

*What's So Great about Reality?*¹

JULIE KIRSCH
University of Toronto
Toronto, ON
Canada M5S 1A2

Life is so wretched that it would be impossible to endure were it not for the luminous beams of illusion that guide us through its darkest moments.² Many argue, perhaps not to this degree (or with this much drama), that were it not for some modest illusions about the world and ourselves we would experience a serious decline in quality of life. On those not-so-good days, this claim strikes me as irresistible. How would I endure life and maintain my sanity if I had to embrace each and every disgusting and disheartening fact about the world? Perceiving the world in a way that is neither clear-eyed nor complete may not be without its advantages. But even if being mistaken about the world or self-deceived causes one to be healthier or more comfortable, it is worth asking whether there is something to be said against a life of this kind. Might there not be some knowledge about ourselves or the world that we *ought* to have? Do not these 'positive illusions,'³ or flattering but inaccurate beliefs that people have about themselves and the world, interfere with

1 In beginning his paper 'Reality Negotiation: From Excuses to Hope and Beyond,' C.R. Snyder explains what prompted him to question the belief that an accurate perception of reality is the *sin qua non* of mental health. In so doing, he recounts his experience of a patient who grumbled, 'What's so great about reality?' (1989, 130) This rhetorical question is the inspiration of my title.

2 I would like to thank Ronald de Sousa, Daniel Goldstick, André Gombay, and Sergio Tenenbaum for their many helpful comments on earlier drafts of this paper.

3 I am borrowing this term from Shelley Taylor (1989).

their ability to live the good life? In the first part of this paper, I will briefly review the psychological literature which suggests that mild self-deception in the positive direction contributes to, rather than compromises, a person's well-being. I will then, in the second part of the paper, examine what three thinkers, Joseph Butler, Immanuel Kant, and W.K. Clifford, have said about the wrongfulness of self-deception. While their arguments call attention to various reasons for thinking that some cases of self-deception *might* give rise on balance to negative consequences, they fail to show that all cases must have this effect. In the final part of the paper, I will consider whether our interest in autonomy might provide us with a better explanation of the supposed wrongfulness of self-deception. I will try to show that in virtue of being mistaken alone, the self-deceiver generally experiences *some* loss of a particular kind of autonomy. Nevertheless, my analysis will leave open the following two possibilities: (a) that the loss of autonomy was autonomously brought about, and (b) that the loss of autonomy in question may result on balance in an increase in the self-deceiver's autonomy.

I

Recently, a number of theorists, such as Jonathan Brown, Keith Dutton, and Shelley Taylor, have argued that mild positive self-deception promotes mental health. Shelley Taylor has challenged the long-standing view that the mark of mental health is that one has a clear perception of reality — one that corresponds with the facts. Taylor argues that not only do most people lack this quality, but that the illusions or 'robust positive biases' that people have are adaptive and promote mental health rather than interfere with it (1989, 4-7). She contends that positive illusions about one's personal qualities, degree of control, and likely future contribute to the satisfaction of the following set of widely accepted criteria for mental health:

a positive view of the self, the ability to be happy or contented, the ability to care for and about others, the capacity for productive and creative work, and the ability to grow and achieve within the context of a challenging and sometimes threatening environment. (48)

Taylor purports to have shown that those who have positive illusions about their personal qualities, degree of control, and the future, are happier than those who do not (49). Moreover, this elevated degree of happiness enables these people to make friends easily and to establish caring interpersonal relations with others (54-5). Positive illusions or an optimistic attitude, Taylor claims, also facilitates creative work, in part

by increasing one's motivation and perseverance to complete the task at hand. In illustrating this point, Taylor calls attention to a remark made by Donald Trump regarding his personal success: 'I never think of the negative. All obstacles can be overcome' (63). The suggestion here is that if you focus your attention on the positive and believe that no problem is insuperable, you will be in a better position to tackle those problems and achieve great things. Finally, Taylor argues that positive illusions generally give rise to personal growth by providing one with a guiding vision of what one wants to become — a vision that can provide one with direction, encouragement, and motivation (85).

Brown and Dutton similarly argue that the importance of accurate self-knowledge has been overstated (1288). They claim that, generally speaking, people neither possess nor actively seek accurate self-knowledge of an evaluative kind.⁴ On their view, this is a good thing because people are typically 'better served by having views of themselves that are a bit more positive than realistic' (1288). Notice here that, like Taylor, Brown and Dutton are endorsing the existence of moderate or mild 'illusions' about themselves and not serious or severe self-aggrandized delusions. Research, they claim, has definitively shown that with respect to nearly all imaginable personality traits, people believe that they are above average or superior to others (1290). Brown and Dutton further point out that on the rather unusual occasions when people do actively seek accurate self-knowledge, it is unlikely that they will find it. Brown and Dutton attribute this fact to our reluctance to give others negative feedback: 'If you don't have something nice to say about someone, don't say anything at all' (1292).

This clear lack of accurate self-knowledge is not an unfortunate state of affairs. In the first place, Brown and Dutton argue that it is rarely the case that inaccurate evaluative self-knowledge has 'punishing or fatal' results. While accurate self-knowledge about one's running ability may serve one well if one is taunting a wild beast, Brown and Dutton rightly point out that it is doubtful that this is information that the casual jogger will need (1293). One might object here that having overly positive views about one's abilities may have a serious negative effect on one's life by causing one to pursue a career or course of life for which one is not sufficiently talented. Brown and Dutton find this reply unconvincing for several reasons. In the first place, they point out that one's natural ability is generally not the exclusive or most important reason why one pursues a certain career and not others. One may pursue a career as a professional

4 The focus of their argument is on evaluative traits of a stable nature, such as one's intelligence, attractiveness, or athletic ability (1288).

sculptor because one loves making sculpture and finds the lifestyle and art-making process genuinely rewarding. Even if one does ultimately fail in one's pursuit of the sought-after career — i.e., one does not sell enough sculptures to make a decent living — it is unlikely that one will regret having given it one's best shot. Finally, Brown and Dutton claim that it is questionable whether accurate knowledge about one's natural ability in the domain of one's chosen career is necessary or desirable. Natural ability is just one factor among several that contribute to one's success. Indeed, an overly positive but false assessment of one's ability may actually bolster one's chance of achieving success by providing one with the confidence necessary to undertake projects or tasks that one would not otherwise undertake (1294). *If* these findings are correct, it seems that we could all benefit from a mild dose of positive self-deception. Or, rather, it would seem that the mentally healthy among us already *are* benefiting from a mild dose of positive self-deception. But might not a deeper analysis of the self-deceiver reveal that these 'benefits' are at the root of a serious evil for the person in question and those around him? It is to this question that we now turn.

II

For the purposes of our investigation, it will be helpful to distinguish between two kinds of self-deception that (among others) are discussed in the literature.⁵ (1) Alfred Mele has shown in some detail that a person's motivational state can have a biasing effect on her beliefs without her trying. It is often the case, for example, that one's wanting that *p* makes it easier for one to believe that *p*. An anti-divisionist/anti-intentionalist account such as this focuses upon the immediate effect that a person's motivation can have upon her beliefs.⁶ Mele is correct in thinking that many ordinary cases of self-deception work this way. (2) In addition to cases of this kind, there are perhaps less common cases of self-deception in which a person initiates a course of action which, if successful, causes her to become self-deceived. In trying to produce the desired belief in

5 To be sure, most theorists of self-deception believe that their analysis of self-deception is the only correct one. However, I am open to the possibility that there may be a number of phenomena deserving of the name 'self-deception.'

6 See Alfred Mele's *Self-Deception Unmasked* (1992). Ariela Lazar also defends an anti-intentionalist/anti-divisionist account of self-deception that stresses the influence of emotions and desires upon belief formation in 'Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs "Under the influence"' (1999).

herself, she may, for example, initiate certain thought patterns and/or associate with people who believe what she wants to believe. A common worry about this and similar accounts is that a person's awareness of her intention would undermine her self-deceptive project. How can a person be fooled by her own trickery, clever as it may be, if she sees it as such or is aware of her intention? Some theorists, such as Donald Davidson and David Pears, have tried to avoid this difficulty by introducing a division in the mind of the self-deceiver that keeps incompatible mental states apart.⁷ But not all theorists are convinced that a division of this kind is necessary. W.J. Talbott uses only 'innocent' or ordinary divisions, such as a division concerning memory, in accounting for the possibility of intentional self-deception.⁸ In what follows, I will discuss and evaluate both Mele-style (anti-divisionist/anti-intentionalist) and Talbott-style (anti-divisionist/intentionalist) cases of self-deception. I will remain silent about Pears- and Davidson-style (divisionist/intentionalist) cases because I do not think that there is any conceptual need for the divisions that they introduce in the mind of the self-deceiver in accounting for ordinary cases of self-deception.

1. *Joseph Butler: The Corruption of Conscience*

Nothing is more surprising, thought Joseph Butler, than the degree to which we are self-deceived (1958, 151). We judge others severely for vices of which we ourselves are guilty; however, we are so blinded by our self-deceit that we cannot see our own culpability (151). The cause of this self-deception, claimed Butler, is a partial and fond regard for ourselves. If it were not for this, it would be much easier for us to know ourselves — our characters, the biases of our minds, and the rightness or wrongness of our actions (152). But it generally does not occur to us that the judgments that we make against others can equally be brought against ourselves. We see ourselves in the best possible light, i.e., one that accentuates our positive qualities and downplays the negative ones if they are even acknowledged. William Ian Miller expresses similar worries about a person's 'blindness' to his own faults in his discussion of Jesus's teachings on hypocrisy and self-deception. Like Butler, Miller points out that while we are quick to call attention to the vices of others,

7 See Davidson's 'Paradoxes of Irrationality' (1982), 'Deception and Division' (1985), and 'Who Is Fooled?' (1998); and Pears' 'The Goals and Strategies of Self-Deception' (1985).

8 See Talbott's 'Intentional Self-Deception in a Single Coherent Self' (1995).

our self-love prevents us from noticing our own moral shortcomings. Following La Rochefoucauld, these considerations prompt Miller to suggest that our negative evaluations of others are far more accurate than our positive evaluations of ourselves (2003, 14). One unfortunate result of our 'blindness' or ignorance is that we are lacking an accurate picture of ourselves that would enable us to change and grow. We delight in our ignorance and are so pleased with ourselves that the possibility of change makes no impression on us.

Not only does this shield of ignorance render change all but impossible, according to Butler it also generates a great deal of malicious and unreasonable behavior that would likely not occur in its absence (156). Our self-deceit can bring us to view the most hideous of our actions and plans of action in an exceedingly flattering and unmerited light. Our self-deceit conceals from us the evil that lies in our planned actions and, in consequence, blocks the repulsion that we would feel in response to the awareness of them — the repulsion whose effect would often be to prevent us from acting in this way. The great danger here is that self-deception can 'carry a man almost any lengths of wickedness, in the way of oppression, hard usage of others, and even to plain injustice; without his having, from what appears, any real sense at all of it' (156). As long as we disguise to ourselves the evil that lies in our actions, there is no limit to what we are capable of doing. When we act maliciously with knowledge of the wrong that we are doing, we do at least preserve our conscience — that 'which is the guide of life' (163). However, when we act maliciously under the influence of self-deception, we not only perform the malicious act that we are self-deceived about, we also corrupt our conscience and thwart its ability to guide us towards the good. Butler is correct in thinking that our conscience cannot do the work that it is supposed to do when we are self-deceived. If we think of one's conscience as a kind of ethical processing machine, it would seem that its conclusions about what I should do will be true only accidentally if the information that I feed into it is false. Still, continuing the machine analogy, feeding the machine false information does not destroy the mechanics of the machine or its ability to draw conclusions in accordance with the body of information that it has. Indeed, the machine functions properly and produces true conclusions, but these conclusions instruct me about how best to act in a world that is fabricated or artificial. In short, self-deception does not corrupt or destroy conscience if we envision it as an ethical processing machine. It does, however, undermine its ability to instruct me about how best to act in the *real* world. It navigates me through a defective simulation of the world in pursuit of the good. If I am fortunate enough to achieve it, this is only an accident and not because my conscience has reached its conclusion based on an understanding of the facts.

2. Immanuel Kant: Inner Annihilation of One's Humanity

Immanuel Kant is not particularly prolix on the topic of self-deception, or what he refers to as 'inner lies,' but he does take the matter very seriously. In the *Doctrine of Virtue*, Kant opens the section On Lying with the following words: 'The greatest violation of a human being's duty to himself regarded merely as a moral being (the humanity in his own person) is the contrary of truthfulness, *lying*' (1996, 182). Although lying in general is the greatest violation of a human being's duty to himself, Kant believed that lying to oneself is the *worst* form of this. But Kant does not explicitly state why lying to oneself is worse than lying to another. Perhaps Kant's reason for stating this lies in his general remarks on the value of knowledge about oneself and its relation to the possibility of moral action. For Kant, moral action requires not merely that I act in *accordance* with duty, but that I am aware of what the moral law requires of me and act *from* duty. But the problem here is that I do not have infallible knowledge about the source of each of my actions and can be mistaken in believing that I am acting from duty (155). Kant, therefore, maintains that one has a wide duty to assess the moral worth of one's actions and to see to it that one's incentive for acting in accordance with duty is the thought of duty (155). Internal lies, at least about one's incentive for acting, conflict with one's duty and, as a result, with one's ability to ensure that one is acting from duty when acting in accordance with it. Kant condemned internal lies because he believed that the 'ill' of untruthfulness involved in such lies spreads into one's relations with other human beings. Mike Martin labels Kant's argument here (as well as W.K. Clifford's similar argument to be discussed in the following section) a 'slippery-slope' argument: 'taking a first, otherwise innocent step in a potentially dangerous direction will initiate a slide down a disastrous slope' (40). Martin considers Kant's argument (like Clifford's) fallacious on the grounds that not every case of self-deception has the potential to lead us down the slippery-slope that Kant has in mind. He adds that inner lies, such as overlooking a lover's obvious faults, do not 'encourage immorality and in some circumstances promises closer bonding and bliss' (40). Martin is correct in claiming that it does not follow from the fact that I will lie to myself about my lover's faults, that I will lie to her or to any other person on separate occasions or engage in any further acts of wrongdoing. However, in Kant's defense, it is arguably likely that this internal lie — that one's lover is better than she really is — will surface in *some way* in one's communicative relations with others. When asked by a friend about the character of one's lover, one will express one's inflated judgment or belief about her. But are instances of this kind best described as cases of other-deception or cases of self-deception? If I have successfully deceived myself, it would seem that

expressing this false belief to you is not a further lie, but the mere result of my self-deceit. Whether or not this is a case of self-deception or interpersonal deception, ‘the ill of untruthfulness’ can certainly spread into one’s relations with others in this way. But can it also spread into one’s communicative relations with others by producing in one a disposition to deceive and thus promoting additional cases of interpersonal deception? Self-deception, especially when it is habitual, may contribute to a general attitude of low regard for the truth. Martin, himself, concedes that slippery-slope arguments of this deflated kind, i.e., those that concern ‘general tendencies or dispositions rather than individual acts or states of self-deception’ (41), are more promising. Indeed, the same lack of esteem for the truth that is at the heart of internal lies can be found, it would seem, in external lies. If I am willing to deceive myself about a particular matter when it suits me, what is to stop me from deceiving you when it is likewise advantageous for me? What stops me, one might argue, is the simple fact that I am deceiving *you* and not *myself*. Although the truth may matter very little to me, I may realize that you hold the truth in high regard and, for this reason, resist deceiving you even when it is convenient for me to do so. In addition, my self-deception may not spill over into my interpersonal affairs because it is localized or limited to a particular subject. If I deceive myself about certain innate abilities for the purpose of enhancing my self-esteem, it does not follow that I will deceive you about the same (or even myself about certain other abilities that are not detrimental to my self-esteem). I may, as it happens, be far more concerned about my own self-esteem and well-being than I am about yours; this, in turn, may cause me to deceive myself about certain of my abilities but not you about yours. Of course, my self-deception *may* promote additional cases of interpersonal deception — depending upon the details of the situation at hand — but I do not see why all cases of self-deception *must* have this effect.

There is, however, an additional and more important reason why Kant condemns falsity as severely as he does. Not only does Kant view falsity as in some way rooted in humanity itself (183), he further states that it (falsity) is inimical to ‘the natural purposiveness of the speaker’s capacity to communicate his thoughts’ (182). On Kant’s view, a human being is bound up with truthfulness and cannot, therefore, *qua* moral being, use himself, *qua* natural being, as a mere means or ‘speaking machine’ by lying (183). The telos of human communication, for Kant, is truthfulness. And so, when a person lies — even to himself — he acts in opposition to his natural purposiveness in communication and, as a result, annihilates his humanity (182). Martin correctly points out here that it does not follow from the fact that something has a natural purpose that it is incorrect or immoral to change it. Moreover, he shows that communicating the truth is not the sole purpose for language; people use language

for multifarious purposes. If this is true, it would follow that it is only in virtue of my being an intellectual being — in virtue of my communicating *for* a purpose — that my acts of communication as a natural being have a purpose at all. I, considered as an intellectual being (perhaps in conjunction with my society or those around me), make the communication of my thoughts purposeful, and the purposes at issue need not be limited to the expression of truths.

3. W.K. Clifford: *The Danger of Credulity*

W.K. Clifford's treatment of self-deception falls within his more general discussion of the moral obligations that one has regarding belief. In 'The Ethics of Belief,' Clifford is well known for advancing the contentious view that 'it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence' (1886, 346). Clifford opens the essay by asking us to consider a shipowner who fears that his rather old and not especially well-built emigrant ship may be in need of repairs. The thought disturbed him so that he contemplated having it 'thoroughly overhauled and refitted,' even though this would be quite expensive. Perhaps for the purpose of putting his mind at ease, the shipowner willfully and deliberately overcame his doubts and convinced himself that his ship was safe and seaworthy. Unfortunately, Clifford tells us that the shipowner was wrong; the ship sank in mid-ocean along with the hopes, dreams, and bodies of the families aboard it (339). Clifford argues that the shipowner in this example was guilty of the deaths of the ship's passengers, for although he produced in himself the sincere belief that the ship was safe, '*he had no right to believe on such evidence as was before him*' (340). He adds that even if the ship had made its voyage successfully, the shipowner would be just as guilty '— he would only have been not found out' (340). According to Clifford, the rightness or wrongness of a person's believing is determined not by whether or not his belief happens to be true, but by how he acquired the belief and 'whether he had a right to believe on such evidence as was before him' (340).

On Clifford's view, no belief is insignificant or entirely disconnected from the rest of one's beliefs. Instead, all of one's beliefs are tightly bound together and form an 'aggregate.' Even if one does not act straightway on a newly acquired belief, the belief affects one's aggregate of beliefs by counting against some and reinforcing or supporting others (342). In this way, a seemingly innocuous belief may infect the rest of one's beliefs and 'some day explode into overt action, and leave its stamp upon our character forever' (342). Beliefs, Clifford continues, can have a similar effect at the social level. Just as no belief is ever insignificant with respect to the rest of my beliefs, no belief is ever insignificant with respect to the

rest of humanity's beliefs. Even seemingly insignificant beliefs (of even seemingly insignificant people, Clifford adds) contribute to our general conception of things and guide our thoughts and actions. It is for this reason that Clifford states that belief is ours *for humanity*, and not merely ours *for ourselves* (343). We abuse and desecrate our ability to believe when we allow our passions and not a careful and unbiased examination of the facts determine our judgment (343).

But the greatest evil to which believing for unworthy reasons gives rise is credulity. Clifford maintains that every time that we believe for unworthy reasons 'we weaken our powers of self-control, of doubting, of judicially and fairly weighing evidence' (344-5). Although this effect is bad enough in itself, we produce a much greater evil when this practice of believing for unworthy reasons becomes *permanent* — when we develop the *character trait* of being credulous (345). When this happens, Clifford claims that society sinks back into savagery (345). He further points out that a careless attitude about what I believe will give rise to a careless attitude in others about what they tell me or otherwise get me to believe. Indeed, if I am careless about the beliefs that I acquire, why should another revere the truth in my mind? If I get myself to believe certain things because I find them pleasant or comforting, and a friend or acquaintance is aware of this, what is to stop him from getting me to believe certain things for the same purpose? Clifford offers here what Martin (again) refers to as a 'slippery-slope' argument against self-deception, or against believing for unworthy reasons more generally. The unsuspecting self-deceiver initially adopts a belief because it is consoling and/or flattering; this belief then, perchance, gives rise to the character trait of credulity and sends him and his society sliding down the slippery slope to disaster. Marcia Baron has recently expressed similar worries about the corrosive effect of self-deception on a person's belief-forming processes. According to Baron, the more we fool ourselves, the less able we are to consider things objectively and from alternative points-of-view (1988, 438). When self-deception has this 'rippling effect,' it can interfere with our ability to act as responsible agents (434). Though this may be true in some cases, all three versions of the slippery-slope argument suffer from the same fundamental flaw: i.e., they fail to show that this is a necessary consequence of all cases of self-deception.

Having agreed to argue on pragmatic grounds, Clifford and Kant⁹ turn the question of whether or not self-deception is wrong into an empirical one. But if the question becomes an empirical one, it will prove difficult

9 This only applies to Kant's slippery-slope argument.

for them to defend the categorical thesis of the priority of truth. Might there be a more promising account of the wrongfulness of self-deception?

III

1. *Interpersonal Deception and Autonomy*

With respect to cases of interpersonal deception, it is generally not difficult to understand how (or at least *that*) the deceiver interferes with the autonomy of the deceived. What is it about the structure of interpersonal deception that has this effect? When I try to deceive someone, I purposefully present him with a distorted picture of the world. If I succeed, this world is internalized in the mind of the deceived and serves as a backdrop for his practical decision making and actions. Joseph Kupfer has called attention to this fact in his discussion of lies. He argues that a successful lie immediately restricts the freedom of the deceived. The lie, he claims, limits 'the horizon or content of his practical reasoning' (1982, 106). He goes on to say that

The lie misdirects the deceived's reasoning about future conduct in two complementary ways: it may suggest choices that are not available or eliminate from consideration live options. Both the illusion and elimination of choice limit the deceived's freedom. The first makes this reasoning practically fruitless, thereby retarding or postponing the deliberations necessary to the exercise of freedom. The latter excludes options from the scope of his volition; they are "shielded" from his view. (106)

Kupfer is particularly concerned with lies, but I think that the same can be said about all kinds of deception.¹⁰ By getting the deceived to hold a mistaken view of the world, the deceiver may cause the reasoning of the deceived to be practically fruitless (or at least *less fruitful*, let us say). The diminishing effect that the deceiver has upon the autonomy of the deceived will be most severe when the deception is persistent and the content of the deception is especially relevant to the practical decision making and actions of the deceived. At the other extreme, there are cases of deception that are relatively minor about matters that do not seem to

10 Kupfer describes lying as a kind of linguistic deception in which a person 'asserts something to another which he believes to be false with the intention of getting the other to believe it to be true' (104). Kupfer believes that lying is more opprobrious than other forms of deception (104).

be practically relevant in any obvious way. Kupfer has pointed out that most so-called white lies do not seriously restrict the freedom of the deceived and are told with his apparent interest in mind (113). Moreover, Kupfer claims that while all lies *immediately* restrict the freedom of the deceived, they do not always have this effect in the long run. Some lies may increase a person's freedom in the future and cause him to think and make decisions more intelligently (108). However, we should not judge such cases prematurely; some seemingly innocuous lies can interfere with the autonomy of the deceived in certain important and unanticipated ways. Thomas Hill has persuasively argued that a number of apparently benevolent lies — lies whose wrongfulness is not always obvious — do just this. For the purpose of his paper, Hill is especially concerned with a certain class of benevolent lies that allow him to focus the discussion on the point at issue. He offers several examples of the kind of benevolent lie with which he is concerned. He describes them as benevolent

because they are intended to benefit the person deceived, for no ulterior motives, and they actually succeed in giving comfort without causing pain. Despite the benevolent motives, there is no denying that deliberate lies were told. We are not dealing with examples of mere silence, evasion, ambiguous response, and the like. The lies, moreover, are not designed to protect incompetents from truths beyond their capacities to handle sanely and responsibly. In our sample cases a lie will protect someone from avoidable pain, but it is not needed to prevent serious physical or psychological damage, violent outbursts, gross misperception of reality, and so on. (1984, 253)

Hill is particularly concerned with a class of lies that falls outside the range of a variety of special problems. By excluding a number of problem cases that raise separate issues, Hill thinks that he will be able to show that many lies are wrong simply because they interfere with the autonomy of the deceived. Hill does not attempt to provide one settled account of autonomy in this paper; rather, he wants to provide us with several ways of understanding the term 'autonomy' that will help us assess the wrongfulness of lying.¹¹ Sometimes philosophers take autonomy to signify 'a capacity for distinctly human values.' Removing us from the pragmatic realm of consideration, the idea behind this conception of autonomy is that human beings (unlike most or all other animals) have the ability to value states of affairs that 'are not for the sake of *any*

11 I will follow Hill in not attempting to provide one settled account of autonomy in this paper. However, for the sake of clarity, I will at times identify the *kind* of autonomy that I have in mind.

anticipated pleasure (or prevention of pain)' (259). That is, those who accept this principle believe that human beings are not simply motivated by what is immediately before them or by what is taken to be agreeable (260). Indeed, Hill claims

that many of us at least implicitly accept the following principle: *First*, in dealing with competent human beings who are not violating anyone's rights, one should not presume that they prefer their own comfort (optimal experiences) over other values (such as the welfare of others, the completion of a project, self-awareness, etc.); and, *second*, when one aims to do something *for* others (say, from gratitude, charity, or love), one should not count their comfort (optimal experiences) as more important than values which they sincerely declare to be important to them, provided at least the declared value is not a violation of others' rights, not the result of momentary impulse, manipulation, obviously false belief, etc. One would violate this principle, for example, if one treated elderly but competent people like small children or pets, caring scrupulously for their comfort and physical well-being but ignoring any desire they might express about what happens in the world outside the confined area which they can experience or check on. (260)

Let us call the principle just stated the 'Important Value Principle,' 'IVP' for short. Hill is correct, I think, in claiming that most of us would accept the IVP. We believe that human beings value more than bare pleasure or comfort and that we should not presume the opposite. In reference to benevolent lies, Hill claims that it is incompatible with this principle to assume that a person's comfort is more important to him than his other values (263). When we tell a benevolent lie to a person (at least of the kind at issue), we do just this: we assume that he prefers comfort over some other value, such as the truth.

Still, it might be thought that in certain cases telling a benevolent lie to a person does not result in any loss of autonomy; i.e., autonomy is simply not an issue. We can imagine situations in which it appears that there is little if anything the person would be able to do on the basis of the relevant true information. If, as in one of Hill's examples, I lie to my mother who is in an extremely advanced stage of a deadly disease about the seriousness of her condition, it might seem that I do not deprive her of the opportunity to make any decisions at all. What is there for her to decide about? In the first place, she cannot do anything to prevent herself from dying. Furthermore, there is no point in her deliberating about what to do given that she does not have the health or time necessary to do much of anything at all. In response to this line of objection, Hill keenly observes that we should not be too quick to pass judgment on such cases. Even when a person faces near and certain death, she may still have a number of serious decisions to make concerning how she will spend the remaining time alive. He adds:

Good novelists and biographers know what philosophers too easily forget, namely, that the most important decisions in life are not always about external behavior, about what to *do* in the public world. How we face death, family tragedy, our own successes and failures, and the way others treat us, is partly a matter of decision, as Sartreans knew but exaggerated. (262)

Hill is quite right about this. Often the most important forms of rational control that a person has over herself pertain to her inner life: how she will interpret an important event, what she will imagine or think about, and which feelings she will try to prevent herself from experiencing.¹²

Among other things, Hill's discussion of benevolent lies highlights the fact that a lie is not always excusable when it is intended to (and actually does) comfort the deceived. Even when there are no decisions to be made on the basis of this information, Hill has pointed out that the deceiver can still interfere with the autonomy of the deceived, where autonomy is understood as a 'a capacity for distinctly human values.' Sometimes we just do value truth for its own sake. That is, even if we have good reason to believe that the information in question will not enter into our practical decision-making and actions, we may just value it in place of a comforting illusion. Robert Nozick's well-known thought experiment regarding an experience machine may help make this point clearer. Nozick asks us to contemplate the existence of an experience machine that, when plugged into it, would allow a person to have any desired experience 'from the inside.' All the while, however, the person having the experience (by way of mental states) will be 'floating in a tank, with electrodes attached' to his brain (1968, 42). Nozick adds that while one is plugged into the machine one will not know that the machine and not the actual world is responsible for his experience. If nothing but felt experience matters, should we not all plug into the experience machine and live out our wildest dreams and fantasies? Nozick replies by denying that felt experience is all that matters to us. The reason why people would not want to live their lives through mental states generated by the experience machine is that they want to do things in the actual world, not simply have the felt experience of doing these things. As James Griffin has captured this point:

I prefer, in important areas of my life, bitter truth to comfortable delusion. Even if I were surrounded by consummate actors able to give me sweet simulacra of love and affection, I should prefer the relatively bitter diet of their authentic reactions.

12 Or, as Hill has pointed out, the person in question may simply value the truth for its own sake.

And I should prefer it not because it would be morally better, or aesthetically better, or more noble, but because it would make for a better life for me to live. (9)

With respect to at least certain aspects of our lives, many of us just do take an interest in reality or veridical experience. Now, I think that something similar can be said about our interest in the truth. Just as we may prefer veridical to illusory 'felt' experience, we may also prefer to have true beliefs and not just comforting but false beliefs about the world. I want to fall in love with an actual human being and not a mere computer simulation of one; similarly, I may only want to believe that another is in love with me if he really is. Nevertheless, this is perfectly compatible with a person's *not* wanting the truth or *wanting* to step into the experience machine occasionally. But at certain times and with respect to certain subjects, a person may just value truth and veridical experience for its own sake.

It is here that the two conceptions of autonomy that I have discussed thus far merge. As Kupfer has shown, when somebody deceives me, he gets me to hold a mistaken conception of reality for the purpose of motivating my reasoning in a particular direction. As a result, I am to a greater or lesser degree cut off from reality and unable to reason and act in response to it; i.e., I experience a decreased degree of what I will refer to as *real world rational self-governance*. But our explanation cannot stop here; for we must understand why 'real world' rational self-governance is valuable. As Hill has shown, it can at times be obvious why one might want real world rational self-governance. Not only is a great deal of accurate information about the world necessary for survival, but it is also necessary in achieving certain internal and external ends that make one's life enjoyable and fulfilling. But even when the information at issue does not enter into one's practical deliberations, one may still value it for its own sake. And so, the fact that one might value reality or real world rational self-governance is often necessary in understanding the diminishing effect that deception has upon the autonomy of the one who is being deceived.

2. *Self-Deception and Autonomy*

Does a self-deceiver experience a loss of autonomy in the same way that the deceived often does in cases of inter-personal deception? In answering this question, let us refer back to a distinction that I made earlier in the paper between two kinds of self-deception: (1) anti-divisionist/anti-intentionalist self-deception, and (2) anti-divisionist/intentionalist cases of self-deception. Now our conclusions about whether or not (or to what degree) a self-deceiver experiences a loss of autonomy will depend upon

the kind of self-deception that we have in mind. An example will help us evaluate both kinds of self-deception:

Harry is a forty-year-old sales representative for a well-known automobile corporation. He is self-deceived in believing that he is more intelligent, charming, and better looking than he really is. Harry's inflated judgment about himself has had a number of seemingly positive effects upon his professional and personal life. He is sure that he can accomplish any job at work and has the motivation to do just this. With respect to his personal life, Harry is always the life of the party and is never alone on a Saturday night. He approaches new people with self-assurance and ease, and the ladies find his confidence irresistible.

Does Harry's self-deception cause him to be less autonomous than he would be in its absence? If Harry intentionally set about deceiving himself at some earlier point in time (in accordance with the *second model* or kind of self-deception) what can we conclude about his autonomy upon accomplishing his goal? Is he less autonomous now than he was before he became self-deceived?

In answering these questions, it will be helpful to consider the example in greater detail. Suppose, if you will, that one day Harry woke up and decided that he was sick and tired of being Mr. Doom-and-Gloom. Harry was sure that many people had more serious reasons for thinking badly about themselves, yet they seemed to get along just fine. He knew plenty of people who were not only unaffected by their various shortcomings, but seemed oblivious to them. Harry longed to join these others in their illusory but blissful state. 'Sure,' he thought, 'I will not know the miserable truth about myself, but I will be far happier as a result.' Although Harry was not all that bright, he thought himself clever enough to devise a workable plan without having to consult self-help books or other sources of advice. If Harry successfully deceives himself in this intentional way, what can we conclude about his autonomy? To some readers, it might appear that Harry has exercised his autonomy in deceiving himself. Moreover, it might be thought that after successfully deceiving himself, Harry is more autonomous than he was before accomplishing his aim. First of all, with respect to Harry's self-deceptive project, it is something that he intentionally initiated after having rationally reflected upon it and his other options. Harry, after contemplating his current state, decided that it would be best to delude himself about the truth and make his life bearable and possibly even enjoyable. What greater expression of personal autonomy could there be? Harry could have passively sat back and allowed the world to dictate his reality to him; instead, he deliberated about what was in his interest and did something about it. Moreover, he realized that self-deception would allow him to accomplish what he would be incapable of accomplishing in its absence. The new Harry is confident, hard working, and sociable; he completes diffi-

cult assignments at work and approaches others at social engagements with ease and self-assurance. Intentional self-deception, on this view, involves the clever exercise of one's autonomy over one's mental states. In pursuing his ends, the self-deceiver is not constrained by the real; rather, he rises above reality and manipulates it in order to get what he wants and take control of himself.

Some readers may not be satisfied with this explanation. While it is true that Harry has exercised his autonomy in deciding how to make his life more enjoyable, his decision is essentially the *rejection* of autonomy. In intentionally deluding himself, Harry has brought it about that he is less autonomous. He has decreased his autonomy *autonomously*, to be sure, but he has decreased it nonetheless. But how has he brought it about that he is less autonomous? After deceiving himself, Harry will experience a decreased degree of real world rational self-governance; he will be to some degree cut off from the real world and unable to reason and act in response to it (in 'real-time,' perhaps). This is in some ways analogous to Ulysses (in Homer's *Odyssey*) ordering his men to tie him to the mast so that he will be able to hear the call of the Sirens without succumbing to it and changing the direction of the ship. Although Ulysses has freely chosen to have his men tie him to the mast, there would seem to be a kind of freedom that he lacks as he pleads with his men to untie him while hearing the Sirens' enchanting call. We might want to say that like Harry, Ulysses has autonomously brought it about at T1 (before tying himself to the mast and hearing the call of the sirens) that he is less autonomous at T2 (while being tied to the mast and hearing the call of the Sirens).

But with respect to Harry, it might be objected here that the information that he lacks has no bearing on his autonomy. In assessing Harry's autonomy, all that matters is the way that he goes about making a practical decision on the basis of the information that he has. However, if we make this move, it would seem that we must also say the same about a person who lacks information as a result of some second person deceiving or manipulating him. But it is commonly thought that interpersonal deception or manipulation is wrong precisely because the deceiver or manipulator does interfere with the 'victim's' autonomy. Therefore, it would seem that the loss of information *itself* — regardless of its source — can decrease a person's autonomy by causing him to reason and act in response to a world that is not entirely as it seems. Marcia Baron has commented on this fact in discussing the similarities and differences between self-deception and inter-personal deception. She correctly points out that the effect that inter-personal deception and self-deception have upon the autonomy of the deceived is the same: namely, they both cause the deceived to operate in response to 'a warped view of the circumstances' (1988, 436). In response to this suggestion, one

might wonder whether the loss of real world rational self-governance produced by this 'warped view' is any worse than the lack of other kinds of autonomy that we all experience as a necessary consequence of any number of practical decisions. For example, in deciding to spend part of one's life in pursuit of a PhD in philosophy, one is simultaneously bringing it about that one will not pursue a PhD in astrophysics during that same time period. While one's decision to pursue a PhD in philosophy has rendered certain possibilities more concrete (e.g., acquiring a teaching position in philosophy), it has also diminished other live possibilities that stood before one in the past. Should we not also complain about *this* loss of autonomy? Given some conception of autonomy, it strikes me as undeniable that in autonomously striving to actualize certain practical possibilities, one will necessarily bring it about that one is less autonomous with respect to other courses of action; i.e., one will be physically or metaphysically cut off from them. Nevertheless, this lack of autonomy describes one's logical, metaphysical, or physical *situation* and is inevitable for this reason: no matter how strong my desire or 'wish' may be, I simply cannot undo the consequences of what I did in the past. In contrast, the loss of autonomy (real world self-governance) that I will experience as a result of intentional self-deception is avoidable and to this extent under my control. To be sure, the loss of autonomy that I experience as a result of intentional self-deception is not more *severe* than that which necessarily results from other intentional actions. However, given that the former is avoidable, we may have certain practical reasons for being more concerned about it than the latter.

Before drawing some general conclusions about self-deception and autonomy, I want to consider cases of self-deception as described by the anti-intentionalist/anti-divisionist model mentioned earlier (self-deception of the first variety). What does this kind of self-deception involve? On this view, most people who are self-deceived do not intentionally bring this about; rather, their motivational state has this effect upon them without their trying. Someone, such as Harry, who is self-deceived in believing that he is more attractive and talented than he really is does not do anything to bring this about. As Mele has convincingly shown, if we suppose that it is much more costly for Harry to believe falsely that he is not very attractive than it is for him to believe falsely that he is, it will often take less to convince Harry that he is attractive. What can we say about Harry's autonomy given this explanation of his self-deception? Harry, on this view, does not demonstrate rational control in getting himself to believe flattering or desirable things about himself. Indeed, his self-deception is not even something that he *does* or brings about at all. Instead, it is something that happens to him; he is the unsuspecting and passive victim of it. What then should we say about this variety of cases? Our answer must be that the loss of autonomy that

the person who is self-deceived experiences is equivalent to that which the person who is deceived by another or simply mistaken experiences.

For our purposes, then, the important and relevant difference between the two kinds of self-deception is causal. That is, the belief (or beliefs) that the self-deceiver is self-deceived in holding either *was* or *was not* autonomously brought about. However, when we focus our attention on the belief (or beliefs) itself, we can see that the loss of real world rational self-governance in the two cases is identical. In order to complete our analysis of self-deception and autonomy, let us now take a closer look at this loss itself. Does it always result on balance in a decrease in one's autonomy?

While holding a grossly mistaken picture of the world will obviously interfere with one's ability to survive, it is less clear that moderate illusions about the world will have this effect. Indeed, as some empirical research has purported to have shown, such illusions may actually promote one's mental health and have a number of positive effects upon one's quality of life. With respect to self-deception, then, the lack of real world rational self-governance that the self-deceiver experiences may not be undesirable or capable of posing a serious risk to his well-being. Following Kupfer, we can also conclude that self-deceiver's loss of autonomy may actually increase his on balance autonomy in the long run. As in the Harry example, a person's self-deception may enable him to accomplish aims that he would be practically unable to accomplish in its absence. Moreover, in certain rare cases, we might even want to say that a person's self-deception does not cause him experience *any* loss of real world rational self-governance. Although the self-deceiver in question lacks an accurate picture of the world, we may be reluctant to say that this results in a loss of real world rational *self-governance*; for we can imagine a case in which awareness of the truth would actually subvert a person's agency or ability to be self-governing at all. Consider the parallel case regarding Ulysses and the call of the Sirens. If we interpret the call of the Sirens as subverting Ulysses' autonomy, we may want to deny that he experiences *any* loss of autonomy after having been tied to the mast. In this case, we might think that there is autonomy to gain but none to lose. However, even in cases of self-deception that share this rough structural similarity, there may be a kind of autonomy that the person would experience upon becoming self-deceived. She may hold the truth about a certain sphere of her life in such high regard that she would rather suffer from a debilitating spell of anxiety or depression than be comforted by an illusion. We can use Hill's conception of autonomy as a capacity for distinctly human values in understanding this possibility.

The risk involved in deceiving oneself is that the lack of information about the world may cause one to make mistakes that interfere with

one's well-being. To speak metaphorically, one's present *self* may not be expressed in one's actions in the real world; for if one's vision were not clouded by illusion, one might very well pursue different courses of action. But as the empirical research discussed above has suggested, when the self-deception in question is moderate, the risk involved may be negligible and the consequences not so disastrous after all. This conclusion is not only at odds with previous conceptions of mental health, but with the personal importance that many of us — philosophers in particular — place upon knowledge about ourselves and the world. Perhaps our natural disposition or lust for knowledge has its own biasing effect upon our beliefs; i.e., it allows our romanticization of the truth to go unnoticed. It causes us to see something that is not there. Still, given that one *does* value the truth even when it will not alter the felt experience of one's life (in conjunction with Hill's conception of autonomy as a capacity for distinctly human values), we can see why this counts for something with respect to one's autonomy and desire not to be self-deceived. I lack some autonomy if I cannot prevent myself from becoming self-deceived or am powerless over my mental states. However, should someone ask me, 'What's so great about reality? Why do you value the truth over illusion?' I must rest content with the reply 'I just do'; for in such cases there can be no further justification.

Received: January 2004

Revised: July 2004

Works Cited

- Baron, Marcia. 1988. 'What's Wrong with Self-Deception?' *Perspectives on Self-Deception*. Brian P. McLaughlin and Amelie Oksenberg Rorty, eds. Berkeley: University of California Press.
- Brown, Jonathon D. and Keith A. Dutton. 1995. 'Truth and Consequences: The Costs and Benefits of Accurate Self-Knowledge,' *Personality and Social Psychology Bulletin* 21 1288-96.
- Butler, Joseph. 1959. *Fifteen Sermons Preached at the Rolls Chapel and a Dissertation upon the Nature of Virtue*. W.R. Matthews, ed. London: G. Bell & Sons.
- Clifford, William Kingdon. 1886. *Lectures and Essays*. Leslie Stephen and Frederick Pollock, eds. London: Macmillan.
- Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon.
- Hill, Thomas. 1984. 'Autonomy and Benevolent Lies,' *Journal of Value Inquiry* 18 251-67.
- Kant, Immanuel. 1996. *The Metaphysics of Morals, Cambridge Texts in the History of Philosophy*. Mary Gregor, trans. and ed. Cambridge: Cambridge University Press.

- Kupfer, Joseph. 1982. 'The Moral Presumption Against Lying,' *Review of Metaphysics* 36 103-26.
- Martin, Mike. 1986. *Self-Deception and Morality*. Lawrence: University of Kansas Press.
- Mele, Alfred. 2001. *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Miller, William Ian. 2003. *Faking It*. Cambridge: Cambridge University Press.
- Nozick, Robert. 1968. *Anarchy, State, and Utopia*. New York: Basic Books.
- Taylor, Shelley E. 1989. *Positive Illusions: Creative Self-Deception and the Healthy Mind*. New York: Basic Books.

