, as we have seen, acting authentically in the senses given by my two specific definitions will often warrant praise mainly because it demonstrates distinct, more familiar virtues than authenticity, such as strength of character and self-respect. I believe it makes for greater clarity to speak about such better-known virtues, whenever

[57] This definition does not presuppose the unity of the virtues. It allows that one might in principle manifest the virtue of authenticity by acting in a way that still isn't the right (i.e. fully virtuous) one in the circumstances. If one accepted the principle of the unity of the virtues, one would have to define authenticity even more stringently, as the reliable disposition to be true to oneself, for the right reasons, *when doing so is appropriate* or the right thing to do.

possible, in relation to the sort of cases that interest us, rather than talk about authenticity as a virtue in the general sense given above.

What do my definitions imply in the various fictional cases I have described in the earlier sections of this dissertation? Given that it presupposes certain value judgments about the relevant aspects of a person's true self, or at least about the expression of these, the verdicts yielded by my analysis will themselves be influenced by such judgments. I would tend to judge Akratic's choice to accept the cigarette inauthentic, because it contradicts her desire to stop smoking, which is likely part of her true self, and because acting in accordance with that desire would have been both praiseworthy and intrinsically valuable. Resisting the urge to smoke when one is addicted to smoking takes an admirable amount of self-control. And there is arguably more value in the desire to promote one's health and longevity than in the desire to experience the ephemeral pleasure of a cigarette regardless of the long-term costs. The same verdict seems appropriate in the case of Influenceable, for similar reasons. Any attitudes she may have that are in accordance with her choice to take the cigarette are probably not stable enough to count as part of her true self. Her desire to stop smoking is a more plausible candidate (and we have noted the likelihood that she is self-deceived). An unrepentant smoker like Rebel, of course, would disagree with the value judgments underlying my verdicts. Nonetheless, we may note that these verdicts coincide with those yielded by the views of Frankfurt and DeGrazia.

That said, divergences will likely arise regarding the other three cases I have described. Rebel's decision to take the cigarette will at least not count as inauthentic on my view, since she is not failing to be true to herself in any sense. I am not sure,

however, that we should declare it authentic either, since on my account such a characterization entails praise, and it isn't clear to me that we should praise unrepentant smokers for consistently indulging themselves – viewing this as a form of integrity, for instance, wouldn't seem particularly plausible. I would say the same about Ex-rebel's case (whose choice is not inauthentic either, since her true self back then was different from what it is now). Finally, Indoctrinated actions at the service of the sect will not count as authentic either on my view, because his desire to obey the guru is not deep enough to count as part of his true self – indeed, he relinquishes it once he is given the opportunity to think critically about the influences behind it. I am not sure, though, that we should declare Indoctrinated's desire to serve the sect inauthentic either in my sense, because it is not clear that acting on it meant betraying his true self – after all, during those ten years, he did not yet believe that the guru was a fraud, though he did eventually come to that conclusion after his relatives got him out. True, he may well have had the desire not to be manipulated, a desire the satisfaction of which was incompatible with his actions at the service of the sect, even though he didn't realize it at the time. But first, this desire need not have been part of his true self. His competing desire to belong to a community may e.g. have prevented it from playing a significant role in his motivational set (which might explain how the sect managed to get him). Secondly, even if we assume that such a desire was indeed part of his true self, we are also assuming that he didn't serve the sect *knowing full well* that they were manipulating him. To echo a point made previously about the Frankfurtian view, if we want to call Indoctrinated inauthentic on the grounds just mentioned, we will also have to conclude that people with a deep desire to live virtuously are acting inauthentically whenever they choose, without realizing it, a course of action that isn't the fully virtuous one. Charging such people with

widespread inauthenticity simply because of the fact of human imperfection strikes me as excessive. It seems both more charitable and more plausible to grant that they are being true to themselves, provided that they genuinely *try* to act in accordance with their commitment to virtue, and do not display weakness of the will.

Here is how we may schematically represent the verdicts reached by, respectively, the Frankfurtian account, DeGrazia's view, and my own account about the choices made by the agents in the main examples I have described:[58]

	Authenticity as wholeheartedness	DeGrazia's account	My own account
Akratic	Inauthentic	Inauthentic	Inauthentic
Rebel	Authentic	Authentic	Neither authentic nor inauthentic
Influenceable	Inauthentic	Inauthentic	Inauthentic
Ex-rebel	Authentic	Authentic	Neither
Indoctrinated	Authentic	Inauthentic	Neither
Unconfident, Opportunist, Penitent & Ex-gay	Authentic	Authentic	Inauthentic

3.6 Objections and replies

A salient fact about this comparison table is that my own view seems much more demanding than the other two, since it yields a verdict of inauthenticity in most cases, and doesn't regard a single one of those agents as making an authentic choice. This might be said to render it implausible. To my defense, however, I have made it clear that people like Akratic, or my four main agents, would have acted authentically if they had made the contrary choice to their actual one, for the right reasons.

[58] I haven't included the sadist's case as it involves the use of an enhancement technology, an issue the discussion of which I am reserving for the final part of this dissertation.

Secondly, given that I have characterized authenticity as an admirable quality and, in some cases, a virtue, that most of the agents in this list do not count as authentic on my view shouldn't really surprise us. Acting virtuously is often difficult, to which we may add that the cases of the composers, Penitent, and Ex-gay were precisely selected as paradigmatic illustrations of inauthentic action, of a sort neglected by views like those of Frankfurt and DeGrazia.

It might also be objected that my criterion for assessing the authenticity of those decisions, to the extent that it rests on contentious value judgments, appears problematic, at least in the cases where its verdicts differ from Frankfurt and DeGrazia. Surely, it might be said, Rebel and Ex-rebel's decisions to accept the cigarette at the party are authentic, whereas Indoctrinated's actions while in the sect are inauthentic. In response to this, I am willing to concede that my view may not be ideally suited to all the cases we have considered. That is not because the verdicts it yields in cases like the three just mentioned are just plain wrong, but rather because when applied to such cases, the question whether the agents' choices are authentic will typically center on a different sort of concern than the one addressed by my analysis. Authenticity seems to be, as Ned Block put it in relation to the concept of consiousness, a "mongrel concept": it admits of a variety of different uses, both among philosophers and non-philosophers (Block, 1995, p.227). One major construal of the authenticity question, when asked from a third-person perspective, is: is the person's decision really *her own*, no matter how exactly this notion is to be cashed out (e.g. by reference to her highest-priority values, etc.)? This is the construal most relevant to cases like those of Akratic or Rebel, and the one to which accounts like those of Frankfurt and DeGrazia are best suited. On the other hand, the question is

also sometimes taken to mean something like: is this person expressing what is good in herself, or realizing the valuable aspects of her nature? This construal is of greater relevance to cases like my four main ones, and I believe is best captured by a version of the true self approach to authenticity like the one I have defended. There is no doubt that different people will disagree on such questions as whether a homosexual way of life can be intrinsically valuable. Let me stress again, then, that the adequacy of the analysis of authenticity I have offered does not crucially depend on my being right in my assessment of Penitent and Ex-gay's self-creation projects as inauthentic. The key assumption underlying my analysis is, rather, that we cannot avoid introducing such value judgments when addressing the authenticity question in cases of that sort, which as we shall see include many cases pertinent to the enhancement debate.[59] People with a conservative outlook, similar to that of the fictional members of Ex-gay's congregation, will disagree with my verdict about his case. From their perspective, he presumably constitutes a paradigm of authenticity or integrity. Yet they can still accept the definition of authenticity I have offered. They may well reason that Ex-gay's fidelity to his religious convictions is an admirable example of moral integrity, whereas embracing his homosexuality would, on their view, represent a serious sin, a way of being true to himself that would have significant *negative* value. With those value judgments as inputs, my analysis will yield what they regard as the correct the conclusion, i.e. that Ex-gay is doing the authentic thing. Knobe's research suggests that people's disagreements in their judgments about authenticity are ultimately often due to such differences in value judgments. It seems to me that

[59] In line with this, I agree with Taylor's claim in *The Ethics* that authenticity as a moral ideal presupposes what he calls "horizons of significance", or the assumption that certain things matter independently of whether we happen to desire or choose them (Taylor, 1991, chap.4). I am less sure that, as he further claims, such horizons must – even if our focus is on authenticity – necessarily entail demands "emanating beyond the self" (p.40). Unfortunately, I do not have the space to discuss this issue in more detail here.

my account helps bring this out.

Let me conclude this section by replying to two further objections to my analysis of Ex-gay's case, and my second specific definition of authenticity. First, I have acknowledged that our true self changes through the life course, to quite a significant degree: our personality and interests at, say, seven years old are not the same as those we have thirty years later. Inspired by Cohen's ideas, I have also suggested that the value embodied in our already *existing* self enjoyed a kind of normative priority, from the perspective of authenticity. Now might our seven-year-old self not instantiate a particular kind of value that our 37-year-old self lacks, such as for instance a greater innocence and capacity to marvel at simple things? Granting that it can, suppose that after your seventh birthday, you were presented with a "Peter Pan pill" that could put your development on "pause", allowing you to retain the true self you had at that age for the rest of your life.[60] Does my account imply that you ought to take the pill? If so, this would arguably count against it. Secondly, while my analysis might imply that Ex-gay is acting inauthentically by choosing to undergo reparative therapy, what does it have to say about him *after* his transformation? I am assuming that he has truly become heterosexual, in other words, that he has developed a new true self. And I have granted that this true self may be just as valuable as his original one. On my view, then, does authenticity require Ex-gay to *preserve* this new, heterosexual self? If he could go through a new course of therapy that reversed the effects of the first one, making him gay again, should he forfeit it in the name of authenticity? Also, can I really claim that his life post-transformation is inauthentic, if he is then expressing his new, heterosexual self in valuable ways? If I cannot, doesn't

[60] Such an example is considered by Nicholas Agar in his book *Humanity's End* (Agar, 2010, pp.187-8).

my view entail that the transformation has actually made Ex-gay's life *more* authentic, since he was initially repressing his homosexual self (and therefore living inauthentically), but is afterwards able to fully live out his new sexual orientation?

In reply to the first objection, I am happy to admit that if there were such a thing as the Peter Pan pill, a child may have a *reason* to take it, precisely because it would then preserve the valuable qualities mentioned above that tend to be the prerogative of children. There doesn't seem to be anything irrational about regretting the loss of the spontaneity, capacity for amazement, or carefree mindset that we had as children but that our more sophisticated adult selves no longer enjoy (or less so). However, my account doesn't commit me to saying that the reason we may have to take the Peter Pan pill is a *decisive* one. We may still think that adult life, and the various goods it makes possible (such as the realization of our best capacities), is superior to the life we lived as children, which makes it a good thing, all in all, to grow up. I thus do not see that my account must have problematic implications in this respect. Regarding the second objection, I agree that Ex-gay will, after enough time has elapsed, acquire a new, valuable true self as the result of his course of therapy, and am even willing to concede that his life after the change might be more authentic than it was before it, since he will then be fulfilling his nature rather than living in repression and guilt. However, it remains that he will not have resolved the problem of the inauthenticity of his life in the right way, since he will paradoxically have done so by betraying himself. As a result, even if his future life can be described as authentic, it will remain tainted by the fact that it was significantly shaped by an inauthentic choice involving a lack of self-respect. Because of this, I believe it could be authentic, full stop, for Ex-gay to undergo a "counter-therapy" to revert back to his previous self, if for instance

he had changed his mind about the morality of homosexuality, and wanted to express his new beliefs and symbolically "undo" his original inauthentic transformation. It would nevertheless depend on his exact circumstances: if he now had a wife and children to whose well-being he was committed, it may well be that the authentic thing to do for him would be to act on that commitment and to preserve his new self, rather than compromising his relationship with them by changing himself again (since his own personal fulfillment would no longer require such a change).

3.7 Authenticity and related notions

Having offered my own analysis of authenticity, let me now say a few more words about how this notion relates to others often associated with it. Integrity is one example; we have seen that it can be understood as a species of (virtuous) authenticity in the sense of valuable self-expression. The same applies to another virtue often associated with authenticity, that of sincerity.[61] To the extent that sincerity essentially consists in presenting oneself accurately to others, it is already included in DeGrazia's concept of honesty, which we have previously discussed. Being sincere, just like acting with integrity, does entail being *prima facie* authentic in the sense given by my first specific definition, and to be insincere is to be *prima facie* inauthentic in that sense – but not necessarily *all things considered*. Consider Benjamin Constant's example of the murderer who knocks on your door and asks if your friend, whom he is after, is inside. You tell him no, even though your friend is actually hiding at your house. Let us assume you are thereby acting on a firmly held principle enjoining you to protect your friends from harm if you can. Your behaviour is insincere, as you are

[61] An association illustrated e.g. by Lionel Trilling's famous book *Sincerity and Authenticity* (Trilling, 1972).

misrepresenting your actual beliefs to the murderer. On that account, it is *prima facie* inauthentic. But surely you are not to be criticized as inauthentic. On the contrary, you deserve praise for your authenticity, to the extent that you are being faithful to your commitment to defending your friends – adherence to which is infinitely more valuable than truth-telling in such circumstances. By contrast, you might count as authentic, full stop, if you were to tell the truth to the murderer, if e.g. you were a strict follower of Kant who thought that concern for the truth should trump your concern for your friend even in such circumstances. However, most of us would think that any integrity you might thereby display would carry little normative weight. Much more significant would be the fact that you would be unnecessarily putting your friend's life at risk, on account of which you would, all things considered, deserve blame. Also, as my four main cases demonstrate, it is possible to act with perfect sincerity and yet still count as inauthentic, all things considered.

We have already touched on other notions often associated with authenticity: originality, autonomy, and self-fulfillment. *Pace* Taylor, I have argued that authenticity need not require originality, even though it is certainly true that people we regard as paradigms of authenticity, such as great innovators in the Arts or Sciences, often show significant originality in their accomplishments and way of life. And despite the close link usually drawn between the notion of autonomy and that of authenticity, I have argued that one can in principle autonomously choose to act inauthentically (in a key sense of this term), and conversely, that there might be examples of authentic actions that aren't autonomous, as illustrated by the Huck Finn case.

Finally, we have seen that, according to Taylor, contemporary Western culture tends to regard authenticity as going hand in hand with self-fulfillment. Taylor himself does not reject that association. He merely thinks that our culture has failed to grasp what self-fulfillment really requires, namely being concerned with demands that emanate from beyond our own desires, and treating our relationships with others as more than just instrumental to our fulfillment (Taylor, 1991, pp.72-3). However, Taylor does not clearly distinguish between two possible meanings of "self-fulfillment", which he both uses himself in *The Ethics*. Alan Gewirth does make that distinction in his book on the concept of self-fulfillment: he distinguishes between aspiration-fulfillment and capacity-fulfillment. Aspiration-fulfillment roughly refers to the satisfaction of our deepest desires, with the resulting feeling of satisfaction that it produces – something many people call happiness. Capacity-fulfillment, on the other hand, is very close to the notion, familiar to psychologists, of self-actualization: Gewirth uses it to describe the realization of one's best capacities (Gewirth, 1998, pp.13ff). The latter form of fulfillment actually constitutes one type of authenticity according to my definition of the concept. The former is a very similar concept, yet it is nevertheless distinct, given that on my view authenticity consists in the valuable *expression* of oneself. And it is possible to express one's deepest desires in one's behaviour without yet fulfilling them, as the rejected suitors of this world know all too well.

When he uses the term "self-fulfillment", Taylor sometimes refers to aspiration-fulfillment, as when he stresses the importance of intimate relationships if we are to fulfill ourselves (Taylor, 1991, p.34). On other occasions he refers to capacity-fulfillment, as when he equates self-fulfillment with personal development,

or realizing one's potential (ibid., p.75). That Taylor should use "self-fulfillment" alternatively in both of these senses without clearly distinguishing them suggests that he views both modes of fulfillment as going together. Yet as Gewirth remarks (Gewirth, 1998, p.16), each of these modes may be present without the other. Consider for instance a virtuoso pianist who, we may assume, would never have achieved the impressive level of skill he had now reached without the tyrannical discipline his father had inflicted on him when he was younger. This person may have achieved capacity-fulfillment without also enjoying aspiration-fulfillment, if his aspirations were for something else than becoming a world-class pianist – perhaps he strongly wishes he had been able to live a more ordinary but also more balanced life. Conversely, someone with modest expectations about himself and life might feel fulfilled without having realized any of his best capacities to the full, i.e. he might have aspiration-fulfillment without capacity-fulfillment.

Yet even once we have distinguished these two meanings of "self-fulfillment", isn't it still the case that living an authentic life will always entail achieving at least *one* of these two modes of fulfillment? No matter how pleasant such an idea might sound, it does not seem to be the case. It is certainly plausible to think that both aspiration- and capacity-fulfillment might naturally *tend* to accompany an authentic life. After all, the fulfillment of one's deepest desires or of one's best capacities are certainly major ways in which one can be true to oneself in a way that will warrant praise. Many examples of authentic individuals embody self-fulfillment in at least one of those two senses. Romantic artists like Percy Bysshe Shelley seem to have achieved both modes of it.

Nevertheless, the connection between authenticity and self-fulfillment in either of these two senses does not seem a necessary one. It is arguably possible to live an authentic life without achieving either aspiration- or capacity-fulfillment. Suppose that Opportunist, untrue to his name, decided that his integrity as an artist was more important than maximizing his career prospects, and consequently refused to sell out. As a result, he loses S.'s support and cannot find anyone receptive to his work as a composer. He is eventually forced to spend the rest of his life supporting himself through a tedious clerical job that does not leave him with enough time for composition. It would then be very difficult to say that Opportunist's frustrating life had been one of self-fulfillment in any of Gewirth's two senses. Yet isn't it nevertheless a heroically authentic life? It is certainly a life of admirable integrity, although, sadly, it doesn't involve the realization of his artistic potential and personal goals. This suggests that some degree of luck is necessary if authenticity is to lead to self-fulfillment in at least one of these two senses.

Hopefully, the preceding sections have managed to throw some light on the concept of authenticity, in the senses most relevant to my four main cases, and to make at least a plausible case for the verdict of inauthenticity in those cases. In the next and final part of this dissertation, I will apply my analysis of authenticity – as well as those of Frankfurt and DeGrazia – to the contemporary debate on human enhancement, with a view to assessing the validity of the concerns about authenticity often raised in the context of that debate. In order to allow for a sufficient depth of analysis, I shall focus on one particular subset of "cosmetic neurology", familiar to the readers of Kramer's book *Listening to Prozac*: the technological enhancement of mood and personality.

4 AUTHENTICITY AND "COSMETIC NEUROLOGY": THE CASE OF MOOD AND PERSONALITY ENHANCEMENT

4.1 Introduction

Every year, millions of people are being treated with antidepressants for a mood disorder, mostly depressive disorder. Such a procedure is typically not considered ethically problematic, even from the specific perspective of authenticity. It is normally assumed that depressive disorder interferes with a person's functioning so as to obscure who they really are, turning them into a mere shadow of themselves. Accordingly, antidepressant use is taken to *restore* the person, i.e. to bring back the true, authentic self that had been masked by depression. As I have mentioned previously, this assumption doesn't seem valid in all cases involving "depression" understood in a general way independently of its particular causes, which means that even antidepressant use for therapeutic purposes may not always be unproblematic for the perspective of authenticity. While I shall have a few words to say on this in this chapter, my focus will nevertheless be on cases of enhancement strictly speaking – particularly in the domains of mood and personality. The authenticity objection to enhancement claims that the use of antidepressants and other interventions into the brain for such cosmetic purposes threatens our authenticity, and that this justifies abstaining from it.

As mentioned previously, Kramer reports such positive transformative effects from the cosmetic use of Prozac that some of his patients demanded to stay on the drug even after their symptoms had resolved. Some of these effects were, on his

account, radical changes in personality. "[W]ith Prozac", he writes, "I had seen patient after patient become...'better than well'. Prozac seemed to give social confidence to the habitually timid, to make the sensitive brash, to lend the introvert the social skills of a salesman" (Kramer, 1994, p.xv). Prozac, he contends, has the capacity to induce the psychological disposition he calls "hyperthymia", the opposite of dysthymia. Hyperthymics tend to experience a predominance of positive affect in their lives, and Kramer describes them as "optimistic, decisive, quick of thought, charismatic, energetic, and confident" (ibid., p.17). He also states that Prozac successfully raised the self-esteem of some patients who had previously tried, without success, to solve their problem through psychotherapy. If Kramer is right, we can say that antidepressants like Prozac have, among other things, the power to increase extroversion and decrease neuroticism even in some "healthy" subjects. Since extroversion is correlated with positive affect and neuroticism with negative affect, it shouldn't be surprising that Prozac can also elevate mood in such people if it can indeed boost the former personality trait and reduce the latter. Though, as we have said, Kramer's claims about the transformative effects of Prozac should be taken with a grain of salt, it nevertheless seems plausible that as our understanding of the underpinnings of mood and personality in the brain grows, we will be able to perfect existing methods of intervention into the brain, up to the point where we will become able to produce transformations at least as impressive as those described by Kramer, with a much greater degree of control.[62] In the present state of affairs, only a minority of Prozac users are supposedly "transformed" by the drug,[63] and among those who are, the effects are not always predictable. Kramer thus mentions the example of a

[62] It should be noted that it doesn't matter for our purposes whether the impact of antidepressants is, as has recently been argued, chiefly due to a placebo effect, rather than to the chemical ingredients contained in the drugs. All that matters is whether the drugs *do* have a significant transformative effect, no matter how it is produced.

[63] As Kramer himself admits: see Kramer, 1994, p.11.

woman, Ms. B., who had been prescribed Prozac to cure her compulsive hair-pulling, yet as a result also lost her long-standing drive to find a spouse, a drive that had so far led to unsuccessful attempts. This case, Kramer writes, "illustrates an important quality of Prozac – namely, that it often surprises us". [64]

The personality changes that people in contemporary Western culture would be likely to seek through cosmetic neurology would precisely be those Kramer credits Prozac with: lower neuroticism and higher extroversion, the traits our culture tends to reward. Higher conscientiousness might also be in demand, to the extent that it typically helps achieve life goals and enhanced work performance. Agreeableness is a trickier case: while people might be drawn to the prospect of making themselves more likeable, the desire for financial success might pull them the other way, as recent research has found that agreeable people, particularly in the case of men, earn less money than their less agreeable peers. The authors suggest that this is largely due to a social norm expecting men to be aggressive and demanding (Judge et al., 2012). Except for low agreeableness, these are also the traits that some ethicists have argued we have good *reasons* to seek out. Transhumanist James Hughes, for instance, argues for the desirability of using enhancement technologies to make ourselves happier. He cites a variety of studies suggesting that "[h]appier people are more likely to get and stay married, have more friends, belong to more groups, and are more likely to volunteer. Happier people are more highly rated by their supervisors and they make more money. Happier people are also healthier and live longer" (Hughes, 2009).[65] He then proceeds to cite a study suggesting a correlation between personality on the one hand, and a host of desirable life outcomes (including better mood) on the other, the

[64] Ibid., pp.265-7.
[65] For the evidence cited by Hughes, see e.g. Lyubomirsky et al., 2005.

gist of his message being that it would be highly desirable to become high in extroversion, agreeableness and conscientiousness, and low in neuroticism.[66] In a similar vein, Mark Walker has argued that we have a moral duty to create and make available "happy people pills", pharmaceuticals that would induce hyperthymia (Walker, 2011). Julian Savulescu also suggests that characteristics like optimism, and having a sunny temperament, are of a kind that makes a person's life go better regardless of what her life projects happen to be, making them prime targets to achieve through the use of enhancements.[67] Finally, it has also been claimed that we may have reasons to directly enhance our *moral* capacities through cosmetic neurology. Examples that have been given include the reduction of a propensity to violent aggression, or of a racist bias (Douglas, 2008). The science of moral enhancement is still in its infancy and the techniques that may allow such manipulations are for the most part speculative, though some suggestive studies have recently appeared, e.g. about the possibility of reducing racial bias by using the beta-blocker propranolol, typically used to treat heart disease (Terbeck et al., 2012).

Do such proposals raise legitimate concerns about the authenticity of our lives? I will now consider this issue in the light of the main accounts of authenticity we have discussed in parts 2 and 3.

4.2 The possibility of involuntary change

While the idea of enhancing mood and personality is likely to appeal to many, this appeal arguably depends on the assumption that neuroenhancers can effect the

[66] The study in question is Ozer and Benet-Martinez, 2006.
[67] In Savulescu et al., 2011a, p.11.

specific sort of changes that the individual desires. Yet as we have mentioned, current enhancers do not allow us to control the relevant changes with much precision. One can of course increase or diminish the dosage of a particular medication, yet this amounts to choosing how much of a particular change one wants to experience. One cannot determine in advance the particular nature of that change, or pick and choose between specific effects, e.g. enhancing self-confidence while leaving baseline mood or sensitivity unchanged. Sometimes, the changes induced by Prozac weren't foreseen at all, as in the case of Ms. B. Though she seemed to have eventually welcomed the disappearance of her drive to find a husband, things might have been different. Suppose for instance that she hadn't been struggling with compulsive hair-pulling, yet had requested to go on Prozac with the hope of experiencing the same sort of transformation as that of Tess, whom Kramer describes as having experienced a dramatic improvement in her social skills, energy level, and sexual appeal to the point that she had three dates a week-end. Had Ms. B. had such expectations, Prozac would apparently not have fulfilled them. In such a scenario, the views of Frankfurt and DeGrazia would still imply that Ms. B.'s choice to take Prozac to change her personality had been authentic, to the extent that it was both honest and autonomous (in the relevant sense of autonomy), yet insofar as she wouldn't identify with her new personality, these authors would presumably regard that personality as inauthentic. Also, it is clear that Ms. B. wouldn't have authentically chosen to acquire *that* particular personality, since she didn't foresee that she would acquire it – on the contrary, she expected a different result. Similar concerns about unforeseen personality changes have been mentioned by bioethicists in relation to other forms of cosmetic neurology such as brain stimulation (Focquaert and DeRidder, 2009).

What if the subject actually turns out to endorse the resulting changes (as the real Ms. B. seems to have done) even though they were not those she had initially anticipated? From the perspective of the self-creation model, it seems that concerns about authenticity should then become irrelevant. If the person wholeheartedly identifies with her new traits, and would still do so after critical reflection, then these traits must be authentic on Frankfurt and DeGrazia's views. Trickier cases, however, would be those in which the person's ulterior identification with her new traits *itself* resulted from the enhancement procedure. Imagine a shy person whose main life project is to promote charitable giving to end world poverty. Believing that becoming more confident and articulate would make her more effective in her campaigning, she starts taking Prozac under the supervision of a psychiatrist. As a result, she does become more confident and charismatic, and starts to excel at public speaking, yet she also experiences what Kramer admits having observed in some of the patients he put on Prozac, namely a certain "numbing of moral sensibility" (Kramer, 1994, p.291). She now concludes that she has been excessively self-sacrificing in the past, and that she should start putting herself first. Accordingly, she embarks on a new career path and finds a well-paid job as a spokesperson for a big tobacco company. We can assume that her former, non-medicated self would be horrified by her new mindset, yet conversely, while on Prozac she regards her former altruistic commitments as immature and excessive.[68] Suppose also that she doesn't wish to go off medication anymore, as she is afraid of collapsing back into her shy and altruistic self, which she no longer regards as her "true" one. It seems that this person could in principle

[68] Such a scenario may be said to be implausible. After all, it might be argued, some of Kramer's patients weren't happy with the changes in their personality induced by the drug, and on that account were glad to stop taking it. However, one of Kramer's main points in *Listening to Prozac* is that different people respond differently to such a drug. Given also what he says about the loss of moral seriousness in some of his patients, it doesn't seem inconceivable that present or future neuroenhancers might have an impact on the user's values – and this possibility seems worth mentioning, as many people would probably find it ethically disturbing.

wholeheartedly endorse her new personality on Prozac. Nothing in the Frankfurtian analysis of authenticity appears to stipulate that this endorsement ought to fit with her former higher-order attitudes for it to be wholehearted. Of course, its stability might be called into question. But if the person keeps taking the drug for long enough, this response becomes unavailable. DeGrazia considers a similar case (though involving permanent change through a "one-off" procedure, a process of brainwashing) and concludes, though tentatively, that the person's higher-order attitudes *post*-transformation ought to have priority from the perspective of authenticity. Such a position might strike some of us as unsatisfactory, for reasons which I believe my analysis of authenticity can explain, as we shall see in a later section.

The concern about unforeseen changes, it should be noted, would disappear if new enhancers were to arrive giving us a greater degree of control over the process of change than current ones do. The main practical implication of this issue, if it is authenticity in Frankfurt and DeGrazia's sense that concerns us, seems to be that were a medical professional to prescribe e.g. an antidepressant to a patient for such enhancement purposes, she would have a duty to make it clear that the results desired by the patient were not guaranteed – the drug might fail to produce any changes in her mood or personality, or it might produce changes she might not like. It would then be up to the patient to make an informed decision as to whether or not to try out the procedure. The unpredictability of current neuroenhancers is one of their main drawbacks, yet provided that they don't involve a risk of dangerous side effects, this unpredictability doesn't in itself seem to warrant prohibiting their use. If someone had made an informed choice to use such an enhancer, and its use had been shown to be safe enough, it would seem to disregard that person's autonomy to forbid her to use it.

4.3 Threats to autonomy

The concern raised in the previous section is partly a concern about the autonomy of people's choice to use personality and mood enhancers. Another autonomy-related concern arises from the existence of social preferences for certain personality traits over others. We have mentioned the premium enjoyed by extroversion and low neuroticism in contemporary Western society. Kramer repeatedly emphasizes that our culture rewards the hyperthymic personality, characterized by high spirits, confidence, assertiveness, and a "thick skin". When new ways become available to modify personality in a safe and affordable manner, people who don't fit the contemporary personality ideal may find themselves under pressure to use neuroenhancers to better comply with it, e.g. in a professional context. We might for instance imagine restaurant employees or flight attendants being expected to take mood brighteners if this allowed them to provide a friendlier service to customers. Not everyone, however, may wish to have their mood or personality modified in such a manner, even if they knew the enhancement procedure to be reasonably safe. If someone were to reluctantly agree to use such enhancers in order to avoid the social or professional costs of a refusal, it may be argued that this person's choice wasn't autonomous, and therefore that it was inauthentic on both Frankfurt and DeGrazia's view.

Should we be concerned about such possible impingements on people's autonomy, and advocate, if not a prohibition on the use of personality and mood enhancers (which itself would constitute another form of autonomy curtailment), at

least the establishment of legal safeguards protecting those who didn't wish to use such enhancers from discrimination? Those who aren't worried by this issue might point out that we already accept some kinds of social pressures as legitimate in the professional context. For instance, someone who categorically refused for some reason to use computers would thereby disqualify himself from being considered for a variety of occupations in our society, from administrative to academic ones. Yet most of us think that the social pressures attached to such occupations are acceptable, given the benefits provided by computer use in such contexts and the need to coordinate the activities of different workers. To this one might counter that the pressure to use neuroenhancers would require us to tinker with more intimate aspects of our identity than the pressure to use computers, and that this is an ethically significant difference. While there does seem to be something to this reply, the issue is nevertheless a tricky one. A University education is a necessary requirement for a variety of high-paying jobs. Does such an education really shape a person's identity to a lesser degree than mood and personality enhancers would tend to do? It isn't clear to me that it must, especially considering that future neuroenhancers might allow us to modulate the duration of their effects quite precisely. While the effects of a drug like Prozac take at least a few days to wane once the drug has been discontinued, we can imagine a future personality enhancer of a shorter-acting sort, closer to a drug like Ritalin. It might be possible for someone to take such an enhancer in the morning to be able to meet the demands for cheerfulness at work, and to then return to his true, graver self in the evening and on week-ends.[69] Also, as Kramer points out, the social norms that would underlie the pressures to use neuroenhancers already exist, and they currently

[69] The risk of undesirable side effects is, of course, another potential reason why pressures to use enhancers might be seen as more problematic than the pressures to use computers or to get a University education. The availability of sufficiently safe enhancers, however, would alleviate that concern.

put at a disadvantage those who, in virtue of their natural constitution, fail to meet them (Kramer, 1994, pp.274-5). Such people might find that the availability of mood and personality enhancers promotes rather than impairs their autonomy, to the extent that it would broaden their social and professional opportunities, making them better able to realize their aspirations in life. Whether enhancement use actually promotes their autonomy would further depend, of course, on the question whether their aspirations were themselves consistent with autonomy.

Still, depending on the exact form that such social pressures might take, they might well justify some degree of concern. The problem might not just be their general nature as social *pressures* placing some restrictions on people's autonomy, but more specifically, their *prejudiced* and arbitrary nature. I shall elaborate on this issue in a later section.

4.4 Are enhanced traits necessarily "fake" or not really "ours"?

4.4.1 Concerns about stability and depth

A common concern raised by the prospect of mood and personality enhancement is that the new features produced by the relevant enhancers would be "fake", or not really our own in some other sense. Remember Elliott's suggestion that changing personality with an antidepressant appears fundamentally inauthentic. In section 2.7.3, however, we have found no plausible way of justifying the claim that Ex-gay's newly acquired heterosexual orientation was fake, or not really "his". Is such a claim more plausible with regards to the improved mood or new personality that neuroenhancers might produce? I don't see that it is. Arguably, neuroenhancers can in

principle genuinely *change* who we fundamentally are, and there is no reason in principle why the new features they might induce shouldn't be able to meet my seven conditions for inclusion into the true self. The fact that such technological interventions may be less familiar to us than psychotherapy, for instance, is irrelevant in this regard.

Yet isn't the fakeness charge still appropriate in relation to at least *some* types of neuroenhancers? As mentioned earlier in relation to the sadist's case, one might for instance stress the temporary nature of the changes produced by drugs like Prozac. Many of Kramer's patients simply reverted back to their original personality style once he took them off medication, which is what led Tess, among others, to declare that she was "not herself" without Prozac, and to ask to get back on it. Therefore, it might be argued, the new personality traits produced by the drug were too short-lived to count as really "hers". This criterion, however, cannot declare inauthentic the features induced through "one-off" procedures like neurosurgery, for instance, since their effects are permanent. It cannot even declare *all* features produced via Prozac and similar drugs inauthentic, since it is in principle possible for someone to keep taking such a medication for the rest of her life.

It might perhaps be argued here that we should still differentiate between procedures with permanent effects like neurosurgery, and agents like antidepressants, which require repeated intake to perpetuate their effects. Because the Prozac user would revert back to her original personality and mood were she to discontinue the drug, the argument goes, her new features do not go *deep* enough to have really become part of her true self, no matter how long she keeps taking the drug. While

seeing the intuitive appeal of this line of argument, I nevertheless find it problematic, insofar as we are assuming that the disposition this person has to revert to her former features without the effect of Prozac is never activated. And it isn't clear why we should give priority to what are now merely potential features over her actual ones when it comes to defining who she really is. We do not, for instance, declare inauthentic the exceptional skill of a chess player or athlete, or the improved mood enjoyed by someone engaging in daily physical exercise, on the grounds that this person wouldn't be able to maintain her level of performance or mood were she to abandon her strict practice regimen, or at least relax it. It might perhaps be retorted that while physical exercise has been shown to lift mood, it cannot produce the sort of personality makeovers that Kramer describes or that future enhancers might allow. Even if that is correct, however, I don't think this justifies calling the Prozac-induced traits fake. Suppose then that physical exercise *did* lead to similar personality makeovers in some people. Surely their physically active self wouldn't be fake, despite being much sunnier and confident than the non-active one. We may feel somewhat puzzled if this person significantly changed, almost in a Jekyll and Hyde manner, when she interrupted her exercise routine, but neither of her two different "selves" would be any less real than the other – it would simply be manifested under different circumstances.

The objector might still insist that my view has unacceptable implications. Imagine a healthy person who has nevertheless been on Prozac since childhood. The drug has given this person a confident and resilient personality, yet let us assume that if he stopped taking it, he would become much more vulnerable and sensitive. On my view, this person's true self is the one on Prozac. If he stopped taking the drug, he

would no longer be himself – just as some of Kramer's patients felt. Even more, given what I have said about Ex-gay in section 3.4, this person may have a strong reason (perhaps even a duty grounded in self-respect) not to discontinue the drug, provided that doing so would mean forfeiting valuable traits. Being authentic in his case would then entail continuing to express his "Prozac self" and sticking to the medication. But this, it might be argued, is implausible: surely we ought to say that if this person stopped taking Prozac, his true self would finally be revealed, and that it involved sensitivity and vulnerability. Though I don't have strong views on this issue, I am prepared to accept these implications of my view. Imagine an intellectual who stopped reading anything – whether books, newspapers or other texts – altogether, and following Henry David Thoreau's example, went to live in a cabin in the woods. Contrary to Thoreau, however, this person makes this a permanent lifestyle and never returns to civilized society and intellectual activities. As time goes by, his interests and preoccupations, and probably some aspects of his personality, would change quite significantly. Yet while this person might well learn new things about himself following this radical move, I see no reason to think that his original interests and traits must have been somehow phony. The intuition behind the objection we are considering seems to be that a person's true self can only be revealed under "normal" life conditions, and that conditions that involve the regular use of Prozac are not normal ones. If this notion of normal conditions could be cashed out in a plausible way, it would vindicate that view. I shall remain agnostic about that possibility, but I am rather sceptical of its prospects of success. The conditions in question cannot simply be defined as the minimum ones required for the healthy functioning of the individual, or else the features we have acquired at least partly through the influence of civilization and enculturation will all have to be declared inauthentic, since it isn't

clear that such processes are fundamentally required for healthy functioning.[70] Also, the relevant conditions must not simply be those that don't involve the use of enhancers like Prozac, or it would become unclear what normative weight we should attach to such conditions. Such a solution would seem to entail a form of pharmaceutical exceptionalism: personality traits produced by such enhancers are inauthentic, because they are produced by such enhancers. This isn't an explanation.

I do share the uneasiness felt by many at the prospect of shaping a child's psychological features early on using enhancers. We might sense that doing so would interfere with the unfolding of the child's "nature", and that this is undesirable (Brock, 1998, p.62). My view might be able to make sense of this, to the extent that even young children already possess certain valuable potentialities which, we might think, it is best to realize, rather than to replace with other, pharmacologically-induced ones. (The idea that the relevant potentialities are *valuable* seems important: if what the enhancement eliminated was, for instance, a disposition to racial prejudice, it is less clear that we should find the intervention problematic, safety concerns aside.) If enhancement use were to take place so early in the child's development that no such potentialities had had the time to form, we may still have some concerns e.g. about the nature of the parents' desire to shape their child, but worries about authenticity would no longer be relevant – indeed, the enhancement would then be *shaping* the child's nature, not thwarting or replacing it in any sense.

[70] We could of course hold an account of healthy functioning that did make such processes indispensable. But then why couldn't we just as well stipulate that Prozac use was necessary for such functioning? This sort of solution would make the authenticity of Prozac-induced features relative to a particular social consensus. A similar difficulty would arise if we appealed to "typical", rather than healthy, functioning: Prozac use might in principle become the norm in a certain society.

4.4.2 The "natural/artificial" contrast

Another common reason why people tend to assume that cosmetic neurology procedures must produce fake results, one related to yet in principle distinct from considerations of stability or depth, may have to do with the distinction between "natural" and "artificial" means of improvement. Weight training or regular practice at chess or tennis, for instance, are often held to constitute natural ways of improving oneself. The outcomes they produce might be judged authentic on that account. By contrast, the use of neuroenhancers, like cosmetic surgery or steroids in sports, are typically considered artificial forms of improvement, which might explain why many people are inclined to regard the resulting traits or performances as phony.[71] But no matter how commonly held such a view might be, it appears highly questionable. Indeed, the sense in which it uses the terms "natural" and "artificial" is not the standard one, and seems rather arbitrary. By standard sense, I mean that given for instance by the *Oxford English Dictionary*, which defines "artificial" as "made or constructed by human skill, esp. in imitation of, or as a substitute for, something which is made or occurs naturally" (Oxford English Dictionary, 2011). The natural, by contrast, is that which occurs in nature without having been purposefully shaped a certain way by human beings. On this standard understanding of the terms, the way the distinction between natural and artificial means of improvement is usually drawn appears inadequate. Practices like education or training regimens for athletes then clearly count as artificial, yet they are ordinarily regarded as ethically unproblematic and even virtuous, and their effects can be at least as far-reaching as those permitted by technological interventions. It is tempting to suspect that the common way of drawing the line here actually involves labelling "natural" the means of improvement

[71] As mentioned e.g. by DeGrazia, 2005a, p.263.

with which we are familiar and comfortable, and "artificial" those that are less familiar to us and about which we consequently feel more uneasy. But if this is indeed the way in which the distinction is commonly drawn, then it doesn't seem able to play the normative role it is supposed to play. The fact that some procedure might be unfamiliar to us, and make us somewhat uneasy, isn't enough to show that it is ethically problematic or, more specifically, that the features it generates must be somehow fake.

We have seen in section 2.7.2 that we do value "naturalness", in the sense of spontaneity, as opposed to affectation. Yet the risk of creating an "unnatural" personality in this sense actually seems *reduced* rather than increased by enhancement technologies, compared to more traditional methods of self-shaping, to the extent that these technologies can produce changes that go deeper, freeing for instance the person from the need to "act the part". Hughes thus cites the example of flight attendants who are pressured to "remain constantly cheerful and bright" while on the job (Hughes, 2009). Their awareness of the discrepancy between their actual mood and outward demeanor can lead such people to experience strong feelings of alienation. The use of neuroenhancers, Hughes points out, could allow them to make their inner feelings congruent with their behaviour and professional requirements, thereby eliminating both their sense of alienation and the need for contrived cheerfulness – possibly resulting in a greater subjective sense of personal authenticity while at work.[72]

[72] Could they however feel strongly alienated from their daytime feelings once they came home in the evenings, and the effects of the neuroenhancer had vanished? This would require empirical verification. It isn't clear, for instance, that ADHD patients on Ritalin who discontinue the medication in their private lives to be their authentic selves necessarily experience a disturbing feeling of alienation as a result.

It may well be that some of those committed to authenticity as "naturalness" actually value *the very fact* of not deliberately shaping their mood or personality, allowing instead their spontaneous psychological dispositions to unfold freely. While it does sometimes seem appropriate to value our natural traits, it isn't clear to me, in the case of features like personality or mood, that we ever have reason to value *their being natural itself*, rather than other, independent properties these features have. Imagine someone who consistently lived by such a commitment to naturalness. This person would presumably never make any effort to manage his negative emotions out of respect for politeness and etiquette, expressing his anger, frustration or bad mood with the freedom of a baby. As a personal ideal, such a wanton way of life doesn't seem to have much to recommend itself.

Finally, perhaps what worries people like Elliott isn't that enhanced traits must fail to be really "ours" in the sense of being fake. Rather, the concern is that they wouldn't be *uniquely* ours. The personality I would acquire on Prozac, it might be thought, would be a "designer personality" that would essentially depend on the pill, and not on the peculiarities of my individual constitution. As applied to Prozac, such a worry seems unwarranted: as we have seen, its enhancing effects, when present, are not fully predictable, and are certainly affected by the individual differences between users. However, as we acquire greater control over our psychological features, it may be that we will eventually become able to shape our personalities exactly as we wish, no matter what our individual characteristics were to begin with. And if the influence of social norms proves powerful enough, we can imagine many people opting for highly similar personalities, in accordance with those norms. While such a prospect does sound disquieting, it seems that even a personality that had been entirely shaped

via enhancement technologies could still be "uniquely ours": one could decide to use those technologies to craft a really unique personality, reflecting one's own personal creativity.

4.5 Direct vs. indirect interventions into the brain

Another way of spelling out the worry that neuroenhancers must necessarily produce fake traits would be to appeal to an idea we have mentioned in section 2.7.3 in relation to the sadist's case: namely, neuroenhancers bypass their users' rational capacities, when those capacities ought to be solicited. This idea might also be phrased in terms of a contrast between direct and indirect interventions into the brain. Neil Levy thus characterizes indirect interventions as involving changing people's psychological features by first influencing their beliefs, via "the presentation of evidence and argument" (Levy, 2007, p.69).[73] Practices like education or life coaching would typically be placed into this category. Perhaps this is the reason why they are also often regarded as "natural" despite involving deliberate human intervention. By contrast, cosmetic neurology procedures change people's traits by *directly* affecting their brains, without any prior impact on their beliefs – or if they do have such an impact, direct interventions nevertheless do not change the person's beliefs by providing her with evidence and argument of any sort. Rather, the relevant beliefs then get modified simply in response to a change in the person's brain processes, e.g. in the way certain neurotransmitters are being reabsorbed into the synapses.

[73] To be specific, Levy himself speaks of direct interventions into the *mind* rather than the brain. This difference of terms doesn't matter for our purposes, as the interventions we are discussing can only intervene into people's minds by affecting their brain functioning as well.

Suppose that Kramer's patient Tess, rather than taking Prozac, chose instead to work with a personal coach who helped her improve her social skills and self-presentation. As a result, she eventually makes new, emotionally healthier friends and manages to turn her romantic life around. Having found a loyal partner and getting much more social reinforcement than before, she now feels fulfilled, loved and supported by others, and as a result becomes just as confident and emotionally resilient as she would have been on Prozac (but without coaching). In this scenario, Tess's new level of confidence would result from the new beliefs she had come to hold about herself and other people's attitude to her, beliefs supported by evidence: for instance, her memory of having performed better recently in social settings, and of receiving demonstrations of kindness and love from her partner. By contrast, the confidence Tess acquires through Prozac might be said to come as it were "out of the blue". Even if we assume that the drug produces its effects by first influencing Tess's beliefs about her own worth, it remains that taking Prozac does not, initially at least, provide her with any evidence that would warrant holding more favourable beliefs about herself. It may *eventually* yield such evidence, insofar as, according to Kramer's account, Tess does start to perform better socially once on Prozac. But the initial change might be considered inauthentic to the extent that it results from a direct intervention into Tess's brain and mind.

While I do believe that this line of argument gestures towards an important concern about cosmetic neurology, the claim that direct interventions into the brain for enhancement purposes must always produce inauthentic traits is too strong. It implies for instance that if the person who regularly engages in vigorous exercise

thereby enjoys a better mood and focus than she would otherwise, these enhanced features must be inauthentic, insofar as physical exercise produces its results by directly affecting the brain processes controlling mood and focus, rather than by providing the person with any argument or evidence. Even procedures like education do not change people's psychology solely through the presentation of evidence. Moral education sometimes involves telling the child that certain kinds of action are simply wrong, full stop, and should not be performed. Arguments in support of such claims cannot always be given to the child, either because she does not yet have the capacity to grasp them, or because the moral claims in question are "basic", not based on further arguments. Yet this does not seem to warrant the conclusion that the moral capacities that the child goes on to develop as a result of such education are inauthentic. Our psychological features don't always need to come about as a response to evidence or argument in order to count as authentic.

The appeal to the distinction between direct and indirect interventions into the brain has more force when applied to the particular psychological features of ours that do presuppose such a response. Before discussing it further, however, we need to clarify the key idea on which this suggestion is grounded, already hinted at in section 2.7.2: the idea that we have reasons to feel a certain way about ourselves, or about certain states of affairs in the world, and to behave in accordance with those feelings.

4.6 The idea of reasons to feel and react

I take it to be a plausible claim that many of our emotions, moods, and at least many of the appraisals we make of others and ourselves, such as self-esteem, admit of

reasons; and that these reasons make the relevant emotions, moods, and appraisals, and the actions they elicit, appropriate or inappropriate, reasonable or unreasonable. Philosophers have thus variously spoken of "appropriate emotions",[74] "affective reasons", or reasons to feel (Kahane, 2011). A complication here is that a certain emotion might sometimes be appropriate, or fitting, in circumstances C, even though what we have most reason to feel in such circumstances, all things considered, is not that emotion but a different one. For instance, it might be imprudent to feel amusement at a joke one has just recollected, because, say, one can reasonably expect this to lead to an uncontrollable bout of laughter, when laughter would be perceived as offensive in the circumstances – say, at a funeral. The joke in question might nevertheless be genuinely amusing, yet all things considered, one has most reason not to feel amused by it.[75] In such a case, we might say that one still has an *intrinsic* reason to feel amused, i.e. the fact that amusement is a fitting response to the joke, but that it is outweighed by other, *instrumental* reasons not to feel amused, i.e. the negative consequences that being amused would produce (Kahane, 2011, p.168). For the sake of simplicity, though, I will confine myself here to cases where the fitting emotion and what one has most reason to feel do not come apart.

Here are examples of what I have in mind. If someone you loved has just died, this gives you a reason to feel grief and to respond to this situation by mourning. If good fortune has befallen you, e.g. if you have won a large sum at the lottery, this gives you a reason to feel joy. Consider also the case of self-esteem. Following Rebecca Roache, we can concisely define self-esteem as a global estimation of one's own worth (Roache, 2007, p.74). This estimation is sometimes assumed to be a

[74] See e.g. Mulligan, 1998, though the idea goes back at least to Aristotle.
[75] As pointed out e.g. by D'Arms and Jacobson, 2000.

positive one, as when people speak of a person's lack of self-esteem, meaning not that this person fails to engage in self-assessment, but rather that his own opinion of himself is a negative one. Sometimes the term is used in a more neutral sense, as in the phrase "low self-esteem", which again indicates a rather negative assessment of one's overall worth. There also seem to be reasons for self-esteem, and these are provided by the qualities and achievements of the agent that *warrant* a favorable appraisal of her own worth. Possessing certain positive qualities, such as tenacity, or having accomplished something significant, such as writing a series of beautiful poems, are reasons for high self-esteem – even though these might co-exist with reasons for low self-esteem, such as a lack of tact, or having forgotten a friend's precious manuscript on a train. In such a case, the balance of those various reasons will determine the level of self-esteem that it is reasonable for someone to have. It may also be that the fundamental property of being a person in itself warrants a certain degree of self-esteem. People can respond to their reasons for self-esteem more or less successfully: e.g. a young person might have great athletic abilities that would give her a reason to have quite a high self-esteem, yet she might fail to respond to it if her parents had impressed upon her the idea that athletic achievements are despicable.

In a recent paper on reasons to feel, Guy Kahane draws a helpful distinction between *responding* and simply *conforming* to our affective reasons (Kahane, 2011, p.171). Imagine a very cynical person who steals a suitcase from his rich grandfather, believing it to contain a large sum of money, and having camouflaged the act to make it look like the result of a burglary. As he drives home in his car, the cynic feels elated at the thought of all the money he has just come into possession of. However, what he

doesn't know is that the suitcase he has stolen is in fact empty – he took the wrong one. Yet by quite an extraordinary coincidence, the cynic's grandfather, whose capacity for critical judgment has unfortunately deteriorated in recent years, has just decided to offer that exact sum of money as a birthday present to his grandson, whom he mistakenly believes to be very meritorious. The cynic hasn't yet discovered the news of the empty suitcase, or of his birthday present either. He now has a good affective reason to feel cheerful, and he actually does feel that way. (Of course, he also has a reason to feel grateful towards his grandfather and guilty for what he did, sentiments quite alien to him.) He may therefore be said to be *conforming* to his reasons to feel in those circumstances. However, he is not *responding* to those reasons, as his cheerfulness wasn't caused by an awareness of the birthday present, the fact that gives him a reason to feel cheerful. Rather, it was caused by a false belief that he had successfully stolen money from his grandfather, a belief which, if true, wouldn't warrant feeling what he is actually feeling.

Finally, we may ask how *precise* the requirements of our affective reasons are. The answer seems to be that this depends on the case. In some cases, only one particular way of responding emotionally to the relevant situation seems appropriate: losing a beloved one seems to warrant feeling significant grief. Feeling only mild sadness, or even indifference, always seems inappropriate – it suggests that one didn't really care about the deceased, which itself may constitute a failure to appropriately respond to one's affective reasons if the person in question really was loveable. On the other hand, there also seem to be other cases that call for a degree of pluralism about what is to count as an appropriate affective response. For instance, it seems plausible to think that there is a *range* of different levels of self-esteem that it is

appropriate for someone to have given her particular set of qualities and accomplishments, rather than *one* specific such level. Suppose Lene experiences a significant boost to her self-esteem on winning the chess tournament organized in her local area. On the other hand, Nina, who won the same tournament the year before, experienced a more modest boost after that particular achievement, and felt something comparable to what Lene now feels only once she became a grandmaster. Nina happens to have more demanding criteria for what counts as a great achievement at chess than Lene. It might be that the latter's more easy-going attitude is more conducive to happiness, yet it is at least not obvious that either of the two is at fault in her response to her victory in the local tournament. Neither of them, let us assume, believes that she deserves knighthood on account of her achievement, or conversely, that this achievement is of no significance whatsoever. There seem to be two views we could take here. Perhaps winning the local chess tournament does warrant a rise in self-esteem of a precise degree, yet it is difficult to exactly identify that degree, on account of which the best we can say is that both of their responses are reasonable, *for all we know*. Alternatively, the win may not warrant such a determinate response, but rather a boost in self-esteem that falls within a certain range – all responses within that range counting as reasonable or appropriate. How we settle this matter is not crucial for my purposes. My main point is that in some cases, it seems we will have to regard a certain (limited) range of affective responses as appropriate or reasonable in light of the person's circumstances – whether this is due to our epistemic limitations, or to the nature of things.

4.7 Direct vs. indirect interventions into the brain, continued

Let us now come back to the suggestion that the use of personality or mood enhancers, as it constitutes a direct intervention into the brain, must necessarily produce inauthentic psychological features. The idea could be that direct interventions into the brain, by their very nature, cannot allow people like Tess to genuinely *respond* to their reasons to feel better about themselves, even when they do have such reasons. The best such interventions can do is make people *conform* to those reasons. This view appears questionable, and stands in need of justification.[76] It seems legitimate to worry that such interventions *might* sometimes achieve their results without yielding a genuine response to affective reasons, but it is another thing to claim that they *must* do so. After all, such a response itself depends on the proper functioning of our brain for its occurrence. There is, at least in principle, no reason why a direct intervention shouldn't be able to improve the functioning of a person's brain so as to allow her to experience genuine insight into her own affective reasons, when her brain wasn't initially so disposed. Or to put it differently, even though such an intervention doesn't change the person's psychology through the presentation of evidence and argument, it might nevertheless help her better *appreciate*, or respond to, the evidence she already has. To take an analogy, recent research has found that it is possible to improve people's mathematical ability by applying Transcranial Direct Current Stimulation (tDCS) to a particular part of their brain, the right parietal cortex (Cohen Kadosh et al., 2010). It would be inappropriate to maintain that the subjects in that study didn't "really" learn to perform mathematical calculations more effectively. They did not as it were pull out the results of those calculations out of nowhere,

[76] David Pugmire, for instance, endorses a view of this kind in relation to the authenticity of certain emotions, like fondness: according to him, fondness produced by a pill wouldn't be real fondness, because it wouldn't originate – as it should – in "appraisal", that is, an evaluation of its object as meriting the relevant emotion response (Pugmire, 1994, p.111). However, he does not provide an argument for his assumption that the relevant appraisal couldn't be generated by pharmacological means.

without needing to engage in mathematical reasoning, after their brain had been stimulated. Rather, the stimulation simply improved their ability to engage in such reasoning.

In response to this, the objector might refine her claim. While conceding that direct interventions can in principle allow someone to genuinely respond to her affective reasons when she initially wasn't able to do so, the objector might nevertheless insist that, in light of the available evidence, this isn't what happened with people like Kramer's Tess. We may remember that, according to Kramer's own account, Tess asked him to get back on Prozac even once her depression had been relieved, as she felt that the inner strength the drug had conferred on her was slipping away. Kramer describes even more extreme cases of patients who, when on Prozac, had healthy self-esteem, yet were overwhelmed again with feelings of inferiority as soon as they discontinued the drug (Kramer, 1994, pp.202-7). While Prozac did cause such patients to hold more positive self-beliefs, the volatile nature of those beliefs might suggest that they did not result from genuine *insight* about how these people really had reason to feel about themselves – and because of this, that their new beliefs were not authentic. Indeed, we usually assume that when genuine insight gives rise to new beliefs about the self, these beliefs will remain stable (provided that the person is not given any reason to change them again, and does not lose her capacity to respond to the relevant affective reasons). If the positive self-beliefs of Kramer's patients had remained stable once they discontinued the drug, it would be appropriate to conclude that Prozac had indeed given them genuine insight into their affective reasons. But since this wasn't the case, we can rule out that possibility, and conclude that their improved self-esteem was "fake".

This more nuanced line of argument still faces the challenge of explaining how we can rule out the possibility that these patients may have enjoyed the ability to respond to their affective reasons only while they were on the drug. Perhaps we naturally tend to assume that people are fundamentally rational beings who necessarily possess a sound enough, even if imperfect, capacity to respond to their affective reasons. But is that assumption always warranted? Robin Dillon coined the term "basal self-respect" to refer to "a more basic sense of worth that enables an individual to develop the intellectually more sophisticated forms, a precondition for being able to take one's qualities or the fact that one is a person as grounds of positive self-worth". Basal self-valuing, she adds, "is our most fundamental sense of ourselves as mattering and our primordial interpretation of self and self-worth" (Dillon, 2010). Kramer appears to have the same notion in mind when he talks about the "visceral sense of self-worth" that Prozac positively affects (Kramer, 1994, p.208). Couldn't it be the case that the basal self-respect of someone like Tess had been damaged by her painful life story, and that Prozac temporarily repaired it, making her capable again of properly appreciating the evidence she had of her own worth? It doesn't seem that we can rule out such a possibility.

True, we shouldn't ignore the risk that neuroenhancement use might sometimes disrupt a person's capacity to respond to her affective reasons. Perhaps some of Kramer's patients acquired a high self-esteem or positive mood that was inappropriate to their circumstances. Or even if it was appropriate, the drug may still have simply caused them to conform to their affective reasons without truly

responding to them.[77] This would be the case if the drug were to raise the user's mood and self-esteem regardless of her personal circumstances, achievements, and behaviour. To assess the soundness of this worry, we would need to know more about the mode of functioning of particular neuroenhancers like Prozac. Kramer, however, does lend some support to that worry when he says that Prozac induces "hyperthymia" in formerly vulnerable, unassuming patients. He cites the words used by psychiatrist Hagop Akiskal to describe hyperthymics, which among some positive epithets also include less flattering ones such as "overoptimistic", "overconfident" or "boastful" (Kramer, 1994, p.165).

In cases where neuroenhancement use yields features that don't constitute a genuine response to the person's affective reasons, does the unresponsiveness of those features itself warrant calling them fake? A positive answer to that question would commit us to a highly normative understanding of the features in question, entailing for instance that authentic self-esteem must be *reasonable*. I personally find it preferable to distinguish the idea of authenticity from the idea of reasonableness. If someone, because of an unlucky genetic endowment or bad upbringing (or both), had only managed to develop a seriously defective capacity to respond to her reasons to feel, this wouldn't seem to me enough to call her resulting affective responses inauthentic. They would simply be inappropriate to her circumstances. I do agree that it seems intuitively plausible to say, in cases where neuroenhancement use has *compromised* someone's responsiveness to her affective reasons, that the resulting features are inauthentic – but not in the sense that they are *fake*. A key assumption in

[77] Though as Kahane points out, this might still count as an improvement if the person wasn't initially able to even conform to her affective reasons: second best might still be better than nothing. See Kahane, 2011, p.171.

such cases is that *she was originally more reasonable* than she ends up being after the enhancement. Maybe she was initially more responsive to her affective reasons; or even if her affective response was partly based on false beliefs, and therefore didn't accord with her *actual* reasons to feel, this response may still have been more reasonable, *in light of those beliefs*, than her eventual one. Under such assumptions, the person's initial or original self was more valuable, in one respect at least, than her new one on medication. This takes us back to the remarks we made about Ex-gay's new sexual orientation in section 3.4: by calling the person's new self "inauthentic", we emphasize that it is unlike her original (i.e. "authentic") one and that we find this problematic or at least regrettable. This may be because her new self is less reasonable than her previous one and is therefore, in that particular respect, less valuable. Or it may be because, while her new self entails affective responses that are just as reasonable as those yielded by her former self (remember the pluralistic view defended in the previous section), this new self nevertheless isn't valuable *in quite the same way* as her former one. If Cohen is right, we may e.g. legitimately lament the loss of our friend's modest disposition, even while acknowledging that the greater assertiveness and higher self-esteem she has acquired on Prozac don't exceed the limits of reasonableness. Ultimately, these considerations are what the concern about inauthentic traits seems to be really getting at. There is no reason to think that the moods and personality traits yielded by neuroenhancers must be any less real or stable than our unenhanced ones, or that they cannot go deep enough.

The above considerations are important, and can sometimes justify regarding the technological manipulation of someone's "authentic", i.e. original personality or mood disposition as undesirable. Nevertheless, in light of what we have said before,

we must also acknowledge the possibility that neuroenhancement use might sometimes *help* people better respond to their affective reasons. In such cases, it seems more difficult to argue for the desirability of preserving the person's "authentic" self. It is true that we sometimes like others partly for their defects. Yet we shouldn't commit ourselves to the implausible view that all forms of self-improvement (including improvements in our affective rationality) are morally problematic, or even something to be regretted. One might also stress the possibility that there might be good reasons for someone's defective responsiveness to her affective reasons: for instance, a pupil being regularly bullied by her classmates may have difficulty appreciating the merits, during a school trip to the theatre, of a play that which she would actually enjoy were she to see it in more propitious circumstances. In such a case, it seems that the right way to improve the person's responsiveness to her affective reasons is to change her circumstances by stopping the bullying, not to give her a mood brightener to lift her spirits enough so as to allow her to enjoy the play.[78] Yet not all cases of defective responsiveness to affective reasons are of that kind.

The idea of reasons to feel and react also seems to play a role in two further, common concerns about the use of mood and personality enhancers. The first one is that their use to make ourselves happy will only lead to an inauthentic happiness. The second one has to do with the idea of *moral* enhancement by technological means, the possibility of which might again be called into question on grounds of authenticity.

[78] We can perhaps imagine variants in which stopping the bullying *and* giving the pupil a mood brightener might be acceptable: suppose that despite the improvement in her circumstances the pupil's mood couldn't return fast enough to a level that would allow her to fully enjoy the play, unless she took the enhancer. This might of course depend on the exact nature and effects of the enhancer in question.

4.8 Inauthentic happiness?

Can authentic happiness be found in a pill, or some other technological variant of mood enhancement? Many would say that it cannot. The President's Council on Bioethics appears to lean in that direction when, discussing the use of "mood brighteners" like Prozac, they write for instance that

> While such drugs often make things better—they often help individuals achieve some measure of the happiness they desire—taking such drugs may also leave many of the same individuals wondering whether their newfound happiness is fully *their own*—and in this sense, fully real. (The President's Council on Bioethics (U.S.), 2003, p.255)

Assessing this worry requires getting clear about what the tricky concept of "happiness" refers to exactly. As Dan Haybron explains, happiness is sometimes understood simply as a *psychological state* of a certain kind, and sometimes as synonymous with *well-being* (in which case it seems to correspond to the Greek term *eudaimonia*).[79] If it is the latter sense that the worry about inauthentic happiness relies on, then it is clear that mood enhancers *on their own* may not be able to produce authentic happiness. On some forms of hedonism, they could in principle do so. Yet the concept of well-being is obviously a hotly contested issue, and there are many other conceptions (such as the Aristotelian one, which places virtuous activity at the centre of well-being) according to which the regular use of mood enhancers will never be sufficient to lead to authentic happiness in that sense. That said, we should note that several philosophical accounts of well-being make at least *some* room for pleasurable feelings as a constituent. To the extent that mood enhancers can increase the occurrence of such feelings, it seems that they could at least make a positive

[79] Haybron, 2011.

contribution to someone's well-being on those accounts. However, on conceptions of well-being that stress the value of a realistic outlook on things, the use of mood enhancers might actually be antithetical to happiness if it led the person to see the world through rose-tinted glasses.

Given the complexity of the philosophical debate about the nature of well-being, as well as the fact that the prospective users of mood and personality enhancers are more likely to be interested in happiness in the "psychological state" sense, it might be preferable to focus on that sense of the term. Unsurprisingly, there is more than one account of happiness in that sense as well, but the two most famous examples are probably hedonism, according to which being happy means having attained a sufficient balance of pleasant over unpleasant experience, and life satisfaction theories, according to which happiness consists in having a favorable attitude toward one's life as a whole (Haybron, 2011). It would seem that mood and personality enhancers, even in their current forms, can at least help people achieve real or authentic happiness in either of these two senses. To claim that they cannot seems to commit one to a more substantive view of happiness than the two just mentioned.[80]

The President's Council's contention that the newfound happiness of those who resort to cosmetic neurology might not be "fully their own" appears somewhat clearer in light of the lines that follow it. This concern, they say, "is even more pertinent, and more disquieting, should one come to feel happy for no good reason at

[80] In the spirit of the self-creation model of authenticity, Mark Walker views a person's happiness (understood as a propensity to positive moods) as authentic if it is in line with her own conception of the person she wants to be (Walker, 2011, p.138). Clearly, neuroenhancers can in principle produce authentic happiness in this sense too.

all, or happy even when there remains much in one's life to be truly unhappy about" (The President's Council on Bioethics (U.S.), 2003, p.255). This suggests that their real concern here is the possibility that pharmacologically-induced happiness might involve a disconnection from the person's affective reasons. They could still understand happiness as a psychological state in one of the two senses above, yet maintain that when such a state fails to be properly responsive to the relevant affective reasons, it is inauthentic. Given what I have said about features like self-esteem in the previous section, I believe that if happiness can appropriately be equated with either pleasant experiences or life satisfaction, it is rather misleading to identify "reasonable happiness" with "authentic happiness". Indeed, the latter phrase suggests happiness that is fully *real*, and surely it is e.g. possible to be genuinely satisfied with one's life even if one's affective reasons don't warrant such satisfaction. The alternative for the Council is to understand happiness in a more robust manner, for example, as *reasonable* satisfaction with life. Something like such a view also seems to underlie Elliott's claim that putting his imaginary accountant on Prozac to rid him of his feelings of sadness and alienation is to compromise the authenticity of his life.

One problem with Elliott's view is that we may not be convinced by his somewhat extreme claim that our inability to find a transcendent framework of meaning in the world anymore – even if this inability is taken for granted – justifies feeling alienated and depressed, and that only "morons" can feel differently. If true, this claim would seem to rule out the possibility of authentic happiness altogether! Furthermore, such a robust conception of authentic happiness as reasonable life satisfaction (or reasonable predominance of positive experience) seems to move away

from the idea of being "true to oneself" to the distinct idea of being true to reality, or to the facts (including facts about our affective reasons). Because of this, it seems to me preferable to speak of "misguided" rather than inauthentic happiness in the sort of cases that Elliott and the Council have in mind. After all, as we have mentioned, some people may be naturally defective in their capacity to respond to their affective reasons, and may fail as a result to respond to their reasons to occasionally feel bad. On the view we are now considering, these people will be incapable of authentic happiness, even if they show a predominance of positive affect and are fully satisfied with their lives, simply because of the natural deficiencies in their affective rationality. As I have mentioned earlier, I would find it more plausible to regard the newfound happiness of Elliott's accountant inauthentic on the assumption that Prozac allowed him to feel content despite having his most fundamental needs and desires frustrated. These would, after all, represent aspects of his true self, which may have found expression in his feelings of sadness and alienation, but which the drug might repress. Even then, however, it may be preferable, for purposes of conceptual clarity, to distinguish happiness from *self-fulfillment* (in Gewirth's "aspiration-fulfillment" sense), and grant that the accountant may be genuinely happy, but deny that he would be *fulfilled*.

The risk that cosmetic neurology might facilitate the achievement of happiness "on the cheap", and at the cost of a disconnection from one's affective reasons, shouldn't be ignored, even if we nevertheless choose to call such happiness authentic. It is the same worry that Robert Nozick had already highlighted with his famous example of the experience machine (Nozick, 1974, pp.42-5). However, this risk still doesn't justify a categorical rejection of mood enhancers even if *reasonable* happiness

is what we are after. Kramer, for instance, argues that Prozac can move people from one normal state to another normal one, only better rewarded socially. If my pluralist assumption about reasons to feel is correct, giving someone a sunnier disposition through neuroenhancement need not disrupt her capacity to respond to her affective reasons. Rather, doing so would then lead her to trade one reasonable affective response to life for another, equally reasonable one. We could of course judge that we then still had a reason to retain the former response on Cohenian grounds – but this would mean moving away from the issue of reasonableness. Furthermore, we have seen before that while neuroenhancers may sometimes be detrimental to our affective rationality, it doesn't seem necessary that they should be – on the contrary, they could in principle enhance our responsiveness to our affective reasons in some circumstances. We should of course be wary before concluding that such a result had been achieved. But if it could be, it seems that mood enhancers could usefully complement more traditional procedures like psychotherapy or cognitive therapy. While such procedures work with our existing rational and affective capacities with the aim of harnessing their potential, mood enhancers could in principle allow us to improve those capacities, in cases where such improvement were desirable. Arguing for the preservation of our "authentic" personality or mood level in such cases would require showing that the original deficiency in our affective rationality was in fact valuable, enough so to be worth preserving – either because it represented a "likeable defect", or because the feature that underlied that deficiency was valuable in other ways, and couldn't retain that value if the deficiency were corrected. I will not try to pursue such a line of argument myself.

4.9 Moral enhancement

In section 2.7.3, we mentioned the suggestion that if the sadist were to take the "morality pill", his new, more civilized personality would be fake, though preferable to his original one. One of the justifications offered for this claim was that the pill wouldn't improve the sadist's behaviour by providing him with genuine moral insight. While this might still be better than the status quo, the fakeness of the sadist's new personality would still make it a sub-optimal outcome. It might be argued that this is what justifies preferring "traditional" methods of moral enhancement, such as education or anger management programs, to technological ones, when the former can be expected to produce similar effects (which admittedly might not be the case for the sadist). Such techniques, one might hold, can truly enable the person to actually respond to his moral reasons, whereas neuroenhancers can at best make him *conform* to these.

What we have said in the previous sections shows the limits of such a line of reasoning. Just as we said earlier in relation to our responsiveness to our affective reasons more generally, there seems to be no reason *in principle* why neuroenhancement use may not at least help someone achieve genuine moral insight, given that the capacity for such insight must itself be somehow grounded in the brain, even though we don't know yet exactly how. True, there have been long-standing disagreements between different philosophical schools about how exactly it is that we grasp moral truths, and about what counts as a genuinely moral *motive*. Moral rationalists, such as Kantians, hold that moral competence and truly moral motivation are primarily a matter of using our rational capacities, rather than being moved by feelings or emotions. Sentimentalists, by contrast, view feelings and emotions of a

certain kind, such as empathy or guilt, as a source of appropriate moral motives, and as the basic foundation of our moral knowledge. Yet which of these schools happens to be right doesn't affect our conclusion. Both our rational capacities and the feelings that sentimentalists regard as crucial to morality are fundamentally dependent on the functioning of our brain. If neuroenhancers could appropriately influence that functioning, they could make us better moral judges or agents on either of these competing views. There is for instance some evidence that the hormone and neurotransmitter oxytocin, which occurs naturally but whose levels can be artificially elevated e.g. by administering it through a nasal spray, can increase moral emotions like compassion and empathy (Rockliff et al., 2011). Accordingly, some ethicists see it as a promising moral enhancer (see e.g. Persson and Savulescu, 2012, pp.118ff). If someone with a naturally defective capacity to empathize with others, such as the sadist in my earlier example (or a similar person who just felt short of a diagnosis of personality disorder), could come to better appreciate his moral duties to others thanks to oxytocin, this would seem to count as a fully authentic moral enhancement.[81]

Besides positively increasing our capacity to respond to our moral reasons, neuroenhancers might also facilitate the workings of an already existing capacity by removing the obstacles to its exercise. Recent research in psychology suggests for instance that a great number of people who sincerely hold anti-racist views nevertheless manifest an unconscious racial prejudice (see e.g. Greenwald et al., 2009). Such people don't lack moral insight: they fully appreciate the fact that someone's belonging to a different ethnicity is no reason for treating this person less

[81] By "naturally defective capacity" for empathy, I mean to rule out e.g. cases in which this defect was merely a response to very harsh personal circumstances that would spontaneously resolve itself were these circumstances to improve (e.g. if the person in question got the love and support he had so far been deprived of).

well. Yet it seems that the deep-seated unconscious prejudice many of us share still persists despite such insight. As Tom Douglas argues, if neuroenhancers allowed us to remove such prejudice where it is present, to the extent that it leads otherwise decent people to harbour unjust attitudes towards people of other ethnic groups, the use of such neuroenhancers would plausibly count as an authentic, i.e. genuine moral enhancement (Douglas, 2008, p.231). Choosing to undergo such enhancement could also count as a fully authentic form of self-transformation on my definition of authenticity, assuming it expressed a praiseworthy commitment to racial equality.

That said, due once again to the limitations of current enhancers, the concern about authentic moral insight may have some force. While oxytocin has been shown to increase trust, empathy and other pro-social emotions, this seems to particularly apply to the members of one's *own group*. This bias can sometimes reflect itself in antisocial behaviour towards out-group members, when such behaviour promotes in-group interests (Persson and Savulescu, 2012, p.120). Such reinforcement of in-group bias might well count as a form of moral corruption, rather than enhancement. Furthermore, even if we agree that emotions like empathy have an important role to play in moral behaviour, both sentimentalists and rationalists will agree that there is more to sound moral judgment than a tendency to feel empathy. The impact of a puff of oxytocin on a person's moral disposition will therefore depend on the particulars of the case. It could in principle genuinely improve the capacity for moral insight of someone like the sadist – but it might also not. Perhaps it would make him *nicer* without making him genuinely *wiser* and more moral, if it simply made him more docile and benevolent, but rather thoughtlessly so, without leading him to understand the wrongness of his past behaviour and the nature of his duties to others. Also, even

in cases where someone's levels of empathy and trust are initially defective, it could be that administering oxytocin would move the person from defect to excess in this regard, rendering him gullible and overly "nice". While trust and empathy are appropriate in many circumstances, they are not so in others, e.g. when dealing with a con artist or with a serial killer who pulls the classic trick of trying to lure his victim into a van by pretending to be injured. Although even current enhancers might sometimes be able to genuinely improve our moral capacities, it is important not to overlook the complex nature of moral life, and the potential negative effects of such interventions if they are used in the wrong circumstances, or if they aren't supplemented with others like moral education.

Given that moral insight does fundamentally involve responsiveness to (moral) reasons, viewing reasonableness as necessary for authenticity seems more plausible in relation to authentic moral insight, or practical wisdom, than it was e.g. in the context of happiness. That said, the worry we have already encountered, according to which the problem with neuroenhancers might not be that they fail to yield "real" new features but rather that these new features involve a loss of value from the original, "authentic" ones, again seems applicable. If Oscar Wilde could have taken a "kindness pill" that would have made his personality more similar to that of the Dalai Lama, he may have become a morally better person, yet at the expense of a loss in non-moral value, e.g. in terms of his scathing wit or of his devotion to his artistic projects. Clearly, though, such a self-transformation couldn't be described as problematic on *moral* grounds. The question whether moral enhancement could paradoxically involve a loss of specifically *moral* value is a more difficult one: if we accept the thesis of the unity of the virtues, which implies that it is never necessary to

sacrifice one moral virtue for the sake of another, then moral enhancement could never threaten existing moral value. On the other hand, one will have to grant that it can if one thinks that some gain in one particular moral virtue, say kindness, can sometimes come at a cost to some other moral virtue, e.g. self-respect. I won't try to address this complex issue here.

4.10 Social prejudice, self-respect, and changing valuable traits

Our discussion of mood and personality enhancers so far suggests that the main problem brought out by the concerns about authenticity raised in that context has to do with the undesirability of changing *valuable* aspects of our present identity, valuable in the sense mentioned earlier in relation to Ex-gay's self-creation project: that is to say, the relevant features are intrinsically valuable, or at least their appropriate expression in our life *could* have such value. These considerations, I believe, help explain the discomfort we may still feel about someone like the altruist turned go-getter on Prozac whom we encountered in section 4.2, even once we have put worries about autonomy to the side: the high moral worth of this person's original character has been lost in the change. Whether or not we think that talk of authenticity is the best language to use to phrase such concerns, I believe that it has the merit of highlighting certain ethical issues that have so far been neglected in the debate on the ethics of enhancement. Authors who have written on authenticity in that context typically focus, like DeGrazia, on the concept of autonomy. While the importance of that concept cannot be denied, this approach nevertheless misses the fact that concerns about autonomy do not exhaust concerns about authenticity. The prospective user of an enhancement technology can still legitimately ask herself: even though I

could autonomously choose to enhance myself in this way, and even though doing so would promote some of my life goals, ought I nevertheless to do it? Or would I thereby be lose an important part of myself?

The data cited by Hughes about personality and life outcomes are striking and cannot be ignored. It is, however, important not to draw hasty conclusions from them. First, showing a correlation between a personality trait and some life outcome isn't yet to show that the former *causes* the latter. If, for instance, the correlation between extroversion and longer life were partly due to the influence of a gene underlying both features, turning someone into an extrovert by stimulating his brain might not yield the full benefits we had expected; for all we know, an intervention at the genetic level might prove necessary. Secondly, the picture of the "ideal" personality painted by Hughes overlooks some of the complexities of the findings in this area. For instance, while several studies do seem to have found a general correlation between happiness and life success, other recent studies, focusing on more specific contexts, have e.g. reported a *negative* association between positive affect and objective college outcomes (Nickerson et al., 2011), and a positive correlation between neuroticism and academic achievement for a particular group of students, those high in "superego strength" – roughly the sense that one has the capacity to successfully confront the challenges one is facing (McKenzie et al., 2000). Our knowledge of the impact of personality traits on life outcomes is still in the making, and more research is needed to understand how these various data fit together. Yet currently available findings suggest that some features that appear disadvantageous from a very broad perspective can actually constitute strengths in specific contexts.

Thirdly, and most importantly for my purposes, even if we assume that people typically have quite strong reasons to use enhancement technologies to make themselves happier, more extroverted and conscientious, or less neurotic, concerns about authenticity highlight the fact that they may also have *other* kinds of reasons that speak against changing their existing personality. Even if some of our personality traits happen to be disadvantageous in some contexts, their appropriate expression may nevertheless have a particular kind of intrinsic value that justifies preserving them, e.g. for the sort of reasons mentioned by Cohen in his defense of the priority of existing value. Let us return to neuroticism, the trait that seems most difficult to defend. If one were only to read Hughes and the research he cites in psychology, one would have to conclude that neuroticism is an unalloyed evil: the less of it one has, the better off one is – at least as long as one doesn't go so low on that dimension as to become dysfunctional, e.g. completely immune to feelings like shame or guilt. Yet while neuroticism may well play an overwhelmingly negative role above a certain limit, there is reason to think that the life of those who aren't the lowest "healthy" scorers can include intrinsic goods that would otherwise not be available to them. Such people might for instance be more keenly aware of the tragic aspects of human existence, making them deeper thinkers. They may also be able to express this awareness in intrinsically valuable creations. It is doubtful that a hyperthymic could have penned such great dark works as Shakespeare's *Macbeth*, Shelley's poem "Ozymandias", or the classic plays of the "Theatre of the Absurd" by Samuel Beckett or Eugène Ionesco; or the thoughtful meditations of Schopenhauer on the cruelty of the natural world; or Tchaikovsky's *Pathétique* symphony. Anti-neuroticism campaigners such as Hugues neglect these considerations entirely. Kramer, to his credit, does recognize them, going so far as to state that "[m]uch of the insight and

creative achievement of the human race is due to the discontent, guilt, and critical eye of dysthymics" (Kramer, 1994, p.165).

This brings us to another possible virtue of neuroticism: the fact that the sense of dissatisfaction it typically entails can provide an incentive to challenge the status quo and improve things, in a way that a very low degree of that trait wouldn't. Hughes denies this. He thus writes that "chemical patience does not imply any less ability to recognize and redress bad situations. Even the normally happiest or most agreeable or extroverted people can recognize and resist illegitimate authority" (Hughes, 2009). Ed Diener and colleagues, however, have found that the "normally happiest" of all people achieve less than slightly less happy people in terms of income, education, and political participation, even though they are more successful in terms of close relationships and volunteer work (Oishi et al., 2007). The authors suggest as an explanation that the happiest people may be too satisfied with their lives to strive for achievement in domains like education and work, thereby being less likely to fulfill their potential, and too satisfied with existing political conditions to want to actively participate in the political process. I have of course no wish to deny the value of close relationships and volunteering work. All I want to say is that, according to the available data (which confirm to some extent what we might have expected), bringing people to the lowest "healthy" point on the neuroticism dimension will involve trading some intrinsic goods for others, and that there doesn't seem to be a personality profile that is "ideal" from any reasonable perspective, contrary to what some of the pro-enhancement literature suggests.

Acknowledging the particular value that a non-hyperthymic personality may have

doesn't commit us to the claim that it must be superior, *overall*, to a hyperthymic one, but only that it *can* be so in at least *some* respects.[82] Some people might simply be hampered by their neuroticism, and be unable to derive anything good from it, in which case they will have no reason to refrain from changing that trait through technological means. Others will be able to express their neuroticism in valuable ways, in which case they will at least have a reason to remain "authentic" in this regard, i.e. not to change this aspect of their personality. This reason might still sometimes be outweighed by their reasons to change, but it will not always be. This will depend on their particular life circumstances, goals and values, and on the exact nature of their personality. While in some cases their neuroticism will make it more difficult for them to achieve some of their life goals, the research we have cited suggests that there will also be cases in which it might actually prove an *asset*. And even when it closes off some potential benefits, the particular intrinsic value attached to their existing personality might still make it reasonable to preserve it as it is. It might also give them a reason to prefer, when possible, ways of securing the various goods mentioned by Hughes, such as health and longevity, that don't involve tinkering with their personality. I believe the same line of argument could be run, with equal plausibility, in relation to other currently "unpopular" traits like introversion.

Yet even if I am right that some people have good reasons to preserve their existing personality even if it isn't the most socially favoured one, is it nevertheless *morally wrong* of them to decide to change it? Not necessarily. The analysis I have offered of Ex-gay's case suggests that self-transformation projects often regarded as inauthentic are morally problematic to the extent that they involve a lack of self-

[82] Kramer fails to appreciate that point when dealing with this issue, for instance, in Kramer, 2000, p.15.

respect. But not all instances of personality or mood enhancement need involve such a moral failing. We have already mentioned the possibility of removing our unconscious racial prejudices using cosmetic neurology. We could also imagine someone, call him the *unhappy introvert*, who has always dreamt of becoming a compere in a comedy club, yet finds that no matter how hard he practices, his personality prevents him from competing with the more extroverted candidates for such jobs. While not judging introversion intrinsically inferior to extroversion, this person simply feels frustrated at not being able to realize his dream, and cannot find full compensation in the introverted activities available to him. If this person were to use a personality enhancer as the last available resort to make his dream come true, his friends may have reason to regret his transformation, yet they may not have grounds for morally criticizing his decision. (And of course, if they are good friends, they will have reason to be happy for him if he then manages to fulfill his dream.) However, the unhappy introvert might indeed be showing a failure to properly appreciate his existing personality if, for instance, he were to regard the prospect of a personality makeover too lightly, as if it were no different from trying out a new shirt, or if he thought that his introversion was, not just less than ideal for his dream job, but a despicable trait that he ought to get rid of if he didn't want to be a "loser" for the rest of his life.

The risk that some types of personality makeover might betray a lack of self-respect is increased by the fact that some of the benefits sought through such interventions appear fundamentally tied to controversial social preferences for certain personality traits. Certainly, *some* such preferences seem legitimate, especially in particular contexts. A person who, as a result of being low in agreeableness or

conscientiousness, constantly gets into confrontations with his work colleagues or is undependable, cannot legitimately complain if his professional career suffers.[83] Similarly, someone with a sombre personal style cannot really complain if he doesn't get accepted by his local samba performance group, on account of his "negative vibes". Still, it is necessary to distinguish between *legitimate* and *problematic* social norms, as was already highlighted by the case of Ex-gay. Homophobic social norms are not legitimate: if some enhancement technology could produce the same results as therapy does in Ex-gay's case, it would be wrong to expect gay people to use it to adapt to those norms – rather, the norms themselves ought to be fought through social and political action, to at least prevent them from exposing gay people to discrimination. The same holds for racist attitudes. As Maggie Little has put it in the context of a discussion of cosmetic surgery, such social norms are suspect because *unjust* (Little, 1998). Social attitudes towards certain personality traits lie in a greyer area, yet they shouldn't be excluded from critical scrutiny. Some parts of the Western world do seem to be moving towards what Elliott has described as a "tyranny of happiness", showing increasingly less tolerance for negative affect of any sort. In some cases, such a mindset is mostly driven by a well-meaning concern to alleviate people's suffering. Yet even then, the norm appears problematic when it increasingly treats *all* forms of negative affect as evils to be removed, including those that constitute a proper response to life circumstances.[84] Similarly, the quiet, contemplative disposition of introverts is often seen in a negative light, as indicating anti-sociality and a lack of assertiveness, dynamism and ability to contribute, even though these prejudices are very often unfounded (Cain, 2012). Such social attitudes

[83] Would he nevertheless have a claim on public resources to allow him to change his unchosen, maladaptive personality? Answering that question would require addressing interesting and difficult issues about moral responsibility.

[84] As noted by Elliott in the writings already quoted, and also by Wakefield and Horwitz, 2007.

may not be on a par with racist, sexist, or homophobic ones. I am for instance not aware of a history of enslavement of introverts by extroverts, or of any systematic discrimination exercised by the latter against the former. Some forms of discrimination, however, can be less visible or recognized than others. The exact status of such social preferences for certain personality traits seems a topic worthy of discussion for ethicists, given the increasing power these preferences will acquire from the arrival of more sophisticated methods of enhancement, allowing people to better conform to them. It would also be good to know more about the extent to which certain disadvantages are tied to intrinsic features of certain personality traits (presumably, low conscientiousness is by its very nature detrimental to health) rather than depending essentially on the impact of social preferences. For instance, are agreeable men being discriminated against at work? Or are they simply being hampered by inferior leadership abilities? The stance we take on such a question should obviously influence how we will view the desire for certain types of personality makeovers.

Even if it is inappropriate to talk about discrimination against introverts or non-hyperthymics, however, it remains that our society shows a degree of *bias* against those traits, a bias often internalized by those who possess them, leading them to view their personality as flawed if not pathological. This might sometimes threaten the autonomy of the choice they will make to change themselves through technology, but it need not always do so. Even when it doesn't, however, this choice might still reflect an insufficient appreciation of their existing identity, and of the intrinsic goods associated with it. This is, I believe, the key issue captured by the concerns about authenticity and personality enhancement that are grounded in the self-discovery

approach, and which accounts focused on the concept of autonomy leave out.

4.11 Can cosmetic neurology help enhance *our authenticity?*

Finally, given that I have made room for the possibility that someone's *enhanced* personality or mood might reflect her true self, it might be asked whether I shouldn't conclude that neuroenhancers can sometimes make people *more* authentic, rather than threatening their authenticity. This would concur with the view of many partisans of the self-creation model, and of some of Kramer's patients, whom we may remember felt more themselves on medication, and requested to remain on it precisely because they felt that without it they were no longer being themselves. In a recent paper, Neil Levy defends precisely that claim, arguing that pharmaceutical enhancers can be seen as enhancing our authenticity even if the notion is understood along the lines of the self-discovery model (Levy, 2011a). Levy's central argument appeals to cases, to which we have previously alluded, of so-called Gender Identity Disorder (GID). Levy notes that many of us who feel drawn to the self-discovery view of authenticity are inclined to accept the descriptions that transsexuals give of their cases: since the person was really a woman trapped in a man's body, and the surgery transformed her body into that of a female, the surgery allowed her to become who she really was (ibid., p.314). But then, Levy says, if we agree that such radically transformative interventions can help the person become who she really is in the sense given by the self-discovery model, this suggests more generally that the use of enhancement technologies – including pharmaceutical enhancements – can be seen as authenticity-promoting even on that model. Levy thus writes:

> The inner voice to which we listen, and which tells us what being human is for us, may not whisper of acceptance. Instead, its message might be that we should change, to bring inner and outer into harmony. Self-discovery might *require* change from us, and to that extent it is entirely compatible with the use of various enhancements. Just as the person suffering from Gender Identity Disorder might come to be who they *really* are by means of an intervention, so the depressed person might become who they are by means of Prozac. There is nothing in the self-discovery view of authenticity that requires us to reject psychopharmaceutical use. (Ibid., p.316)

Levy does have a point. It does seem possible to find examples of cosmetic neurology that support his claim. Consider a shy person who nevertheless has an extroverted personality and a great sense of humour (the *shy extrovert*). Suppose her shyness, rooted in negative beliefs about herself, prevents her from expressing her true personality in public: as a result of having been brought up by harsh parents who repeatedly criticized her for minor mistakes, she feels unworthy of other people's company and (mistakenly) assumes that they would find her boring and unlikeable if they got to know her. This person starts taking Prozac and soon becomes much more confident and assertive. Her self-esteem experiences a dramatic boost, so that she now feels free to express her extroverted nature, initiates conversations with strangers and publicly shares the jokes she used to keep to herself. As a result, she makes a number of new friends. It does seem plausible to say that in one respect at least, the drug has allowed this person to become who she really was, namely a sociable, fun extrovert. Also, provided we do not want to regard shyness as a disease, this case is plausibly construed as one of enhancement.

However, we should note that idea that authenticity has been enhanced in such a case is supported by the fact that *prior* to taking the drug, the shy extrovert already possesses certain behavioral dispositions that are just not activated, because of her inhibitions. The drug, we are assuming, does not give her *new* preferences and

dispositions, but only allows already existing ones to be expressed. Not all cases of cosmetic neurology will be of that kind. Consider again the unhappy introverted wishing he could become a professional compere. Suppose it is not the case that he already harbours extroverted tendencies that, for some reason, he is unable to express, or is not even aware of. He scores like an introvert on personality tests and behaves like one, is naturally quiet, thinks carefully before speaking, and enjoys socializing in smaller groups. Yet because of his dream to become a compere enjoying the applause of a comedy audience, he *wishes* he were more outgoing, gregarious, and capable of snappy repartee, and he turns to neuroenhancement to acquire those preferences and capacities. Now compare such a case with the example of Jan Morris, the author mentioned by Elliott who underwent sex reassignment surgery in the 1970s. In her memoir *Conundrum*, Morris reports the characteristic experience of people with GID: she realized in her early childhood years that she had been "born in the wrong body, and should really be a girl".[85] This idea seems crucial for explaining why we are inclined to regard sex reassignment surgery as allowing such people to "become who they really are" even if we accept the self-discovery framework. Namely, we believe that the person's gender identity, which involves a certain set of affective dispositions (e.g. feeling comfortable wearing female clothing but not male clothing), fundamentally defines who she is. Her biological features, by contrast, we take to be more shallow characteristics, perhaps only "skin deep", and less relevant for her self-definition. True, someone reporting a mismatch between her gender identity and her biological characteristics could in principle be self-deceived. He might merely *wish* he were transgender, and e.g. enjoyed wearing lipstick and female clothing, without really having such preferences. But assuming the person's self-report is accurate, we

[85] Quoted in Elliott, 2003, p. 30.

will agree that sex reassignment surgery has allowed her to become who she really is, because of the priority we give to the person's "inner" features over her physical attributes when it comes to defining her identity.

In the case of the unhappy introvert, on the other hand, we do not seem to find anything like a discrepancy between the person's self-perception and other, more superficial features of himself. I am not aware of any reported cases of "personality identity disorder" where someone appears to have a certain personality trait, e.g. introversion, in view of his test scores and behaviour, but nevertheless claims to have always felt that he was in fact an extrovert. Of course, an introvert may always have *wished* he were an extrovert, but this is a distinct idea, just as people with GID are not just people who wish they were born as a member of the opposite sex, e.g. because of the greater opportunities available to men in our society.

And if someone like the unhappy introvert were to maintain that he felt he "really" was an extrovert, it is unclear that we could take his claim at face value from the self-discovery perspective. Being an introvert just *is* having a certain set of affective and behavioral dispositions. If we assume person X has such a set of dispositions, whether X perceives himself as an introvert or an extrovert becomes irrelevant to the question of which personality trait we should attribute to him. Admittedly, if someone we had always regarded as an introvert assured us that he was in fact much more extroverted than we thought, we should not discount such a self-report simply because it surprised us. This person might have introspective access to features of himself we were not aware of, and his self-report might well be accurate:

he might e.g. be a shy extrovert. But such a case would be different, as such a person would in fact *not* be the introvert he had seemed to be.[86]

The self-discovery model, as I understand it, doesn't seem to make room for the idea that the unhappy introvert has in fact always been an extrovert, and that his "true" personality just required the ingestion of a pill to be made manifest. On that model, while someone like Jan Morris uses surgery to reveal on the outside who she *already* is deep inside, i.e. a woman, the unhappy introvert on Prozac simply acquires new psychological features he did not have before. The status of his introverted personality is more akin to Morris's female gender identity than to the male physical features she relinquished, as it defines who he fundamentally is – with the difference, of course, that the introvert isn't happy with his personality, just as a transgendered or homosexual person could in principle be unhappy about the preferences she had and might try and deny them. And if after being put on the drug, the introvert eventually reported, like some of Kramer's patients, that only now did he really feel like himself, from the self-discovery perspective we would have to regard such a statement with scepticism, for the reasons just mentioned. It would seem more plausible to interpret his claim as meaning that he had at last become the person he always wanted to be. On the self-discovery model, *subjective* authenticity isn't a guarantee of *objective* authenticity: a person can in principle experience the latter and yet be mistaken. This means that the self-reports of Kramer's patients, for instance, cannot simply be taken at face value, and do not, in the absence of further supporting evidence, justify the view that their authenticity has indeed been promoted.

[86] While I am deliberately using "pure" cases here for the sake of clarity, I do not wish to suggest that introversion and extroversion must come in such clear-cut forms. Being overall an introvert is compatible with possessing *some* extroverted features, and vice-versa. I do not see this as a problem for my argument: my remarks will apply to any introvert who takes a medication to change some of his introverted traits, as opposed to trying and express any extroverted features he might already have.

It might be replied here that even though there may indeed be such a structural difference between the unhappy introvert case and GID cases, this does not threaten the core of Levy's argument, as the introvert may still be described as modelling himself after a particular "inner voice": the voice telling him that he should do what it takes to acquire the sort of personality he needs to succeed. This is a valid point, yet it still doesn't support the idea that neuroenhancement has helped the introvert become who he really is even on the self-discovery model of authenticity. Indeed, on that model, it is not plausible to claim that just because some device helps you acquire certain traits you want to have, it thereby enhances your authenticity. Proponents of the self-discovery framework typically do not wish to say that a hypothetic drug that boosted your IQ by 80 points, or extended your lifespan by several centuries, would allow you to become who you already are. On the other hand, from the self-creation perspective, such a case of personality change can easily be seen as promoting authenticity: neuroenhancement is allowing the unhappy introvert to shape himself in accordance with his ideal vision of himself.

While the authenticity of the unhappy introvert isn't enhanced on my view, his self-creation project might in principle count as authentic, depending on how we assess the value of the way of life he is striving for compared to the "introverted" alternatives available to him. Yet it might also not count as such. It would not, for instance, if his self-creation project was guided instead by a desire to comply with questionable social pressures. These sorts of considerations, however, are irrelevant to the issue of authenticity from the self-creation perspective. Because of this, I cannot agree with Levy when he writes that "for the purposes of assessing the authenticity of

enhancements, we do not need to settle the dispute: on either view, the use of enhancements is acceptable" (Levy, 2011a, p.317). He is right that the authenticity objection to enhancement, in its sweeping form, should be rejected. Yet this objection doesn't exhaust all concerns about authenticity: even if we agree that some enhancement use is compatible with, even sometimes promoting of, authenticity, we may nevertheless want to know, in the case of any *specific* use of pharmaceutical enhancers, whether or not it might compromise our authenticity. And the above analysis suggests that there is more reason to worry about the potential impact of pharmaceutical enhancers on our authenticity if we accept the self-discovery model, including my version of it.

5 CONCLUSION

Enhancement technologies promise to give us an unprecedent degree of control over our personal characteristics. This is certainly an exciting prospect. We have seen some of the potential benefits that cosmetic neurology might bring, e.g. when discussing moral enhancement. While the debate on the ethics of human enhancement can easily become polarized, I hope to have shown that a nuanced view on those matters was required. The claim that all enhancement use or, more specifically for our purposes, all use of cosmetic neurology should be rejected as threatening our authenticity is untenable. However, I have also tried to show the limitations of the self-creation approach to authenticity, both in the context of the enhancement debate and more generally. This approach is undeniably helpful and does capture some important uses of the concept. It is probably the one most relevant to questions of public regulation. Indeed, by giving the notion of autonomy a central place, it

highlights both the need to protect people from coercive pressures to use enhancers, for instance in the professional context, and the need not to interfere with their autonomous choice to use enhancers to shape themselves as they wish, provided that they do not thereby harm others.

DeGrazia has also offered valid criticisms of some of the existing versions of the true self approach to authenticity. I hope to have shown, however, that this approach could still be defended. The concept of the true self can be given a firm grounding if we combine philosophical analysis with the resources of contemporary psychology. Also, the true self approach can capture important ethical concerns that the self-creation model leaves out, including in the context of the enhancement debate. It has turned out, however, that in that particular context, these concerns could be most clearly phrased by referring not just to authenticity but to other ethical notions, such as self-respect, the preservation of value embodied in our present identity, or responsiveness to our affective reasons.

Discussing the concerns raised by Kramer about Prozac, Elliott writes that "what worries him is not something as simple as Prozac making sad people happier but less interesting or less creative. What is more deeply worrying is that for at least some of the patients on Prozac, their personality changes really do seem to be for the better" (Elliott, 1998, p.177). If the analysis I have presented is correct, however, no matter how "simple" the former worry might be, it is to a large extent what the authenticity concern about cosmetic neurology is really getting at. As we have seen, there is no reason to think that the features produced through cosmetic neurology must necessarily be "fake", or otherwise inauthentic, except in the sense that they are not

our *original* features. One key insight of authenticity concerns about cosmetic neurology is that such interventions can sometimes threaten the specific kind of value realized in our current identity – in our personality, or our ability to respond to our affective reasons, even when doing so isn't always optimal from the point of view of positive affect.

Insofar as I have defined authenticity as an admirable quality, which can amount to a virtue, my own analysis seems less pertinent to the regulation of enhancement use than that of DeGrazia, for example. As citizens of liberal democracies, we don't normally assume that the aim of such regulation should be to make people virtuous. I certainly didn't mean to recommend a paternalistic imposition of my own value judgments on the rest of society. Yet that doesn't mean that discussing authenticity in that particular sense is irrelevant. I have argued that the social preferences in some parts of the Western world today are biased against certain personality traits and mood dispositions, and threaten the self-respect of those who have these features. While we cannot, nor should, legally coerce people not to entertain such biases, or forbid forms of enhancement use that we might think demonstrate a lack of self-respect, what we can do is point out such biases, make the case for the value of various personality types, and encourage people to be tolerant of individual differences and to properly value themselves even when their personal features are not those favoured by the dominant social *ethos*. (If a new pill could help people cultivate such attitudes, it might count as a form of moral enhancement!) I have made this point specifically about personality and mood, but I believe it carries over, for instance, to domains like cosmetic surgery. In countries like South Korea, we see the influence of another kind of disturbing social norms stating that looking more

Caucasian means being more attractive. If our aim is to encourage responsible enhancement use, a critical examination of such norms, as well as a discussion of the value of the identities they are affecting, seem in order.

Word count: 74’968 words (excluding bibliography and Appendix)

6 REFERENCE LIST

AGAR, N. 2010. *Humanity's End : Why We Should Reject Radical Enhancement,* Cambridge, Mass., MIT Press.

ANNAS, J. 2011. *Intelligent Virtue,* Oxford, Oxford University Press.

ARPALY, N. 2000. On Acting Rationally against One's Best Judgment. *Ethics: An International Journal of Social, Political, and Legal Philosophy,* 110 (3)**,** 488-513.

AUDI, R. 1990. Weakness of Will and Rational Action. *Australasian Journal of Philosophy,* 68**,** 270-81.

BENNETT, J. 1974. The Conscience of Huckleberry Finn. *Philosophy: The Journal of the Royal Institute of Philosophy,* 49 (188)**,** 123-34.

BENSON, P. 1991. Autonomy and Oppressive Socialization. *Social Theory and Practice,* 17 (3)**,** 385-408.

BEROFSKY, B. 1995. *Liberation from Self : a Theory of Personal Autonomy,* Cambridge, Cambridge University Press.

BLOCK, N. 1995. On a Confusion About the Function of Consciousness. *Behavioral and Brain Sciences,* 18**,** 227-47.

BOLT, I. & SCHERMER, M. 2009. Psychopharmaceutical Enhancers: Enhancing Identity? *Neuroethics,* 2**,** 103-11.

BOUCHARD, T. J. 2004. Genetic Influence on Human Psychological Traits: a Survey. *Current Directions in Psychological Science,* 13 (4)**,** 148-51.

BROCK, D. 1998. Enhancements of Human Function: Some Distinctions for Policymakers. *In:* PARENS, E. (ed.) *Enhancing Human Traits: Ethical and Social Implications.* Washington, DC: Georgetown University Press, 48-69.

BUBLITZ, J. C. & MERKEL, R. 2009. Autonomy and Authenticity of Enhanced Personality Traits. *Bioethics,* 23 (6)**,** 360-74.

BUSS, D. M. & CRAIK, K. H. 1983. The Act Frequency Approach to Personality. *Psychological Review,* 90 (2)**,** 105-26.

CAIN, S. 2012. *Quiet : the Power of Introverts in a World that Can't Stop Talking,* London, Viking.

CHOPRA, D. 2011. The Problem with Socrates. *The Huffington Post*, 29th June. Available: http://www.huffingtonpost.com/deepak-chopra/reason-enlightenment_b_883374.html [Accessed 27/03/2013].

COHEN, G. A. Unpublished. A Truth in Conservatism: Rescuing Conservatism from the Conservatives. Available: http://politicalscience.stanford.edu/sites/default/files/workshop-materials/pt_cohen.pdf [Accessed 08/04/2013].

COHEN, G. A. & OTSUKA, M. 2012. *Finding Oneself in the Other,* Princeton, N.J. ; Oxford, Princeton University Press.

COHEN KADOSH, R., LEVY, N., O'SHEA, J., SHEA, N. & SAVULESCU, J. 2012. The Neuroethics of Non-invasive Brain Stimulation. *Curr Biol,* 22 (4)**,** R108-11.

COHEN KADOSH, R., SOSKIC, S., IUCULANO, T., KANAI, R. & WALSH, V. 2010. Modulating Neuronal Activity Produces Specific and Long-Lasting Changes in Numerical Competence. *Curr Biol,* 20 (22)**,** 2016-20.

CONGREGATION FOR THE DOCTRINE OF THE FAITH. 1986. Letter to the Bishops of the Catholic Church on the Pastoral Care of Homosexual Persons. Available:

http://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_19861001_homosexual-persons_en.html [Accessed 07/04/2013].
COTTINGHAM, J. 2010. Integrity and Fragmentation. *Journal of Applied Philosophy,* 27 (1)**,** 2-14.
COX, D., LA CAZE, M. & LEVINE, M. 2012. Integrity. *In:* ZALTA, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition).* Available: http://plato.stanford.edu/archives/spr2012/entries/integrity/ [Accessed 19/09/2012].
COX, D., LA CAZE, M. & LEVINE, M. P. 2003. *Integrity and the Fragile Self,* Aldershot, Ashgate.
CRAVER, C. F. 2012. A Preliminary Case for Amnesic Selves: Toward a Clinical Moral Psychology. *Social Cognition,* 30 (4)**,** 449-73.
D'ARMS, J. & JACOBSON, D. 2000. The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research,* 61 (1)**,** 65-90.
DANIELS, N. 2000. Normal Functioning and the Treatment-Enhancement Distinction. *Cambridge Quarterly of Healthcare Ethics,* 9 (3)**,** 309-22.
DEGRAZIA, D. 2000. Prozac, Enhancement, and Self-Creation. *Hastings Center Report,* 30 (2)**,** 34-40.
DEGRAZIA, D. 2005a. Enhancement Technologies and Human Identity. *J Med Philos,* 30 (3)**,** 261-83.
DEGRAZIA, D. 2005b. *Human Identity and Bioethics,* Cambridge, Cambridge University Press.
DILLON, R. S. 1992. How to Lose Your Self-Respect. *American Philosophical Quarterly,* 29 (2)**,** 125-39.
DILLON, R. S. 2010. Respect. *In:* ZALTA, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition).* Available: http://plato.stanford.edu/archives/fall2010/entries/respect/ [Accessed 20/07/2012].
DOUGLAS, T. 2008. Moral Enhancement. *J Appl Philos,* 25**,** 228-245.
DRYDEN, J. 2010. Autonomy: Overview. *Internet Encyclopedia of Philosophy*. Available: http://www.iep.utm.edu/autonomy/ [Accessed 28/08/2012].
DWORKIN, G. 1976. Autonomy and Behavior Control. *Hastings Center Report,* 6 (1)**,** 23-8.
ELLIOTT, C. 1998. The Tyranny of Happiness: Ethics and Cosmetic Psychopharmacology. *In:* PARENS, E. (ed.) *Enhancing Human Traits: Ethical and Social Implications.* Washington, DC: Georgetown University Press, 177-86.
ELLIOTT, C. 1999. *A Philosophical Disease : Bioethics, Culture, and Identity,* New York ; London, Routledge.
ELLIOTT, C. 2003. *Better Than Well : American Medicine Meets the American Dream,* New York ; London, W.W. Norton.
FOCQUAERT, F. & DERIDDER, D. 2009. Direct Intervention in the Brain: Ethical Issues Concerning Personal Identity. *Journal of Ethics in Mental Health,* 4 (2)**,** 1-7.
FRANKFURT, H. 1988a. Freedom of the Will and the Concept of a Person. *The Importance of What We Care About.* Cambridge: Cambridge University Press, 11-25.
FRANKFURT, H. 1988b. Identification and Wholeheartedness. *The Importance of What We Care About.* Cambridge: Cambridge University Press, 159-76.

GEERTZ, C. 1973. Person, Time and Conduct in Bali. *The Interpretation of Cultures.* New York: Basic Books.
GEWIRTH, A. 1998. *Self-Fulfillment,* Princeton, N.J. ; Chichester, Princeton University Press.
GIMLIN, D. L. 2002. *Body Work : Beauty and Self-image in American Culture,* Berkeley, Calif. ; London, University of California Press.
GOWANS, C. 2006. Standing Up to Terrorists: Buddhism, Human Rights, and Self-Respect. *In:* ALLEN, D. (ed.) *Comparative Philosophy and Religion in Times of Terror.* Lanham: Lexington Books, 101-22.
GREENWALD, A. G. ET AL. 2009. Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology,* 97 (1)**,** 17-41.
HAYBRON, D. 2011. Happiness. *In:* ZALTA, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition).* Available: http://plato.stanford.edu/archives/fall2011/entries/happiness/ [Accessed 30/09/2012].
HERDER, J. 1877-1913. *Ideen*, vii.I. *In:* SUPHAN, B. (ed.) *Herders Samtliche Werke.* Vol. XIII, Berlin: Weidman.
HOPE, T., TAN, J., STEWART, A. & FITZPATRICK, R. 2011. Anorexia Nervosa and the Language of Authenticity. *Hastings Center Report,* 41 (6)**,** 19-29.
HORWITZ, A. V. & WAKEFIELD, J. C. 2007. *The Loss of Sadness : How Psychiatry Transformed Normal Sorrow into Depressive Disorder,* Oxford, Oxford University Press.
HUGHES, J. J. 2009. Group Pressures for Technological Management: What's Wrong with Society Wanting Us to Be Happy and Friendly? *Free Inquiry*, 29 (5), 28-32. Available: http://ieet.org/archive/20090725-hughes-mood management.pdf [Accessed 04/10/2012].
HURSTHOUSE, R. 2010. Virtuous Action. *In:* SANDIS, C. & O'CONNOR, T. (eds.) *A Companion to the Philosophy of Action.* New York: Wiley, 317-30.
JEFFRIES, S. 2002. The Quest for Truth. *The Guardian*, 30th Nov. Available: http://www.guardian.co.uk/books/2002/nov/30/academicexperts.highereducation [Accessed 07/04/2013].
JUDGE, T. A., LIVINGSTON, B. A. & HURST, C. 2012. Do Nice Guys – and Gals – Really Finish Last? The Joint Effects of Sex and Agreeableness on Income. *Journal of Personality and Social Psychology* 102 (2)**,** 390-407.
JUENGST, E. T. 1998. What Does Enhancement Mean? *In:* PARENS, E. (ed.) *Enhancing Human Traits: Ethical and Social Implications.* Washington, DC: Georgetown University Press, 29-47.
KAHANE, G. 2011. Reasons to Feel, Reasons to Take Pills. *In:* SAVULESCU, J., TER MEULEN, R. H. J. & KAHANE, G. (eds.) *Enhancing Human Capacities.* Oxford: Wiley-Blackwell, 166-78.
KARP, D. A. 2006. *Is It Me or my Meds? : Living with Antidepressants,* Cambridge, Mass. ; London, Harvard University Press.
KILTY, J. M. 2011. Tensions Within Identity: Notes on How Criminalized Women Negotiate Identity Through Addiction. *Aporia,* 3 (3)**,** 5-15.
KNOBE, J. 2005. Ordinary Ethical Reasoning and the Ideal of "Being Yourself"'. *Philosophical Psychology* 18 (3)**,** 327-40.
KNUTSON, B., ET AL. 1998. Selective Alteration of Personality and Social Behavior by Serotonergic Intervention. *Am J Psychiatry,* 155**,** 373-9.

KRAEMER, F. 2011. Authenticity Anyone? The Enhancement of Emotions via Neuro-Psychopharmacology. *Neuroethics,* 4 (1)**,** 51-64.
KRAMER, P. D. 1994. *Listening to Prozac,* London, Fourth Estate.
KRAMER, P. D. 2000. The Valorization of Sadness. Alienation and the Melancholic Temperament. *Hastings Cent Rep,* 30 (2)**,** 13-8.
LANGSTROM, N., RAHMAN, Q., CARLSTROM, E. & LICHTENSTEIN, P. 2010. Genetic and Environmental Effects on Same-Sex Sexual Behavior: a Population Study of Twins in Sweden. *Arch Sex Behav,* 39**,** 75-80.
LEVY, N. 2007. *Neuroethics,* Cambridge, Cambridge University Press.
LEVY, N. 2011a. Enhancing Authenticity. *Journal of Applied Philosophy,* 28 (3)**,** 308-18.
LEVY, N. 2011b. Expressing Who We Are: Moral Responsibility and Awareness of our Reasons for Action. *Analytic Philosophy,* 52 (4)**,** 243-61.
LIAO, S. M. 2011. Parental Love Pills: Some Ethical Considerations. *Bioethics,* 25 (9)**,** 489-94.
LINDEMANN, H. 2001. *Damaged Identities, Narrative Repair,* Ithaca, N.Y. ; London, Cornell University Press.
LITTLE, M. O. 1998. Cosmetic Surgery, Suspect Norms, and Complicity. *In:* PARENS, E. (ed.) *Enhancing Human Traits: Ethical and Social Implications.* Washington, DC: Georgetown University Press, 162-76.
LITTON, P. 2005. ADHD, Values, and the Self. *Am J Bioeth,* 5 (3)**,** 65-7; discussion W10-2.
LOBO, I. & SHAW, K. 2008. Phenotypic Range of Gene Expression: Environmental Influence. *Nature Education*, 1 (1). Available: http://www.nature.com/scitable/topicpage/phenotypic-range-of-gene-expression-environmental-influence-581 [Accessed 26/02/2013].
LOUGHREY, D. 1998. Second-Order Desire Accounts of Autonomy. *International Journal of Philosophical Studies,* 6 (2)**,** 211-29.
LYUBOMIRSKY, S., KING, L. & DIENER, E. 2005. The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? *Psychol Bull,* 131 (6)**,** 803-55.
MACKIE, J. L. 1977. *Ethics : Inventing Right and Wrong,* Harmondsworth, Penguin.
MAHER, B. 2008. Poll Results: Look Who's Doping. *Nature,* 452**,** 674-75.
MCCRAE, R. R. & COSTA JR, P. T. 1995. Traits Explanations in Personality Psychology. *European Journal of Personality,* 9**,** 231-52.
MCCRAE, R. R. & JOHN, O. P. 1992. An Introduction to the Five-Factor Model and its Applications. *J Pers,* 60 (2)**,** 175-215.
MCKENZIE, J., TAGHAVI-KHONSARY, M. & TINDELL, G. 2000. Neuroticism and Academic Achievement: the Furneaux Factor as a Measure of Academic Rigour. *Personality and Individual Differences,* 29**,** 3-11.
MISCHEL, W. & SHODA, Y. 1995. A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. *Psychol Rev,* 102 (2)**,** 246-68.
MULLIGAN, K. 1998. From Appropriate Emotions to Values. *The Monist,* 81 (1)**,** 161-88.
MULLIGAN, K. 2009. Was sind und was sollen die unechten Gefühle? *In:* AMREIN, U. (ed.) *Das Authentische. Zur Konstruktion von Wahrheit in der säkularen Welt.* Zürich: Chronos Verlag, 225-42.
NEWMAN, G., KNOBE, J. & BLOOM, P. Under review. The Moral Nature of the True Self.

NICKERSON, C., DIENER, E. & SCHWARTZ, N. 2011. Positive Affect and College Success. *Journal of Happiness Studies,* 12 (4)**,** 717-46.
NOZICK, R. 1974. *Anarchy, State, and Utopia,* Oxford, Basil Blackwell.
OISHI, S., DIENER, E. & LUCAS, R. E. 2007. The Optimum Level of Well-Being: Can People Be Too Happy? *Perspectives on Psychological Science,* 2 (4)**,** 346-60.
OSHANA, M. 2006. *Personal Autonomy in Society,* Aldershot, Hants, England ; Burlington, VT, Ashgate Pub. Ltd.
OXFORD ENGLISH DICTIONARY. 2011. artificial, adj. and n. *OED Online*, Sept. 2011, Oxford University Press. Available: http://www.oed.com/view/Entry/11211?redirectedFrom=artificial [Accessed 09/10/2011].
OXFORD ENGLISH DICTIONARY. 2012. authenticity, n. *OED Online*, Dec. 2012, Oxford University Press. Available: http://ezproxy.ouls.ox.ac.uk:2277/view/Entry/13325?redirectedFrom=authenticity - eid [Accessed 24/02/2013].
OZER, D. J. & BENET-MARTINEZ, V. 2006. Personality and the Prediction of Consequential Outcomes. *Annu Rev Psychol,* 57**,** 401-21.
PARENS, E. 2005. Authenticity and Ambivalence: Towards Understanding the Enhancement Debate. *Hastings Center Report,* 35 (3)**,** 34-41.
PARENS, E. 2011. True Human Enhancement: On Nicholas Agar's *Humanity's End: Why We Should Reject Radical Enhancement*. *Science Progress*, 11th March. Available: http://scienceprogress.org/2011/03/true-human-enhancement/ [Accessed 24/02/2013].
PARFIT, D. 1984. *Reasons and Persons,* Oxford, Clarendon Press.
PERSSON, I. & SAVULESCU, J. 2012. Unfit for the Future? The Need for Moral Enhancement. *Uehiro Series in Practical Ethics.* Oxford: Oxford University Press,.
PUGMIRE, D. 1994. Real Emotion. *Philosophy and Phenomenological Research,* 54 (1)**,** 105-22.
REPANTIS, D., SCHLATTMANN, P., LAISNEY, O. & HEUSER, I. 2009. Antidepressants for Neuroenhancement in Healthy Individuals: a Systematic Review. *Poiesis & Praxis,* 6**,** 139-74.
RIIS, J., SIMMONS, J. P. & GOODWIN, G. P. 2008. Preferences for Enhancement Pharmaceuticals: The Reluctance to Enhance Fundamental Traits. *Journal of Consumer Research,* 35**,** 495–508.
ROACHE, R. 2007. Should We Enhance Self-Esteem? *Philosophia,* 79**,** 71-91.
ROCKLIFF, H. ET AL. 2011. Effects of Intranasal Oxytocin on 'Compassion Focused Imagery'. *Emotion,* 11 (6)**,** 1388-96.
RORTY, A. & WONG, D. 1990. Aspects of Identity and Agency. *In:* FLANAGAN, O. & RORTY, A. O. (eds.) *Identity, Character and Morality.* Cambridge, Mass.; London: MIT Press, 19-36.
SALMELA, M. 2005. What Is Emotional Authenticity? *Journal for the Theory of Social Behavior,* 35 (3)**,** 209-30.
SANDEL, M. J. 2007. *The Case Against Perfection : Ethics in the Age of Genetic Engineering,* Cambridge, Mass., Belknap Press of Harvard University Press.
SARTRE, J.-P. 1957. *Being and Nothingness : an Essay on Phenomenological Ontology,* trans. by H. E. Barnes, London, Methuen.
SAVULESCU, J., TER MEULEN, R. H. J. & KAHANE, G. 2011a. *Enhancing Human Capacities,* Oxford, Wiley-Blackwell.

SAVULESCU, J., SANDBERG, A. & KAHANE, G. 2011b. Well-Being and Enhancement. *In:* SAVULESCU, J., TER MEULEN, R. H. J. & KAHANE, G. (eds.) *Enhancing Human Capacities.* Oxford: Wiley-Blackwell, 3-18.
SCHECHTMAN, M. 1996. *The Constitution of Selves,* Ithaca, NY, Cornell University Press.
SCHLEGEL, R. J., HICKS, J. A., ARNDT, J. & KING, L. A. 2009. Thine Own Self: True Self-Concept Accessibility and Meaning in Life. *J Pers Soc Psychol,* 96**,** 473-90.
SINGH, I. 2005. Will the 'Real Boy' Please Behave: Dosing Dilemmas for Parents of Boys with ADHD. *American Journal of Bioethics,* 5 (3)**,** 34-47.
SKEEM, J. L., POLASCHEK, D. L. L., PATRICK, C. & LILIENFELD, S. O. 2011. Psychopathic Personality: Bridging the Gap Between Scientific Evidence and Public Policy. *Psychological Science in the Public Interest,* 12 (3)**,** 95-162.
SOUTH, S. C. & KRUGER, R. F. 2008. An Interactionist Perspective on Genetic and Environmental Contributions to Personality. *Social and Personality Psychology Compass,* 2 (2)**,** 929–48.
SYNOFZIK, M. & SCHLAEPFER, T. E. 2008. Stimulating Personality: Ethical Criteria for Deep Brain Stimulation in Psychiatric Patients and for Enhancement Purposes. *Biotechnol J,* 3**,** 1511-20.
TAYLOR, C. 1989. *Sources of the Self : the Making of the Modern Identity,* Cambridge, Mass., Harvard University Press.
TAYLOR, C. 1991. *The Ethics of Authenticity,* Cambridge, Mass ; London, Harvard University Press.
TAYLOR, C. 2007. *A Secular Age,* Cambridge, Mass. ; London, Belknap Press of Harvard University Press.
TELFER, E. 1968. Self-Respect. *The Philosophical Quarterly,* 18 (71)**,** 114-21.
TERBECK, S. ET AL. 2012. Propranolol Reduces Implicit Negative Racial Bias. *Psychopharmacology (Berl),* 222 (3)**,** 419-24.
TETER, C. J., ET AL. 2006. Illicit Use of Specific Prescription Stimulants among College Students: Prevalence, Motives, and Routes of Administration. *Pharmacotherapy,* 26**,** 1501-10.
THE PRESIDENT'S COUNCIL ON BIOETHICS (U.S.) 2003. *Beyond Therapy : Biotechnology and the Pursuit of Happiness,* New York, ReganBooks.
TRILLING, L. 1972. *Sincerity And Authenticity,* London, Oxford University Press.
WALKER, M. 2011. Happy People Pills. *International Journal of Well-Being,* 1 (1)**,** 127-48.
WILLIAMS, B. A. O. 1981. *Moral Luck : Philosophical Papers, 1973-1980,* Cambridge, Cambridge University Press.
WILSON, T. D. 2002. *Strangers to Ourselves : Discovering the Adaptive Unconscious,* Cambridge, MA ; London, Belknap Press.
WOLF, S. R. 1982. Moral Saints. *The Journal of Philosophy,* 79 (8)**,** 419-39.
WOLF, S. R. 1990. *Freedom Within Reason,* New York, Oxford University Press.
ZIMBARDO, P. G. 1977. *Shyness : What It Is, What to Do About It,* Reading, Mass., Addison-Wesley Pub. Co.

7 APPENDIX: MAIN FICTIONAL CASES DISCUSSED (IN THE ORDER IN WHICH THEY APPEAR IN THE DISSERTATION)

The accountant's case. An accountant living in Downers Grove, Illinois, suddenly comes to himself one day and asks himself: "Jesus Christ, is this it? A Snapper lawn mower and a house in the suburbs?". He has come to feel deeply alienated from his way of life. He talks to his psychiatrist about his problems, who diagnoses him with depression and, at his patient's request, puts him on Prozac. The accountant soon feels much better, not alienated anymore but now at peace with himself and his run-of-the-mill life.

Akratic. Akratic is addicted to smoking. She would like to quit but has been unsuccessful so far. At a party, when someone offers her a cigarette, she yields to her desire to have one even though she does not identify with it – in fact she would prefer not to have it.

Rebel. Rebel, just like Akratic, is offered a cigarette at the party, which she accepts. Yet she is an unrepentant smoker who deems it perfectly rational to sacrifice some of her life expectancy for the sake of a more pleasurable life (and has no second-order attitudes that conflict with this outlook). Accordingly, she wholeheartedly endorses her desire to accept the cigarette.

Ex-rebel. Ex-rebel also accepted the cigarette she was offered at the party, which she attended while in her twenties. At the time, she used to adhere to a rebellious, pro-smoking attitude. This is no longer the case, however: she abandoned this mindset in her mid-thirties, as her priorities in life had gradually changed with the years.

Influenceable. At the party, Influenceable, a repentant smoker like Akratic, finds herself surrounded by friendly hedonists who encourage her to join them in their smoking and to share their *carpe diem* philosophy. Temporarily falling under their sway, she accepts the

cigarette she is being offered, and at that time feels no scruples at all about her choice – after all, you only live once, she tells herself. However, a few days later, contemplating again her decision to abandon her efforts to quit smoking, Influenceable finds that she can no longer endorse it. She now feels she was misled by the carefree spirit that prevailed at the party.

Indoctrinated. For ten years Indoctrinated has belonged to a sect that subjected its members to an intensive process of brainwashing. During these ten years he has had a constant desire to obey the demands of the sect's guru, including those of a financial nature. He has always fully identified with this desire, and this higher-order attitude did not conflict with any others he had. Indoctrinated's relatives, however, eventually manage to take him away from the sect's environment for two weeks. They push him to critically think about his situation and talk to several "outsiders". He finally concludes that the sect has manipulated him, exploiting his desire to belong to a community. He severs all links with the sect, cancels his standing order to donate money, and no longer endorses the desires on which he had been acting for ten years. He now wishes that he had never been lured into joining the sect.

Unconfident. Unconfident is studying musical composition under the supervision of Professor Arnold S., the leading figure of a well-regarded musical school. Unconfident happens to have ideas that would really break new ground in contemporary music. However, he is not sure whether he ought to include them into his compositions, as they are unlike any of the music he can find around him. Also, he finds it more challenging to create coherent pieces out of his own audacious ideas than by sticking to the canons of S.'s school. Unconfident eventually decides to be "reasonable" and to take the latter path, staying away from any "risky" endeavours. The results earn him accolades from his mentor and others, although he ends up being yet another follower of S. who doesn't stand out very much in the musical world.

Opportunist. Opportunist is an aspiring composer in all respects similar to Unconfident, with the exception that he is not intimidated at all either by the boldness of his ideas or by the prospect of working to give them coherent shape. Yet his mentor S., a rather close-minded

figure, is of a different opinion. He tells Opportunist that only by sticking to the musical style he himself champions will he be able to fully realize his artistic potential. As the time comes to submit his graduation piece, Opportunist reflects that if he complies with S.'s recommendations, his life as a musician is likely to be much easier, as he will then have the support of an eminent authority in the musical field. Even though he regards S.'s views as backward, he decides to "play it safe" and scraps everything in his work that was personal and original so as to fit his mentor's artistic outlook.

Penitent. Penitent is a 20-year old gay man. He is also the member of a religious group who regards homosexuality as an immoral perversion. Accordingly, the group preaches that gays and lesbians should "purify" themselves from their desires, allegedly imposed on them by the devil, through spiritual exercises and "reparative therapy". Penitent sincerely accepts the creeds of his religion, including this one, and loathes himself for his sexual orientation. He decides to embark on the supposed purificatory path recommended by his community. After several years of painful work on himself, Penitent eventually manages to get his homosexual inclinations completely under control. He still occasionally experiences them, but never acts on them. He enters a heterosexual marriage, and while he doesn't feel sexually fulfilled, he nevertheless develops a strong bond with his wife, who supports him in his efforts to keep his gayness at bay.

Ex-gay. Ex-gay is in all respects similar to Penitent, with the difference that he is offered the latest form of reparative therapy, more effective than any of its predecessors, which allows him to completely abolish his homosexual desires. With great relief, he can now pronounce himself "cured", and enjoy all the benefits of a purely heterosexual relationship.

The sadist, basic case. He is offered a pill that will stop from further enjoying harming innocent people, and acquire a minimum sense of empathy, in exchange for parole.

The sadist, "tyrannical prison governor" case. The sadist is pressured to take the morality pill by a tyrannical prison governor who is only interested in asserting his authority.

Juan's case. Juan is a farmer in Colombia. He is descended from a long line of farmers, yet his parents were keen to ensure that he would have reasonably broad career prospects when reaching adulthood. Juan thus spent more years in education than they did, and his good marks would have allowed him to move to the city and get training for a variety of different jobs there. However, after careful reflection, he finally chose to stick to the family tradition and managed to persuade his parents to let him come and help them on the farm instead. He has now taken over from them, and has never regretted his choice: on the contrary, being a farmer feels like a vocation to him, and he senses that he would not have enjoyed life in the city nearly as much. His existence involves a repetitive daily routine that does not leave much space for creativity of any sort, but Juan is fully happy with his life and has no desire to change it.

The unhappy introvert. This person has always dreamt of becoming a compere in a comedy club, yet finds that no matter how hard he practices, his personality prevents him from competing with the more extroverted candidates for such jobs. While not judging introversion intrinsically inferior to extroversion, he simply feels frustrated at not being able to realize his dream, and cannot find full compensation in the introverted activities available to him. This person takes a personality enhancer as the last available resort to make his dream come true (we can assume, successfully).

The shy extrovert. This person's shyness, rooted in negative beliefs about herself, prevents her from expressing her true personality in public: as a result of having been brought up by harsh parents who repeatedly criticized her for minor mistakes, she feels unworthy of other people's company and (mistakenly) assumes that they would find her boring and unlikeable if they got

to know her. She starts taking Prozac and soon becomes much more confident and assertive. Her self-esteem experiences a dramatic boost, so that she now feels free to express her extroverted nature, initiates conversations with strangers and publicly shares the jokes she used to keep to herself. As a result, she makes a number of new friends.