Self-Deception Won't Make You Happy

1. Introduction

An influential literature in psychology claims that self-deception is characteristic of mental health. Most notably, Shelley Taylor, in a series of works that touches three decades, argues that "positive illusions" contribute to the production of better mood, better popularity, better ability to care for others, creativity, productivity, resilience from stress, and ultimately happiness.[1] So Taylor is enamored of the "adaptiveness" of human positive illusions.[2] Not having them is the hallmark of depression. Having them is a great boon (with a few qualifiers for when they get too extreme). Her work is not explicitly normative, but it seems to me to have the implication that many illusions she discusses are to be encouraged. Fully realistic assessments of oneself are not prized.

The question I shall be concerned with *is* normative. I shall ask: If we desire happiness, is it practically rational to pursue policies of self-deception? Although Taylor does not address this question specifically, her work might inspire one to answer in the affirmative.

To be precise, the reasoning I wish to refute is as follows. Having beliefs about oneself that are positive (flattering, make oneself out to be admirable, and so on) leads to happiness; work such as Taylor's in social

[1]Taylor doesn't distinguish positive illusions from self-deception. See section 2 for a distinction between self-deception and a type of bias I call "self-inflation." The *locus classicus* of the series I'm referring to is Shelley E. Taylor, *Positive Illusions: Creative Self-Deception and the Healthy Mind* (New York: Basic Books, 1989). This is the piece to which I will be most concerned to respond, since it is highly influential and attempts to integrate the most research. But there are a number of other pieces that purport to support the conclusions of *Positive Illusions*. See, for example: Shelley E. Taylor and Jonathon D. Brown, "Positive Illusions and Well-Being Revisited: Separating Fact from Fiction," *Psychological Bulletin* 116 (1994): 21-27; Shelley E. Taylor, Margaret E. Kemeny, Geoffrey M. Reed, Julienne E. Bower, and Tara L. Gruenewald, "Psychological Resources, Positive Illusions, and Health," *American Psychologist* 55 (2000): 99-109; Shelley E. Taylor, Jennifer S. Lerner, David K. Sherman, Rebecca M. Sage, and Nina K. McDowell, "Are Self-Enhancing Cognitions Associated with Healthy or Unhealthy Biological Profiles?" *Journal of Personality and Social Psychology* 85 (2003): 605-15.

[2]Taylor doesn't use "adaptive" in the sense offered by the theory of natural selection, but rather in sense of *useful in life* or *productive of mental health*.

psychology shows that this is true even when such beliefs depart from the truth in being *overly* self-flattering. In order to achieve happiness, therefore, it is a good idea to pursue a *policy* of self-deception concerning the self and one's situation, making self and situation out to be better than frank appraisal would support. (Again, I don't attribute this reasoning to Taylor herself, but hold that her work suggests it.)

I argue, on the contrary, that self-honesty is a superior policy to self-deception, if one wishes to achieve happiness. There are three sides to my overall argument.

First, I argue that Shelley Taylor's work, empirical and otherwise, cannot properly be used to support the target reasoning. This is because: (i) Taylor misapplies the notion of illusion to cases of self-fulfilling belief that, although possibly conducive to happiness, needn't be illusory; (ii) Taylor misconstrues *imagining* positive outcomes as part of planning as illusion; and (iii) Taylor fails to show that the beliefs resulting from what we might call *self-inflation bias* are the kind of beliefs needed for happiness.

Second, I give three arguments that a policy of self-deception does *not* produce choiceworthy happiness. The first two of these arguments even support the contrary claim: self-deception leads to *un*happiness. I hold: (a) self-deception, as a producer of false belief, undermines one's ability to satisfy one's desires; (b) self-deception creates an anxiety-provoking internal tension in the mind of the self-deceiver; and (c) self-deception could only produce happiness in a thin sense, which I call *Matrix happiness*.

Third, I argue that an effective route to happiness lies in *honest imagining*, which involves honesty with oneself about one's own abilities and the offerings of one's environment, and positive imagination about what to do with abilities and environment.

2. The Ideas in Question

Before turning to these arguments, let's stabilize the notions under investigation with some conceptual work on happiness and self-deception.

2.1. Happiness

What do I mean by "happiness"? Some hold that an array of pleasurable or positive feelings constitutes happiness. Others hold that having certain genuine "external" goods—health, friends, activities, enjoyable possessions—is required as well.

I hold that both positive feelings and the possession of genuine external goods are constituents of happiness. Nor are they neatly separable.

Human relationships (having friends) would, by most lights, fall in the category of external goods. But a relationship is not one we desire without mutually held positive feelings; so the relationship, usually thought of as an external good, does not exist without the pleasurable sentiment, usually thought of as an internal good. Relationships thus don't fall neatly into either category. Furthermore, *activities* need a favorable environment to be possible and have mental components; they defy the dichotomy as well. Nevertheless, let's treat external goods and internal positive sentiments as separable for now, since doing so will help us see some useful distinctions.[3]

One may have positive sentiments or not; and one may have genuine external goods or not. There are thus four categories of interest.

First, having neither external goods nor positive sentiments is *misery*.

Second, suppose one is arrayed with a fine panoply of genuine external goods but largely lacks positive sentiments in life. I call this state of affairs the *Woody Allen condition*, after Woody Allen's character in the film "Manhattan," who is largely successful in life and is dating a beautiful young woman, but who lacks positive sentiments about his state of affairs. The Woody Allen condition is not just a form of depression; depression involves loss of motivation and hence leads to inactivity, but neither of these problems (necessarily) obtains in the Woody Allen condition.

Third, one may have positive, pleasurable sentiments without having much or anything by way of genuine goods. I call this *Matrix happiness*, after the film "The Matrix," in which humans are kept in cells by artificially intelligent robots and kept in a (relatively) positive-feeling state of mind by a massive computer program that influences their conscious experience. The happiness of most humans in that film did not include genuine external goods.

Fourth, if one has genuine external goods and positive sentiments, one has *choiceworthy happiness*. I say "choiceworthy" because even the advocate of the view that only sentiments are needed for happiness would choose this form of happiness, if given the option.[4] Choiceworthy

---

[3]My way of framing things here owes much to Aristotle's *Nicomachean Ethics*, although I do not want to present any of my views as being an interpretation of Aristotle.

[4]Alfred Mele has pointed out in personal communication that there are some complications here. The various types of happiness come in degrees and vary along multiple dimensions. In saying that even the proponent of Matrix happiness would choose choiceworthy happiness, I am assuming a choice in which the latter has everything the former has and more. The situation gets trickier if we're faced with a choice between a very intense Matrix happiness and a low-level form of choiceworthy happiness. I won't attempt to give a systematic theory of how to negotiate that trade-off here, as doing so would be an entirely different project. But my arguments in later sections are relevant to this issue. Basically, most Matrix happiness that would come in a trade-off with choice-

happiness is, I believe, more stable than either the Woody Allen condition or Matrix happiness. Matrix happiness is likely to be intruded upon often by the natural connection between people's perception of the world around them and their sentiments. The Woody Allen condition is likely to devolve into depression and then misery.[5]

We can summarize these types with the following chart:

	Have positive sentiments	Lack positive sentiments
Have worthwhile external goods	*Choiceworthy happiness*	*Woody Allen condition*
Lack worthwhile external goods	*Matrix happiness*	*Misery*

A couple of clarifying points are in order.

First, concepts of happiness or unhappiness can be applied *globally* or *locally*. Globally, one discusses a person's overall state of happiness; locally, one speaks of a person's happiness in a certain area of life. It's generally assumed that one's global happiness is largely a function of aggregated local areas of happiness. I make the analogous assumptions using the distinctions I draw here. For example, if one has local Matrix happiness in many areas of life, the person will be Matrix happy globally; and so on.

Second, just as *believing* one has external goods is not sufficient for having them, *believing* one has positive feelings or sentiments is not sufficient for having them. Beliefs about feelings may often be right, but we'll easily be misled and run afoul of much empirical psychology if we assume that people's beliefs about their own mental states, including feelings and sentiments, are always right.[6]

---

worthy happiness is unstable, since reality gets in and impinges on the affect of even the most willful self-deceiver. See section 4.1.1 for elaboration.

[5]As seems to happen in "Manhattan." The Woody Allen condition will not, from here on out, play a big role in this essay. One reason, in addition to systematicity, that I include it in this section is that I want people to see that much of the unpleasantness they face in life is from the Woody Allen condition. Many Westerners are unhappy, despite being well fed, sheltered, and having friends to relate to. Compare this to the attitude of Duke Senior in *As You Like It*, whom I discuss in section 7.

[6]On this point, see especially Richard E. Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, N.J.: Prentice-Hall, 1980).

The distinctions of this section will play into the conclusions of this article as follows. The only kind of "happiness" that pursuing a policy of self-deception is at all likely to promote is Matrix happiness. Now you can divide Matrix happiness into two types. Some Matrix happiness will lead to choiceworthy happiness, since positive sentiment can breed positive action that yields genuine external goods. But some Matrix happiness simply sets one up for a fall and disillusionment. The Matrix happiness that a policy of self-deception produces is likely to be the latter sort.

Keep in mind that I'm focusing on the question of what's the best *policy*. On the issue of whether self-deception can contribute to happiness, people often raise individual examples of how self-deception *can* bring about happiness in particular circumstances. But to justify a policy of doing *x* for the sake of *y*, it is not enough for it to be *possible* for an instance of *x* to play a causal role in the occurrence of an instance of *y*; rather, making a policy of doing *x* for the sake of *y* is only justifiable, in general, if the occurrence of *x* can lead to a rational expectation of *y* in the circumstances one is actually in. So one example of a self-deception that led to happiness, choiceworthy or otherwise, will not refute my argument that it is not instrumentally rational to self-deceive for the sake of happiness. Suppose someone argues that it is bad to give sharp knives to children under the age of four. Would it refute this argument to show *one possible case* of a three-year-old child who did something good with a knife, for example, cut carrots? Or is that an exceptional behavior that in no way justifies the practice of giving children knives? Thus, to assess the policy, we'll have to attend to the systematic features of the states in question, self-deception and happiness. Cases are relevant, but only insofar as they represent actual patterns.

2.2. The policy of self-deception for the sake of happiness

How might one hope to deceive oneself to become happy? One cannot turn self-deception on and off like flipping a switch, so a policy will be necessary. So one must adhere to habits that cultivate *beliefs* that one *believes* will cause happiness.[7]

In particular, one wishes to believe that certain negative propositions are not true (although the weight of one's information suggests they are) and that certain positive propositions are true (although the evidence suggests otherwise). In order for it to be a policy of *self-deception* that one is pursuing, one must have beliefs already that are in tension with the desired beliefs. (I'll explain why shortly.) This is an important difference

[7]Immediately the worry arises: what if the self-deceptive policy also corrupts one's ability to judge *which* beliefs will make one happy? I'll develop versions of this worry below.

from most of the positive illusions that Taylor discusses, for her "healthy" subjects seem to be not at all *divided* on questions of whether they are, for example, more liked than most people (even though they aren't) or are better drivers than most (even though they aren't).

Along these lines, let's draw a distinction for use in the wider argument. "Self-inflation bias" will refer to a *general* tendency to form beliefs about oneself that are more flattering than reality justifies, where those beliefs are not necessarily controverted by other information in the agent's mind. For example, believing one is an above-average cook, where one has not much evidence for or against this, is most likely the product of self-inflation bias. "Self-deception," on the other hand, refers either to a process or a state. "Self-deception" (as a state) occurs when a belief exists that is contrary to the evidence and epistemic norms an agent has, where *motivations* topically related to the content of this belief are causally implicated in its having come about.[8] Self-deception (as a process) is how one comes to be in this state. Believing one's wife has been faithful, when one is aware (at some level) of evidence to the contrary, is thus a state of self-deception, whenever the belief is caused by motivations such as wanting it to be the case that the wife has been faithful or wanting to believe that she has been. Pre-theoretically, one might refer to self-inflation bias as a form of self-deception, but keeping them distinct will be important for this paper at certain points. In sum, the main differences between self-inflation bias and self-deception are two: (i) the self-deceiver is in some way epistemically divided, while this isn't typically the case for someone merely in the grips of self-inflation bias; (ii) self-deception is caused by specific motivations or desires, while self-inflation bias is a general tendency.

Given this distinction, we can sharpen the focus of the paper. Pursuing a policy of self-deception will mean attempting to form beliefs that are contrary to negative information about the self *that one already possesses*. It is *this* that I shall attempt to show does not lead to happiness. Nevertheless, some critical examination of Taylor's work on positive illusions, which are often from self-inflation bias, will be useful for showing that her work doesn't justify the policy of self-deception that I am opposing.

A policy of self-deception characteristically involves the following. One has views about which beliefs are desirable to have and which are desirable to get rid of. Desirable beliefs on this policy might have contents such as: *I am highly intelligent*, *my friends are attractive*, *my rela-*

[8]See D.S. Neil Van Leeuwen, "Finite Rational Self-Deceivers," *Philosophical Studies* 139 (2008): 191-208, for a fleshed-out version of this definition, and "The Product of Self-Deception," *Erkenntnis* 67 (2007): 419-37, for distinctions among types of self-deception.

tionship is entirely healthy, and so on. The policy also involves a commitment to ignoring (or explaining away) information that counts against the desired beliefs, and to actively attending to information that seems to confirm them.

Such commitments are common. If one has a rich Mafioso benefactor, one might commit to ignoring evidence that the man is a killer, while actively seeking out evidence that he does good things for the community. Sports fans, in the grips of the view that they must believe their team is the best, actively seek out evidence that would support this belief while ignoring evidence to the contrary. Mountains of contrary evidence are ignored; teaspoons of supporting evidence are celebrated.

About which topics will someone pursuing a policy of self-deception for the sake of happiness have practices of selective attention? One might, for example, ignore the fact that one rarely understands books, and focus on the clever comment one made the other day. But if one is already secure in one's intelligence, there will be no point in inflating one's estimation of it. If one is to pursue the policy I'm discussing, one must focus the self-deception on areas about which one is insecure. The hope is that once one has formed the self-deceptive belief that is meant to promote happiness, the insecurity will fall away.[9]

To summarize, there are four main elements of a policy of self-deception for the sake of happiness.

(1) Awareness of areas in one's life that are felt to be lacking or about which one has insecurities.
(2) On the basis of awareness of the sort mentioned in (1), selection[10] of which beliefs will promote happiness by working against the tendency toward negative affect the awareness engenders.
(3) Commitment to attending to information that seems to confirm the beliefs selected in (2).
(4) Commitment to ignoring evidence that disconfirms beliefs selected in (2).

The possible success of this policy will depend on the persistence of the commitments mentioned in (3) and (4) after the awareness mentioned in (1) has been extinguished, since that awareness is precisely the sort of thing that the policy is meant to extinguish. It's not incoherent to suppose

[9]This hope is not entirely vain, since there is empirical evidence that attempts at memory suppression are to some extent effective. See: Michael C. Anderson, Kevin N. Ochsner, Brice Kuhl, Jeffrey Cooper, Elaine Robertson, Susan W. Gabrieli, Gary H. Glover, and John D.E. Gabrieli, "Neural Systems Underlying the Suppression of Unwanted Memories," *Science* 303 (2004): 232-35. It is not clear, however, to what extent suppressed memories actually go away as opposed to simply having a "cap" inhibiting access.

[10]It's not necessary that this selection be conscious.

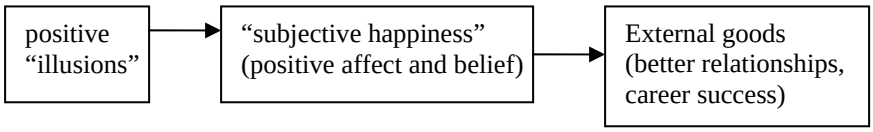
that persistence is possible, so I'll rest my argument against this sort of policy on other grounds. Now let's proceed with the wider argument to see whether Taylor's empirical work offers anything that would help justify this sort of policy.

3. Does Taylor's Work on Positive Illusions Support Self-Deception?

In the next three subsections, I'll argue there are serious problems with using work such as Taylor's in social psychology to infer that self-deception can contribute to choiceworthy happiness. Such inference involves a confused application of the term "illusion," where the confusion can be of three varieties: (i) conflation of illusion with self-fulfilling confidence that one can achieve a positive outcome in a certain domain, (ii) conflation of illusion with future-oriented imagining, and (iii) confusion of self-flattering belief, which is often illusory, with a justified sense of self-worth.

3.1. "Illusions" of success?

Taylor argues that positive illusions contribute not only to having positive feelings, but also to greater success in life. In my terms: Taylor holds that positive illusions, such as the belief that one is more well-liked than is actually the case, contribute to the attainment of choiceworthy happiness. How are the illusions supposed to bring about their beneficial effects? The model is something like this:[11]



What's the argument that the first link obtains? Taylor cites a number of empirical studies that suggest that people who are more depressed have more accurate self-perception and others that suggest that people with positive illusions are happier.[12] We may therefore accept that having un-

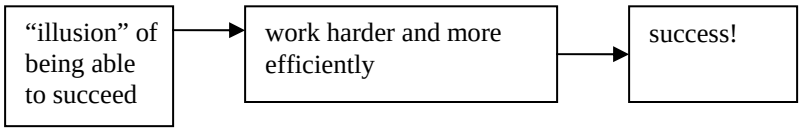
[11]Sometimes Taylor talks as if positive illusions in the form of beliefs *cause* the positive feelings. Other times she talks as if the feelings are just *part* of the illusions. I'll assume the first interpretation here.

[12]Taylor cites, among others, the following studies to support the correlation between depression and accurate appraisal: Nicholas A. Kuiper, "Depression and Causal Attributions for Success and Failure," *Journal of Personality and Social Psychology* 36 (1978): 236-46; Nicholas A. Kuiper and M.R. MacDonald, "Self and Other Perception in Mild Depressives," *Social Cognition* 1 (1982): 233-39; Ross Rizley, "Depression and Distor-

realistically flattering views of the self is *correlated* with not being depressed. This is of course a far cry from having shown that unrealistic self-assessment *causes* positive feelings about the self. It may be that the two have a common cause, or that the positive feeling is causing the unrealistic self-assessment.[13] Nevertheless, let us grant this link for the time being so that we can get a better sense of the structure of Taylor's view.

What about the second link in the diagram? What's the argument that that obtains? The general idea, which is true, is that having a positive affect can have good consequences. Positive affect can make it easier to engage in relationships, easier to be productive, and easier to engage in healthy behaviors. So if these links obtain (granting the first), does that show that positive illusions cause genuine external goods?

Let's focus on the word "illusion." Often, the "illusions" Taylor refers to actually turn out *true*, at least if the model is correct. Thus it seems to be a conceptual confusion to call them "illusions," since illusions are not things that turn out true. Let's look at a specific example. In Chapter 2 of *Positive Illusions*, Taylor argues that illusions that one can succeed will lead to success. So we have:



It may seem astounding that the word "illusion" is used at all here. But there is a more nuanced version of the view that Taylor is suggesting. The idea would be that having an unrealistically high estimation of the probability of success raises the probability of success, although not to the level of the unrealistically high assessment, which therefore still deserves the name "illusion." This version of the view doesn't suffer from the conceptual confusion that the simpler version suffered from, but I see no evidence that this is what is happening. In particular, Taylor presents no evidence that people in the kinds of contexts she focuses on are actually making *probability* estimates or anything like them. It may be that outcomes envisioned are brighter than those that typically come about,

---

tion in the Attribution of Causality," *Journal of Abnormal Psychology* 87 (1978): 32-48. One example study cited in support of the correlation between positive affect and positive illusion is Neil D. Weinstein, "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology* 39 (1980): 806-20. See Taylor's *Positive Illusions* for more references. The follow-up literature is also substantial.

[13]I suggest that the direction of causality goes from positive affect to self-deception or illusion in my "Spandrels of Self-Deception," *Philosophical Psychology* 20 (2007): 329-48, and argue that it does below.

but that doesn't indicate that there are any false *beliefs* about the present state of affairs or that any false beliefs that may exist are beneficial to the attainment of genuine external goods in the future. It just means that envisioning or imagining a brighter future is helpful. In fact, there are some interesting remarks that Taylor and Jonathan Brown make in response to criticisms of their views that suggest precisely that, although they don't seem to realize the implication:

> When individuals are in a deliberative mindset, attempting to make a decision, their positive illusions are quite modest; but when they are in an implemental mindset, attempting to put a decision into effect, illusions increase dramaticially ... there may be windows of realism during which people suspend their illusions, at least somewhat, in favor of a more realistic vantage point.[14]

There is some truth to this passage: individuals operate with more than one reservoir of representations, and they switch, depending upon what kind of practical setting they're in. But the tendency to label as "illusion" every reservoir that's not entirely reality-tracking is a mistake. Taylor and Brown appear to admit that when it's needed, normal individuals are fairly good at keeping track of reality. Of course, when they need to "implement" or produce actions, their representations of the future become exceptionally positive. But that doesn't mean those representations are illusions, and the fact that reality is gripped in the background suggests they're not. This takes us to the next section.

3.2. Belief or imagining?

People no doubt have many illusions about themselves and their lives. Taylor is right to study them. But in claiming that illusions lead to positive future outcomes, she also conflates having illusory beliefs about the self with having future-oriented *imaginings*. Let's first consider an example of genuinely illusory belief. Taylor writes:

> When people whose driving had involved them in accidents serious enough to involve hospitalization were interviewed about their driving skills and compared with drivers who had not had accident histories, the two groups gave almost identical descriptions of their driving abilities. Irrespective of their accident records, people judged themselves to be more skillful than average, and this was true even when the drivers involved in accidents had been responsible for them.[15]

The term "illusion" certainly applies to the majority of minds of at-fault drivers who judge themselves to be above-average drivers. But compare

---

[14]Shelley E. Taylor and Jonathon D. Brown, "Positive Illusions and Well-Being Revisited: Separating Fact from Fiction," *Psychological Bulletin* 116 (1994): 21-27, pp. 25-26.

[15]Taylor, *Positive Illusions*, p. 11.

that with the optimism about the future that Taylor discusses in her chapter entitled "Escape from Reality: Illusions in Everyday Life."

> One of the more charming optimistic biases that people share is the belief that they can accomplish more in a given period of time than is humanly possible. This bias persists in the face of innumerable contradictions. Perhaps the most poignant example of unrealistic optimism is the daily to-do list. Each day, the well-organized person makes a list of the tasks to be accomplished and then sets out to get them done. Then the exigencies of the day begin to intrude: phone calls, minor setbacks, a miscalculation of how long a task will take, or a small emergency. The list that began the day crisp and white is now in tatters, with additions, cross-outs, and, most significantly, half its items left undone ... the pattern persists day after day, completely unresponsive to the repeated feedback that it is unrealistic.[16]

It is surely problematic to categorize such future-oriented list-making as illusory. Of course, illusion may be in the background sometimes. But one needn't be convinced that one will get through the entire list in order to make it up. Oftentimes the list just reflects a set of priorities and is made so that one not forget them. Where is the illusion in that? Arguably, it makes sense to write down more than one can actually do, in case things go more smoothly than usual. It costs little to add an item to the list and can be beneficial.

The point is general, applying not only to the categorization of mental representations underlying list-making. Representation of future events needn't be illusory, even if their contents are unlikely to end up true. It could simply be positive future-oriented imagining.

By lumping under the single label of "positive illusions" both self-inflated positive beliefs and imagined future outcomes that are useful for planning, Taylor has done a disservice. Imagining can be valuable, even in cases in which the contents of one's imagining depart widely from reality. When I exercise I sometimes imagine I'm an Olympic athlete, which motivates me to continue. There's no illusion here, just imagining. The beneficial effect of positive imaginings *does not* show that real illusions (like the driver illusion) lead to positive future outcomes, as Taylor's equivocal use of "illusion" suggests.

3.3. What beliefs contribute to happiness?

Does believing you're a better driver than you are make you happier? Does believing you're smarter than you are make you happier? Does believing your son is smarter than he is make you happier? It may be that beliefs like these are *correlated* with positive affect—with both Matrix and choiceworthy happiness. Do we have reason to believe that such be-

[16]Ibid., p. 34. She does not refer to the unrealistic optimism as "illusions" in the text, but the fact that they are discussed in a chapter with the heading "Illusions in Everyday Life" indicates that Taylor is thinking of them as such.

liefs *cause* positive affect to the extent that they must, if it is to make sense to have a policy of inculcating them via self-deception?

Many people *know* they are of above average intelligence and are depressed nonetheless. An experienced taxi driver, who is surely an above average driver and knows it, can be depressed. Many unhappy fathers in the world know their sons or daughters are highly intelligent. These are all examples of people with *true* beliefs with contents like those posited in Taylor's positive illusions. So I'm not convinced that such beliefs (illusory or not) actually contribute causally to positive affect in more than a very temporary way. Conversely, many happy people know they're not intellectual stars, know they're not the best drivers, and know their children may even be below average. So such beliefs are clearly not *needed* for happiness.

One might respond to these points by saying that although one positive belief or so might not yield happiness in any sense, having many positive beliefs in several domains would. But if self-deception were to be the means to *that* state of affairs, one would have to be committed to such a range of self-deceptions that it would put one in the category of delusional. Furthermore, a number of findings in psychology suggest that when high self-esteem is inflated or unfounded, individuals respond to perceived threats to their self-esteem with hostility.[17] A disposition toward hostility characterizes someone who is the *opposite* of sanguine. So self-esteem by self-deceptive inflation, which might really just be *acting* as if one has self-esteem, seems (again) a bad route to positive affect.

What I think is going on in the empirical studies that Taylor cites is that nondepressed people have a systematic tendency simply to overestimate various aspects of personal worth. But a likely interpretation of this finding is that the positive affect causes the overestimation (and not, as Taylor holds, vice versa). What's the argument for this? Simply, if beliefs that attribute positive features cause positive affect, we'd expect not to see so many intelligent or good-looking or successful people who are unhappy. But there are many such people. But if positive affect causes self-flattering beliefs, then it's possible for one to have beliefs that attribute positive features to oneself without having positive affect, where these beliefs arrived by another route. Furthermore, we'd still expect to see a correlation between positive affect and self-flattering belief, which is precisely what we find.[18]

---

[17]Roy F. Baumeister, Laura Smart, and Joseph M. Boden, "Relation of Threatened Egotism to Violence and Aggression: The Dark Side of High Self-Esteem," *Psychological Review* 103 (1996): 5-33; Edward A. Johnson, Norah Vincent, and Leah Ross, "Self-Deception versus Self-Esteem in Buffering the Negative Effects of Failure," *Journal of Research in Personality* 31 (1997): 385-405.

[18]One thing that may be going on in the correlation between positive affect and overly

Try this. Give the question of how good a driver you are serious consideration. Come up with an estimate in terms of percentile of the population on the basis of your driving history—accidents, tickets, near misses, and success in pulling into parking spaces. Knowing what you know about people's tendencies to overestimate their driving abilities, you should probably lower the estimation you just made by about 15 percentile points. Now face it, *that's* your driving ability. Does this realization make you any less happy? When I do this, I end up in the 50th percentile (initial estimate 65th, then adjust for bias). The realization that I'm not as good a driver as I thought has no effect whatsoever on my positive affect. I doubt it will for you either. I take this to be evidence that overestimation of one's abilities doesn't have much influence on affect either way.

But it is not simply wrong that beliefs about the self with positive contents are needed to have positive affect. Depressed people are known for having repetitive thoughts with contents like *I'm a loser* and to dwell on their faults.[19] What I wish to propose here is that belief in one's worth as a human being is necessary for happiness. But this is a different kind of belief from beliefs with contents like *I am above-average popular* or *I am above-average good-looking*. It is even problematic to call the awareness of one's own worth as a human a belief. It is certainly not a belief with factual contents over which inferences are drawn of the sort discussed in contemporary epistemology. Let's simply refer to what I'm talking about as *sense of self-worth*.

Do the positive illusions that Taylor discusses contribute to a sense of self-worth? Here again equivocation on "positive illusion" gives a misleading impression. There is nothing illusory in a normal human being's having a sense of self-worth. This is not to say that people lack faults, but simply that human beings do have worth and are justified in believing so. If it were possible for slugs to have a sense of self-worth, it would perhaps be illusory for them to have it, but it is not illusory for humans—no matter how average or below average in looks, intelligence, or driving ability. Nor do illusions on such issues plausibly contribute to a sense of self-worth.

* * *

Let me summarize the critical points made thus far. Although there is empirical evidence that (many of) the "illusions" that Taylor studies are

---

positive views concerning the self is that the agent is *interpreting* her positive mood. If one feels a certain way, one seeks to explain that feeling, which causes one to look around and find positive or negative propositions about the self that may be believed.

[19]For an example study, see Jennifer A. Harrington and Virginia Blankenship, "Ruminative Thoughts and Their Relation to Depression and Anxiety," *Journal of Applied Social Psychology* 32 (2002): 465-85.

correlated with positive affect—and thus happiness—her use of "illusion" is fraught with equivocation. An "illusion" one can succeed is no illusion, if it leads to success. Such a doxastic state is better labeled something like *true belief in one's abilities* or, more generally, *self-fulfilling belief*. Second, *imagining* future outcomes for the sake of planning or giving oneself an orientation is not illusion. Third, a sense of self-worth, which does contribute to choiceworthy happiness, is not illusory, since all humans have worth. So the illusions Taylor discusses *that actually are illusions* have not been shown to be worthy contributors to positive affect or, more seriously, to choiceworthy happiness.

4. Three Arguments that Self-Deception Does Not Produce Happiness

I've argued that the empirical evidence that prima facie favors a policy of self-deception for the sake of happiness does not actually support that policy. It is a mistake to conflate positive illusions with a policy of self-deception, and the illusions themselves haven't been shown to do anything for happiness. Support for the policy has thus been removed. Now I'll offer arguments for rejecting it altogether.

4.1. What does self-deception do to desire satisfaction?

Having true beliefs enables people to accomplish things. I need to believe what my options are before I can choose among them. Having beliefs about what the outcomes of chosen options will be guides choice. If any of the beliefs mentioned in this general schema is false—either about the options or outcomes—then I'm likely to end up dissatisfied. So false belief tends toward dissatisfaction. But self-deception leads to false belief. So self-deception tends toward dissatisfaction. Dissatisfaction tends away from happiness (both Matrix and choiceworthy). So self-deception tends away from happiness. A person wishing to promote her happiness, therefore, would be mistaken to pursue a policy of self-deception for that end.

This line of reasoning seems so basic that I find it remarkable that anyone could champion self-deception as a means to happiness at all. How might that champion of self-deception respond?

The champion of self-deception notes, first of all, that there are some things in life over which one has no control, but which affect happiness nonetheless. One can't control one's innate components of looks or intelligence, for example. So the general recipe for self-deception to promote happiness, so the line goes, involves not self-deception about those things one can control, but about those things one can't control. The idea is that since one is stuck in a given position, there's no point in having

true beliefs about it, especially if they make one feel bad. So argues the champion of self-deception.

Three points, however, are noteworthy. First, we're out of the territory of vying for choiceworthy happiness, so the champion of self-deception who takes this line has resolved to settle for Matrix happiness, at least locally. Furthermore, the Matrix happiness is unlikely to last, since evidence of the reality one is denying is likely to come in at some point. Second, even though one can't change the innate components of intelligence or looks (or whatever), having false beliefs on these matters can still undermine desire satisfaction. Oftentimes, knowing one's limitations is essential to making the optimal choices in compensating for them. Third—and this point is often overlooked—self-deception also undermines one's ability to know *which* are the situations one has control over and can make improvements on and which aren't. In short, self-deception (assuming it could be useful at all in the short term) undermines cognition of the conditions under which it could (supposedly) be useful. Let's go over each of these points in more detail.

4.1.1. Self-deceptive Matrix happiness is temporary

Suppose that the aspects of one's appearance that one feels bad about are not possible to alter through exercise or other means. And suppose that deceiving oneself plausibly could make one feel better about those aspects. (I shall argue in section 5 that self-deception is unlikely to succeed even at this, but let's grant the champion of self-deception this point for now.) All that is accomplished by this self-deception is a slight contribution to global Matrix happiness, without a genuine increase in external goods. There is certainly no increase in attractive appearance, and, given that, there won't be any increase in the other external goods one might hope to obtain as a consequence of being good-looking. Thus, the Matrix happiness is likely to be intruded upon. Suppose the practice of self-deception were successful. One is likely to face perpetual disappointment and alarm at not being treated as good-looking, given the self-deceptive belief that one is. Thus, this is the kind of Matrix happiness that tends in the direction of its own undoing. And the disappointment at not being treated as good-looking may even be *greater* than if one were self-honest, given the *expectations* set up by the self-deceptive belief.

The champion of self-deception might respond that the extra boost of self-confidence one gets from self-deceptive belief in one's good looks will have *other* positive consequences, such as being more relaxed in social situations, which will arguably lead to being more liked and having more friends. But the claim is tenuous. There would need to be links between self-deceptive belief in one's good looks and self-confidence

and between the resulting self-confidence and social success. But it is not likely that the self-deceptive belief will actually produce the required self-confidence, given that the state of self-deception involves, at some level of cognition, information to the contrary of the self-deceptive belief in one's attractiveness. Furthermore, most people find those with inflated opinions of themselves off-putting, particularly when the inflated opinion is based on false belief that masks insecurity.[20] So both purported causal links (between the self-deceptive belief and the "self-confidence" and between the self-confidence and the social success) are dubious. In any case, why would one choose the tenuous strategy of improving social relations by self-deceptive belief that one is good-looking as opposed to something straightforward, like being friendly to people?

Analogous considerations apply to the elements of Matrix happiness that one might seek to construct on the basis of other self-deceptive beliefs, such as a self-deceptive belief that one is more popular than one actually is. Disappointment is likely.

4.1.2. More ways self-deception can undermine desire satisfaction

Recall that the champion of self-deception is arguing that there are situations to which the reasoning I started section 4 with doesn't apply. (The reasoning was that self-deception undermines happiness because false beliefs undermine one's ability to get what one wants. The response was that there are some situations one can't change, so one might as well be self-deceived about them to feel better.) My present point is that even if situation A can't be changed, having true beliefs about A might be needed for getting what one wants in situation B. So it doesn't follow that the fixedness of A makes self-deception about A immune to the problem of undermining desire satisfaction by causing false beliefs.

Suppose I am self-deceived that my son has great intellectual abilities, when he is in fact of quite ordinary intellect. The champion of self-deception says: "Why not be self-deceived? It might make you feel better. And you can't at this point change those portions of his DNA that influence native intelligence anyway." But, although one can't change a person's native intellectual gifts, having an accurate assessment of the

[20]One main upshot of the longitudinal and laboratory studies presented by Randall Colvin et al. is that individuals with a tendency toward self-enhancement (a form of self-deception) display behaviors that seem detrimental to positive social interactions, as judged by third-party observers. The authors' interpretation of why this is so meshes well with what I say here: "A deep albeit perhaps unrecognized and unacknowledged sense of uneasiness consequently may pervade the self-enhancer, hardly a condition conducive to mental health." C. Randall Colvin, Jack Block, and David C. Funder, "Overly Positive Self-Evaluations and Personality: Negative Implications for Mental Health," *Journal of Personality and Social Psychology* 68 (1995): 1152-62, p. 1161.

intellectual abilities of a child is needed for finding him the kind of help needed for his fullest possible development, given whatever his native abilities are. If I believe he is a natural at math, that might cause me to make two kinds of mistake. On the one hand, I might not get him the help he needs, which thwarts my desire for him to develop as much as possible; this is a sin of omission, resulting from the failure to have an accurate assessment. On the other hand, I, as a self-deceived father, might push to have him in a more difficult class than he can handle, which would result in great difficulty for him and much disappointment for me; this is a sin of commission, resulting from having an inaccurate assessment.

It's true that one can wrongly estimate a child's abilities without self-deception and make mistakes on how to raise the child on the basis of nonself-deceptive false beliefs. But the self-honest person is in a far better position to correct her mistakes than the self-deceiver, since the self-honest person has a commitment to updating beliefs on the basis of incoming information, while the self-deceiver must be committed to ignoring, discounting, or tendentiously reinterpreting incoming information. Thus, although one can't change a child's natural intellectual abilities, having an honest assessment of them is critical to making the right choices on further matters.

Perhaps another example is more favorable to the champion of self-deception. Suppose you have compelling evidence that by most evidential lights would get one to believe that your wife had an affair several years ago. Should you focus on the scanty evidence that she was faithful (she says she was) and ignore the mountain of evidence that she wasn't (emotional unavailability, lack of interest in sex, unexplained absences, frequent sightings of her with a male co-worker), and thereby, in the supposed interest of happiness, pursue a policy of self-deception on the matter? What good could having true beliefs on the matter bring about? It might seem that believing the truth in this case would be pointless, since the mooted event is far enough in the past, while believing falsely she was faithful could (it seems) alleviate your uneasiness on the matter. But good relationships require communication, and lack of honesty on one issue will lead to silence on many others. Inability to talk about something that happened in the past will lead to inability on her part to talk about the sorts of things that made the relationship problematic in the first place. In short, if you desire communication in the relationship, you should at least be open to the kind of information that self-deception would shut out. So again, self-deception is likely to undermine getting what you want.

There are two general points here: (i) true beliefs are needed for desire satisfaction often in ways that aren't obvious or predictable, and (ii) the shutting-off of information that's needed for self-deception under-

mines knowledge beyond the specific proposition about which one is self-deceived. The consequent *lack* of knowledge is also damaging to desire satisfaction as much as the specific false belief implicated in the self-deception.

4.1.3. How does the self-deceiver know what to be self-deceived about?

Finally, the champion of self-deception is now relying on the argument that *in certain situations* (those over which one has no control) self-deception can contribute to feeling better. The question is: How is the person pursuing a policy of self-deception to know the difference between situations the policy applies to and ones it doesn't apply to? Take again the case of the father self-deceived about his son's intelligence. On complicated matters, like the intelligence of a child, there will be ways in which one can contribute and be helpful, as well as aspects of the situation one cannot change. To know the difference between the aspects of the child's mind one can help and the aspects one cannot, one has to be responsive to the evidence the child provides of his abilities. A policy of self-deception is deliberately contrary to such responsiveness. One is likely to end up being self-deceived not just about the native abilities of the child, but also about abilities one could help improve. In short, self-deception, even on the assumption it ever *could* be helpful for happiness, undermines awareness of the conditions for its own helpfulness.

* * *

To sum up this section: Reasoning based on true beliefs is needed for desire satisfaction. Self-deception is contrary to true belief and hence tends away from desire satisfaction and happiness. Arguments in favor of self-deception ignore the damage that self-deception does to truthful cognition on matters of practical importance. Whether or not self-deception promotes Matrix happiness, it is contrary to choiceworthy happiness, since it undermines the attainment of genuine external goods.

I argue next that self-deception even undermines Matrix happiness.

5. The Internal State of the Self-Deceiver

Recall the distinction between the self-inflation bias and self-deception. Self-deception involves an internal tension in the mind between the weight of the evidence and the self-deceptive belief, while in self-inflation bias this need not be the case. Now, if you're to make a *policy* of self-deception for the sake of happiness, "self-deception"—not "self-inflation bias"—is the correct term for what you're doing. Self-inflation

bias is a general tendency not wedded to specific desires, while self-deception is by nature motivated by specific desires, like the desire for happiness. Having seen this, let's consider the internal mental state of the self-deceiver and its consequences for affect.

The cognitive state of the self-deceiver is, in a way, quite ordinary. Every ordinary cognizer has a large body of beliefs and stored information that she is not accessing for the purposes of choice and inference at any given time. So it is with the self-deceiver. What marks the difference between the ordinary cognizer and the self-deceiver is that the former, when a given piece of unattended knowledge becomes relevant, has no obstacles to attending to that knowledge. When you're hungry, representations of nearby restaurants cross your conscious mind; when you're tired, you naturally think of where the nearest bed or couch is; when your child has a cut, you recall where the band-aids are. The self-deceiver, on the other hand, resists considering unattended knowledge that runs contrary to the self-deceptive belief. It may seem remarkable that the self-deceiver can do this. And perhaps it is. But it is much more remarkable that attention works as well as it does in the first place. You have information on thousands of topics stored in long-term memory; it is quite mysterious how and by what process the right bits of information come into consciousness at the times when they're relevant.

Evidently, unattended beliefs, or unconscious bits of information stored in long-term memory, have a property that makes them salient to attention when their contents become relevant to the practical interests of the agent. Let's designate this property of unattended beliefs their *call*. I do not make any theoretical claims about what the call is like or whether it is a property intrinsic to each unattended belief or a property of the system as a whole. I only claim that such a property must exist if we are to account for the efficacy with which previously unattended information becomes conscious whenever it is relevant to a matter under consideration.

What happens in self-deception is that the call of unattended but relevant beliefs is ignored, and beliefs contrary to the agent's total evidence and epistemic norms are formed on the basis of what limited evidence is attended to. Attending to the call is avoided either because it promises immediate discomfort or because attending would be contrary to a self-deceptive policy, as in the sort of self-deception we're now considering.

Ignoring the call does not make it go away. When one is in a state of self-deception, there are two relevant features in play. First, the agent has evidence, stored in the form of unattended beliefs, that something is actually the case that's contrary to her desires or what she wishes to believe. Second, the call to attend to these beliefs is ignored.

To better understand what this psychic state is like, consider what happens when the first feature obtains while the second doesn't. That is,

states in which one has unattended beliefs that indicate something is true that's contrary to one's wishes, but one *does* go on to attend to the relevant (previously unattended) subject matters. We have several phrases to describe the combination of cognition and affect that ensues from this. One says, on accepting the truth of the undesirable consequence: "It's been weighing heavily on my mind." Or: "I've had a sneaking suspicion." Or: "I feared it was the case." Or: "It's been eating away at me." Each of these phrases indicates that a judgment has been formed. What's notable for present purposes, however, is that in each case the speaker seems to admit having been in some sort of unpleasant psychic state *prior* to the formation of the judgment—in other words, prior to answering the call of the unattended beliefs. Suppose now that one does not answer the call and thus leaves the beliefs unattended. It stands to reason that the unpleasant state of mind will simply continue. Far from producing a sanguine state of blissful Matrix happiness, therefore, self-deception is likely to result in a prolonged state of low-grade anxiety.[21]

The existence of this prolonged state of anxiety is further suggested the by phenomenon Freud observed known as *reaction formation*. Reaction formation is a tendency stridently to assert or advocate the contrary of what the agent suppresses (ignores the call of). For example, Henry Adams, Lester Wright, and Bethany Lohr have produced evidence that males who exhibit the greatest amounts of homophobia also exhibit greater levels of arousal in response to gay pornography than nonhomophobic males.[22] The present point is that the defensiveness and vitriol of the person who exhibits reaction formation point to an unpleasant psychological state, not a state of one who is relaxed or happy.

The champion of self-deception will say "ignorance is bliss." But regardless of whether that slogan is ever true, the person in a state of self-deception is not simply ignorant of the matter about which she is self-deceived. Unlike the merely ignorant, the self-deceiver has possession, at some level, of information that is "weighing on her mind" or "eating away at her," even if she won't admit it.[23]

To conclude this section, insofar as the mental state designated by such phrases as "eating away at" is unpleasant, self-deception results in an unpleasant mental state and thus tends away from Matrix happiness as

---

[21]See the previous footnote for a reference to outside support for this claim.

[22]Henry E. Adams, Lester W. Wright, Jr., and Bethany A. Lohr, "Is Homophobia Associated With Homosexual Arousal?" *Journal of Abnormal Psychology* 105 (1996): 440-45.

[23]There is also evidence that avoiding confrontation with unpleasant cognitions can be unhealthy bodily as well as mentally. See, for example, D.A. Weinberger, "Not Worrying Yourself Sick: The Health Consequences of Repressive Coping," paper presented at the 100th Annual Convention of the American Psychological Association, Washington, D.C., August 1992.

well as choiceworthy happiness. The champion of self-deception will respond that the psychic stress from admitting the unpleasant truth is greater than the unpleasantness of having the ignored information eat away at one. I doubt this is so. But it is an empirical question, and I know of no studies or metrics that weigh this precise trade-off. For my part, I've felt *relief* in the past upon being disabused of various self-deceptions. But in any case, ignoring information in one's mind that points in an unpleasant direction does indeed have a psychic toll. Given that, the rational thing is to attend to that information, as this gives the best prospects for figuring one's way out of a hard situation.

6. What Kind of Happiness?

There is a curious tension in the position of the champion of self-deception. On the one hand, she defends self-deception as a means to happiness by saying that deceiving ourselves into positive affect will give us the bright outlook that will positively influence the world around us. On the other hand, part of the strategy for defending self-deception has been to say that sometimes we are powerless to change a situation, and, given that, we might as well convince ourselves that the situation is better than it actually is. I have attempted to cast doubt on both of these positions in the course of this essay: (i) it's not clear that *illusory* positive self-inflation contributes to the right kind of positive affect; (ii) it's not clear that letting unpleasant evidence fester in the mind unattended is pleasanter than accepting the truth and moving on. But my point here is that both defenses of self-deception cannot be made for the *same* self-deception. The "positive affect changes the future" defense *presupposes* the agent has some sort of control over the surrounding environment on the matter about which she deceives herself. The "can't change the situation anyway" defense presupposes that the agent can't change the environment about which she deceives herself.

The question then is: In which kind of situation is self-deception conducive to improving happiness, situations in which there is control over the environment or situations in which there isn't?

The champion of self-deception will exuberantly say "both!" The champion of truth will say "neither!" The question here is whether there is any incoherence in saying "both." It seems to me that there is. The person who holds that it is both advisable to self-deceive in situations one has control over and in situations one has no control over adopts a policy of self-deceiving just about whenever there is something unpleasant in life. Now, such a committed self-deceiver trying to make a bad situation out to be better than it is will self-deceive into believing that she can positively *change* the bad aspects of the situation. It follows that the per-

son who says "both" is committed not just to self-deceiving in both kinds of situation, but also to confusing which type of situation she is in. But then what should the contents of her self-deceptions be? That the bad situation is good? Or that she can change it? The self-deception of the person who is in a situation she can't change is supposed to have contents that make the situation out to be better than it actually is, which would tend in the direction of *not* trying to change things. The self-deception of the person who is in a situation she *can* change for the better is supposed to make the situation out to be better than it is in terms of its potential for being positively influenced, that is, changed. The former self-deception tends in the direction of idleness; the latter tends in the direction of action. But the person committed to the utility of self-deception in *both* situation types, we have seen, undermines her ability to tell which situation type she is in and hence which sort of self-deceptive contents she should pursue. Her policy has become incoherent. She is a fly in the glass jar of her own self-deceptions.

It remains to the champion of self-deception to retreat to the position that self-deception is useful in one situation type or the other. In the next section, I'll argue that any benefit that might come from self-deception in situations that one has some control over can be better achieved through *honest imagining*. So I'll take up here one last time the position that self-deception might be of benefit in those situations over which one has no control.

Let's assume for the time being that self-deception-produced Matrix happiness is the best I can hope for if I'm in a bad situation I can't change. Suppose I'm in prison, with rough fellow inmates, no library, and bad food. It may be that self-deceptively convincing myself of the goodness of this situation will make me feel good temporarily. But this will be an unstable state, constantly intruded upon by confrontation with the actual inmates, lack of resources, and food. It follows that if there is a situation one is in about which it makes sense at all to self-deceive, the self-deception will be an ongoing endeavor. Furthermore, the belief that the situation is better than it is sets one up with a host of expectations, which are sure to be disappointed. So the Matrix happiness that ensues from self-deception will have at least three undesirable properties:

(1) It will be frequently interrupted by facts of the situation.
(2) It will be hard to maintain.
(3) It will be accompanied by disappointment one could have avoided were it not for the self-deceptive belief.

The question immediately arises: is this sort of "happiness" worth pursuing even in situations over which one has no control? I conjecture it is

not. It is, again, an empirical question—one which we may not even have the concepts to investigate properly—whether the psychic toll of admitting the badness of one's situation is greater or less than the psychic toll of maintaining the uneasy self-deception needed for local and temporary Matrix happiness, coupled with its added disappointments that ensue from false expectations. But the present point is that the Matrix happiness ensuing from self-deception (in the only situation types in which it *could* make sense to self-deceive) is not a desirable thing in and of itself, irrespective of whether it compares well or badly with the self-honest option in unalterable bad circumstances.

So the policy of self-deception for the sake of happiness seems to have a chance of fairing better than self-honesty only in situations that are bad and in which one can't alter the circumstances, and even then the outcome of the policy is not that good. But the problem is that the policy of self-deception undermines one's ability to tell whether one is in the kind of situation for which self-deception might (dubiously) be appropriate. Thus, if one follows the policy, one might end up pursuing it in situations in which far better and rationally defensible options are available. It seems to me to follow that a policy of self-deception for the sake of happiness is a bad policy.

7. A Self-Honest Approach

In Shakespeare's *As You Like It*, the Duke Senior and his friends are driven from their court by the Duke's nefarious brother, Duke Frederick. They take up residence in the Forest of Ardenne, where the Duke Senior gives this speech:

DUKE SENIOR

Now, my co-mates and brothers in exile,
Hath not old custom made this life more sweet
Than that of painted pomp? Are not these woods
More free from peril than the envious court?
Here feel we not the penalty of Adam,
The seasons' difference, as the icy fang
And churlish chiding of the winter's wind,
Which, when it bites and blows upon my body,
Even till I shrink with cold, I smile and say,
"This is no flattery. These are counselors
That feelingly persuade me what I am."
Sweet are the uses of adversity,
Which, like the toad, ugly and venomous,
Wears yet a precious jewel in his head.
And this our life, exempt from public haunt,
Finds tongues in trees, books in the running brooks,
Sermons in stones, and good in everything.

Is the Duke deceiving himself? A simplistic reading would say that he is, that he is trying to convince himself that a bad situation is good and is thereby at least attempting to deceive himself. But notice that his thinking does not have the usual form of self-deception. He is not ignoring the cold that bites him, nor is he attempting to convince himself that it is less harsh or "not so bad," as the self-deceiver might. Rather, the Duke is fully cognizant of the harshness of his new environment and embraces it. Furthermore, what makes it sweet to him is that he can *learn* from the very adversity. This theme of embracing adversity in its role as a teacher ("sermons in stones") and thereby finding "good in everything" is in fact in direct opposition to the idea that self-deception brings happiness.

The Duke is attending to the reality around him and using his imagination to craft a path from the situation to a state of well-being. Nor is the path that his imagination constructs at all unrealistic. He *can* learn from his harsh environment, and is aided in doing so by not being encumbered by the many distractions of the court. In order to conceive the Duke's mindset as a state of self-deception, one must conflate two different kinds of state: (i) imagining a positive future outcome that can be constructed out of present adversity, and (ii) believing (self-deceptively) that the present circumstances are not adverse. (This is, of course, very much like the conflation I pointed out in Taylor's work earlier in this essay. Recall that she put daily list-making in a chapter entitled "Escape from Reality: Illusions in Everyday Life." It is as if Taylor doesn't have the attitude of imagining in her psychological ontology, and thus lumps together anything that the agent happens to represent in the category of *belief*, which would lead one to think of many imaginings as illusions.)

I propose that we call the Duke's state of mind *honest imagining*. This is a state with four components:

(1) One attends to all aspects of the situation one can that are of practical interest;
(2) One straightforwardly applies her usual epistemic standards to what is cognized via the process referred to in (1) in order to form beliefs;
(3) One uses imagination to envision one or more possible ways to be happy in the circumstances one acknowledges one is in via the beliefs referred to in (2);
(4) One uses imagination to envision a route from the present situation to one of the ways to be happy envisioned via the process referred to in (3).

I hold that whatever gains in choiceworthy happiness one might hope for out of self-deception can be far better gotten through honest imagining. Whatever advantages there are to the positive illusions of the sort

that Taylor discusses, they most likely accrue simply from the fact that the agent has bothered to represent something involving a good outcome. Representation helps the agent choose actions that will help bring the good outcome about. If one is self-deceived, one gets the representation of the positive outcome, but one is then crippled with respect to learning whatever else needs to be brought about or changed in order to achieve it. This is not an advantageous position to be in. If one has honest imagining, one has the advantages of representing the positive outcome, without the crippling effects of not being responsive to relevant incoming information.[24]

8. Conclusion

The kind of mistake Taylor makes at a theoretical level can be made at a practical individual level. In fact, this is what might lead one into a policy of self-deception in the first place. People often hold that they must *believe* a good outcome will come about in order for it to come about. This leads to self-deception about the indicators that suggest the good outcome won't happen. But there is both something right and something wrong about the dogma that one must believe a good outcome will happen. What's right is that one must have a representation in mind of the good outcome, otherwise one couldn't select actions that construct a path toward it. But what's wrong is the idea that the representation must be a belief.

Let's review the argumentation of this essay. The project has been to assess the view that a policy of self-deception would be a practically rational way to promote happiness. The idea is simply that if we paint things as better for ourselves than they actually are, we'll be happier. One might take further support for this stance from the research program in social psychology that suggests positive illusions are conducive to mental health. I have suggested that this line of reasoning rests on a number of confusions. Matrix happiness must be separated from choice-worthy happiness; self-deception must be separated from self-inflation

[24]There is a complication here that I won't go into for lack of space, but feel compelled at least to point out. One might ask: *How* extravagant or conservative should the contents of the imaginings referred to in steps 3 and 4 be? If the outcomes imagined are too improbable, they risk being not useful. If too conservative or easy to bring about, they risk not being a genuine path to happiness. Two things can be said. First, one should probably imagine a *range* of possibilities, from the more conservative to the more extravagant. This is one advantage that imagining has over belief: greater latitude in the range of normatively acceptable possible contents. Second, it is unlikely that there will be a systematic theory of how extravagant the imagining should be. Rather, the Aristotelian notion of the *mean* seems to apply. Thanks to Ville Paukkonen for raising this issue.

bias; illusion must be separated both from true belief in one's abilities and self-fulfilling prophesy, on the one hand, and from imagining, on the other. In addition to disentangling the confusions, I argued that the policy in question is: (i) unsupported by the actual empirical evidence there is, (ii) destructive to the satisfaction of desires and goals, (iii) self-undermining, (iv) likely to contribute to a state of mental unrest, and (v) inferior to the policy of honest imagining.

Why is the idea that self-deception can contribute to happiness so seductive? There seem to be three main reasons. It has an air of insightful-seeming paradoxicality which works like revelation on minds that fail to see the right distinctions. More importantly, however, the prevalence of self-deception leads people to assume that it must have *some* function. Hence Taylor calls the illusions she describes "adaptive," a term she uses in the sense of "useful." But it doesn't seem to me that self-deception is adaptive, either in the evolutionary sense or in the sense of being useful for everyday life. I have argued extensively elsewhere that the capacity for self-deception is a byproduct, or spandrel, of features of the human mind that are beneficial for reasons other than their being implicated in producing self-deception.[25] Whether or not this is true, the simple possibility that self-deception is a spandrel should alleviate the pressure we might feel to give it a functional interpretation. And even if we do give it a functional interpretation, we may still be quite wrong that the function is happiness. It is commonplace in evolutionary thinking about the mind that adapted features of cognition and volition often do not tend in the direction of happiness. The final reason why the idea I have attacked here is seductive is, I believe, laziness. It seems easier to change a cognition than to change reality, so we might as well self-deceive to become happy. But mismanagement arising from laziness often leads to more work in the long run than if one had been diligent in the first place; so it is with cognitive mismanagement, or self-deception. There is a mental toll to ignoring the call of relevant but unsettling information, and there is a toll to missing opportunities to interact positively with the external world one has thereby ignored. In short, self-deception won't make you happy.[26]

D.S. Neil Van Leeuwen
The Center for Humanities
Tufts University
neilvl@gmail.com

[25]Van Leeuwen, "Finite Rational Self-Deceivers" and "Spandrels of Self-Deception."

[26]I'd like to thank Wes Holliday, Alfred Mele, John Perry, and Allen Wood for their insightful comments on an earlier draft. I wrote this paper while supported by a fellowship from the Center for Cultural Analysis at Rutgers University.