What's Bad About Bad Faith?

Simon D. Feldman and Allan Hazlett

Forthcoming in the *European Journal of Philosophy*

Our contemporary common sense ethical framework is committed to an ideal of authenticity, according to which there is something bad about being inauthentic, and something good about authenticity. Call this the BASIC INTUITION about authenticity. There is a related, but distinct facet of the ideal of authenticity: we are critical of the inauthentic person, such that to call someone a phony, a fake, or a hypocrite is to accuse her of something, to offer some kind of censure of her or of her behavior, and to charge her with doing or having done something wrong. And we praise authenticity even when manifested in behavior that we would otherwise condemn; someone's bad behavior is vindicated for us on the grounds that, at least, it was authentic, that the vicious actor was "being herself." The common sense ideal of authenticity, therefore, also says that the inauthentic person is doing something wrong and is culpable for her inauthenticity. This is why the characterization of another as inauthentic takes the form of a charge or an accusation. Call this the CULPABILITY INTUITION about inauthenticity.

Supposing the BASIC INTUITION is right, what explains its truth? In what sense, exactly, is inauthenticity bad? On standard philosophical accounts of authenticity, inauthenticity is bad because of what the inauthentic person does *to herself*. Charles Taylor writes that (according to the common sense ideal of authenticity) being authentic is a "better or higher mode of life" (1991, p. 16) than is being inauthentic, and that to fail to live up to this ideal is to live a life that is "somehow wasted or unfulfilled." (p. 17) Developing his own account of inauthenticity, or what he calls "bad faith," Sartre maintains that bad faith is a species of self-deception: the inauthentic person does herself harm by denying her own freedom. (1956, p. 86-96) The common commitment here is that inauthenticity is bad *for the inauthentic person*, in the sense that inauthenticity is a threat to the goodness or quality of her life. We have in mind here the sense of quality or goodness evoked by Derek Parfit's question: "What would be best for someone, or would be most in this person's interests, or would make this person's life go, for him, as well as possible?" (1984, p. 493)

Consider, then:

> WELLBEING ACCOUNT: Inauthenticity is bad because it always detracts from the wellbeing of the person who is inauthentic; the authentic person would always be better off, all things considered, were she not inauthentic. More exactly, keeping everything else about the world fixed, any internal change in an agent in the direction of authenticity would always result in the agent being better off, all things considered.

We do not take authenticity and inauthenticity to be an all-or-nothing affair; a person can be more or less authentic or inauthentic. And authenticity and inauthenticity are plausibly relative to certain aspects of a person; a person is not authentic or inauthentic tout court, but rather authentic or inauthentic in such-and-such a way.

There are two related theses we will not consider here. First, it could be argued that inauthenticity is bad in the sense that, keeping everything else about the world fixed, any

internal change in an agent in the direction of authenticity would always result in the agent being better off *with respect to authenticity*. So, the idea would be, authenticity is good, as such, and inauthenticity bad, as such. But this would not capture Taylor's idea that an inauthentic life would be a life "wasted or unfulfilled," for it would leave open the possibility that authenticity is (as it were) a luxury good – something good, as such, but by no means necessary or typically needed for living a good life. Our arguments against the WELLBEING ACCOUNT, below, will be compatible with the claim that inauthenticity may sometimes detract from the goodness of a person's life.

Second, we could drop the "keeping everything else about the world fixed" qualification, maintaining that there is always *some possible world* such that an internal change in an agent in the direction of authenticity would result in the agent being better off, all things considered. This would be quite different from the WELLBEING ACCOUNT, as formulated: to say of an inauthentic person that there is a possible world in which she would be better off, were she authentic, is not the same as to say that she would be better off, were she authentic. This latter claim vindicates the idea that authenticity is an "ideal" – something we ought to seek. If I would always be better off, were I more authentic, then it is obvious in what sense authenticity is an ideal. But if this is not the case, and it is merely the case that there is always some possible world in which I would be better off, were I authentic, then the sense in which authenticity is an ideal vanishes. Compare: there is always some possible world in which I would be better off, all things considered, were I to wear a bow tie, namely, that possible world in which people are systematically and reliably rewarded for wearing bow ties. But this hardly suggests that wearing a bow tie is ideal, for me, given the way things actually are.

In §1 we consider the metaphysically suspicious claim that authenticity requires fidelity to one's "true self." In §2 and §3 we criticize two versions of the WELLBEING ACCOUNT, based on an intuitive account of inauthenticity as pretending to be other than you are (§2) and on the Sartrian idea that inauthenticity is a species of self-deception (§3). Moreover, we argue that even when inauthenticity (as it is conceived on these two accounts) is prolonged (i.e. spanning a significant portion of a person's life) or substantial (i.e. involving deep or fundamental aspects of a person), it is not always inimical to wellbeing. This is sufficient, we submit, to defeat the idea that an inauthentic life (so conceived) is "wasted or unfulfilled." In §4 we consider cases which suggest that, nevertheless, there is something bad about inauthenticity. We argue that appearances are deceiving: that these cases do not involve the badness of inauthenticity, as such. Finally, in §5 we reject the idea that the badness of inauthenticity is to be explained in terms of the wellbeing of the inauthentic person. We propose that inauthenticity is bad because the inauthentic person culpably runs the risk of deceiving other people. On our diagnosis, then, the BASIC INTUITION is vindicated by the fact that the inauthentic person objectionably runs the risk of treating others in morally problematic ways. And because we conceive of the badness of inauthenticity in moral terms, our account, unlike the WELLBEING ACCOUNT, makes sense of the CULPABILITY INTUITION.

1. INAUTHENTICITY AS INFIDELTY TO "TRUE SELF"

What is it for a person to be inauthentic? Consider what seems like a paradigm case of inauthenticity. Smith has always been a passionate baseball fan, having spent many Sunday afternoons at Fenway Park with her family as a child. She regularly attends games, wears

Red Sox t-shirts, orders special baseball cable channels, and so on. But when Smith moves from Providence to Boston, to work at a new job, she finds that her new colleagues don't share her interest in baseball. They are, however, obsessed with basketball, and they repeatedly invite her to join them at games, to go out to watch the games on television, to talk about basketball at lunch, and so on. Smith, wanting to fit in with her new colleagues, decides to fake it: she buys some Celtics gear and joins her colleagues, pretending to like basketball, and her relationships with her new friends flourish.

What makes this a paradigm case of inauthenticity? A natural way to articulate the notion of authenticity is to say that authenticity is being true to yourself, so that being inauthentic involves some kind of betrayal of oneself. But what do *these* notions amount to? What does it mean for a person to be true, or false, to herself?

One kind of answer to this question is offered by Taylor:

> Herder put forward the idea that each of us has an original way of being human. Each person has his or her own "measure" is his way of putting it. […] There is a certain way of being human that is *my* way. I am called upon to live my life in this way, and not in imitation of anyone else's. [T]his gives new importance to being true to myself. If I am not, I miss the point of my life, I miss what being human is for *me*. (1991, pp. 28-9)

In the language of contemporary popular psychology, each person has a "true self," which is something (or someone?) distinct from her. A person's behavior may or may not manifest her "true self," but her life will be "somehow wasted or unfulfilled" (as Taylor puts it) unless she is faithful to it. Taylor goes on to speak of the importance of "a kind of contact with myself, with my own inner nature." (p. 29) So authenticity amounts to fidelity to my "inner nature" or "true self."

But there are decisive metaphysical problems with this account. If people have "true selves" or "inner natures," these can't be capable of changing. Suppose Smith's story ends like this: After a few months of watching basketball games with her new friends and after her friends explain a few things about the game, Smith finds herself enjoying it, and appreciating aspects of basketball she hadn't understood before; eventually, she becomes as genuinely passionate about basketball as any life-long Celtics fan. The natural thing to say about this is *not* that Smith "betrayed her true self" (who loved baseball), nor is it to say that she finally "found her true self" (who loves basketball). The natural thing to say is that *Smith changed*. The defender of the idea of an "inner nature" or "true self" faces the question: which of Smith's selves was her "true self"? And what makes that version of Smith the "true" one, and the other one "false"?

It's not plausible to say that the "external" origin of Smith's interest in basketball is the culprit. If we change the story so that Smith's overbearing father is responsible for her previous love of baseball (she only went to the games to please him, but eventually found herself enjoying them), shall we then say that she was betraying her "true self" even before her meeting her new friends? And if an origin like that is sufficient for some aspect of a person not being part of her "true self," what elements of our identities could be said to be authentic? Our loves and hates, our interests, and our values always have "external" origins.

What we care about is the result of a variety of diverse partial causes: our interactions with other people, growing up and living in a certain community or communities, our interactions with various social and political institutions, facts about our physiologies, whether idiosyncratic or resulting from our species-specific biological natures, and so on. So an "external" origin can't explain why Smith's concern for basketball isn't part of her "true self." And, in general, being changed and influenced by interacting with other people is a normal and unobjectionable part of most human lives.

For these reasons, we are reluctant to take talk of "true selves" or "inner natures" seriously. If talk of "inner natures" and "true selves" is meant as metaphor, that is fine, but we need a non-metaphorical articulation of what being authentic amounts to, if we are to seriously consider whether or not inauthenticity always detracts from a person's wellbeing.

2. INAUTHENTICITY AS PRETENDING

Although it is obscure what it would mean to say that Smith's "true self" prefers baseball to basketball, there is a non-obscure truth in the neighborhood, namely, that *Smith* prefers baseball to basketball. In what way does she betray herself, then, by pretending to like basketball? Perhaps her inauthenticity amounts to nothing more than that: to the fact that she is pretending. In the passage from Taylor, quoted above, we focused on the appeal to "inner natures." But perhaps the emphasis should have been on the idea that authenticity involves avoiding imitating other people. Consider the following "paradigm case of inauthenticity," from David Velleman:

> This person laughs at what he thinks he is supposed to find amusing, shows concern for what he thinks he is supposed to care about, and in general conforms himself to the demands and expectations of others. The motives that his behavior is designed to simulate are motives that he doesn't genuinely have. (2002, p. 97)

As Lionel Trilling puts it, authenticity requires "a congruence between avowal and actual feeling." (1972, p. 2, see also p. 27) Let's conceive of inauthenticity, then, as pretending to be someone or something other than you actually are. Given this conception of inauthenticity, is it the case that inauthenticity always detracts from wellbeing, in the sense that any internal change in an agent in the direction of authenticity would always result in the agent being better off, all things considered?

We think not. Consider Jones, a detective playing the "bad cop" role in an interrogation. Jones glares angrily at the suspect, she slaps a cigarette out of his mouth, she declares that she "doesn't have time for scum like you," and storms out of the room. The motives that Jones' behavior is designed to simulate are not motives that Jones genuinely has; Jones suffers from a lack of congruence between avowal and actual feeling. She is pretending to be other than she is, and thus is inauthentic, with respect of the interrogation. But Jones seems no worse off for her inauthenticity. Or, at least, we can easily imagine that she is no worse off: if she enjoys playing the "bad cop," or if she just doesn't care about it and sees it as part of her job.

Might it be the case that prolonged pretending – pretending that takes up a significant portion of your life – always detracts from the goodness of your life? Or that substantial pretending – pretending about some deep or fundamental aspect of yourself – always detracts from the goodness of your life?

Is so, then people who have certain careers – undercover police officers, politicians, actors, café waiters – always live worse lives, in virtue of having those careers. All these jobs involve prolonged pretending: pretending to be someone else, pretending to have feelings you don't have, pretending to be happy when you aren't, and so on. And in many cases this pretending will be about deep or fundamental aspects of the person who engages in it. The professional actor may repeatedly enact a character entirely unlike her, with different deep values, with a completely different personality and character traits, and so on. An undercover cop might identify herself most deeply as an officer of the law, but spend most of her time pretending to be a criminal.

It would be a mistake to accept the conclusion that people who have these careers are always worse off on account of the pretending involved. Prolonged and substantial pretending is a normal and unobjectionable part of many human lives.[1]

This doesn't mean that there aren't cases in which pretending significantly detracts from the goodness of a person's life. Consider Jews passing as Christians during the Inquisition or the Holocaust. It could be maintained that no good life involves *this* kind of pretending; that these Jews were deprived of the chance of living a good life, by being forced or coerced into pretending to be Christians. This kind of pretending might also detract from the goodness of a person's life in virtue of a powerful and unpleasant feeling of alienation that might go with it.[2] Our point is that pretending, *as such,* is not incompatible with living a good life. For someone coerced by threat of violence into pretending not to be Jewish, what threatens the goodness of her life is not her inauthenticity, but that fact that her inauthenticity and the accompanying feeling of alienation were the result of being coerced by threat of violence. And being coerced by threat of violence to do anything is a threat to the goodness of a person's life. (We return to related issues in §4.)

Smith's inauthenticity seems to threaten the goodness of her life precisely if we add details to her story that make it look more like a case of forced pretending. If she feels powerless to resist the "peer pressure" to become a basketball fan, or if she finds the prospect of not getting on with her colleagues so terrifying that she reluctantly accepts the only slightly-less horrible option of pretending to enjoy basketball, then her life is not going well. But, as above, in that case, the threat to Smith's wellbeing is not coming from her inauthenticity, as such, but from the fact that she is being forced to be inauthentic.

[1] Note that our argument doesn't depend on the idea that pretending is required for many (wage-earning) jobs. Someone could unproblematically pursue acting as a creative or artistic project, not merely as a means to earn a living, but as a meaningful ground project.

[2] Given this, we can distinguish between being painfully alienated from the pretending involved in one's job (which may threaten the goodness of one's life) and simply pretending, as part of one's job (which does not). But see §4 for cases of virtuous alienation.

On the other hand, Smith's inauthenticity seems not to threaten the goodness of her life at all if we add details that reduce or eliminate the sense of coercion. If she consciously recognizes that she prefers friendship to friendless baseball fandom and deliberately sets out to make friends by feigning an interest in basketball; if she thinks of her pretending as an opportunity for desired change; or if she is open to learning to eventually like basketball, and so on, then her inauthenticity does not detract from the goodness of her life. Indeed, pretending may be an essential part of the life-expanding experimentation that is characteristic of the very best lives. It seems virtuous to try things we are unsure about (or even downright suspicious of). And when we do, we will often commit to them in our avowals or our behavior, if not (yet) in our feelings. We call this "going for it" and we praise people for their open-mindedness and their adventurous spirit when they do.[3]

Pretending may also be an unavoidable part of becoming who we are. When a college freshman changes her habits of dress in the process of finding a comfortable social place, her high school friends may accuse her of being a phony (and she may even feel as if she is faking it); but we can far more charitably describe her as growing up, as figuring out who she is (as believers in "true-self" metaphysics might put it), or simply as changing over time. Such change, resulting from experimentation, is familiar, typical, and unobjectionable.[4]

Given all this, we conclude that the WELLBEING ACCOUNT fails, if we conceive of inauthenticity as pretending. We sympathetically return to this conception of authenticity, but not to the WELLBEING ACCOUNT, in §5.

3. INAUTHENTICITY AS SARTRIAN BAD FAITH

In §2, being inauthentic was conceived of as a way of acting or behaving; on the present conception, being inauthentic is a way of thinking, and in particular a way of thinking about oneself.[5] In a well-known passage from *Being and Nothingness*, Sartre describes what he takes to be a paradigm case of "bad faith": the case of a woman (called, in the literature, alternately a "flirt" and a "coquette") whom we'll call Michelle. Michelle "has consented to go out with a particular man for the first time."

> She knows very well the intentions which the man … cherishes regarding her. She knows also that it will be necessary sooner or later for her to make a decision. But she does not want to realize the urgency; she concerns herself

---

[3] See §5 where we offer a related comment on Sartre's ambivalent young woman on a date. If she is required to be wholehearted (or at least to behave as if she is), the call to authenticity may itself turn out to require a certain kind of pretending.

[4] Marcia Baron says something similar about self-deception: it is often best to "stumble along confusedly … settling on *some* plan of action and *some* picture or other of what is going on, in an attempt to end the painful confusion and unease and get on with life. In fact, it is not that easy to figure out what we are about, discovering it may at times have to arise *through* self-deception." (1988, p. 441)

[5] Though see Gardner 1993 for a broader conception of bad faith as not merely mental misrepresentation of self (in self-deception) but as a form of behavior that involves "self-misexpression." (p. 37) Our account of the badness of inauthenticity, in §5, is compatible with this conception of inauthenticity as falsely expressive behavior.

> only with what is respectful and discreet in the attitude of her companion. […] But then suppose he takes her hand. This act of her companion risks changing the situation by calling for an immediate decision. To leave her hand there is to consent in herself to flirt, to engage herself. To withdraw it is to break the troubled and unstable harmony which gives the hour its charm. The aim is to postpone the moment of decision as long as possible. We know what happens next; the young woman leaves her hand there, but she *does not notice* that she is leaving it. (1956, p. 96-7)

Sartre maintains that Michelle engages here in a particular species of self-deception. (In §5 we shall offer an alternative account of what Michelle's inauthenticity consists in.) He develops this account in the context of a philosophical framework with which we have substantial sympathy. Sartre rejects the idea of a "true self" and of an antecedently given "unique way of being human" for each of us, and recognizes the difficulty in giving a metaphysically adequate account of what it would be for a person to be authentic or inauthentic. He proposes that bad faith amounts to self-deception about one's human condition – self-deception about the kind of being that one is. He says, of Michelle:

> [W]hile sensing profoundly the presence of her own body … she realizes herself as *not being* her own body, and she contemplates it as though from above as a passive object to which events can *happen* but which can neither provoke them nor avoid them[.] The basic concept which is thus engendered utilizes the double property of the human being, who is at once a *facticity* and a *transcendence*. These two aspects of human reality are and ought to be capable of a valid coordination. But bad faith does not wish either to coordinate them or to surmount them in a synthesis. (1956, p. 98)

Michelle's inauthenticity amounts to her refusal to recognize her own status as a free agent (a "transcendence"); she treats herself as though she were merely an inanimate, unthinking object (a "facticity"), thereby refusing to face up to the necessity of making a choice about whether to respond positively or negatively to her date's advances. Inauthenticity is a "lie to oneself" (p. 89) about one's own freedom; it involves "hiding something from oneself" (p. 94), and it is from one's own freedom that one hides.[6] Summing up this existentialist account of inauthenticity, Beauvoir writes that the person in bad faith "cannot efface the agonizing evidence of his freedom." (1948, p. 45) Authenticity, then, would amount to avoiding this species of self-deception, and wholeheartedly accepting one's status as both a facticity and a transcendence.[7]

---

[6] This is the most common form of bad faith: fleeing from one's transcendence. (See, for example, Detmer 1986, pp. 23-4, Gardner 1993, pp. 36-7, Santoni 1995, pp. 29-34, and Kruks 2001, pp. 68-9) We ignore here a second, less common form of bad faith: fleeing from one's facticity. (See, for example, Flynn 2003, §4.) Thanks to John Davenport for this point.

[7] Although Sartre is far from clear on this. See his 1956, pp. 112-6, and Santoni 1993, pp. 89-109, and Catalano 1993.

Given this conception of inauthenticity, is the WELLBEING ACCOUNT plausible? In other words, does Sartrian bad faith always detract from wellbeing?[8]

Many discussions of bad faith, including Sartre's, have focused on the question of whether and how self-deception is possible (Wood 1988, Catalano 1990, Santoni 1995, pp. 28-46). Others have focused on the question of whether and to what extent bad faith can be avoided (Bernstein 1971, pp. 149-52, Detmer 1986, pp. 102-31, Catalano 1993, Santoni 1995). We assume self-deception is possible, and take no position on whether it can be avoided; our question concerns whether bad faith always detracts from wellbeing (cf. Baron 1988).

3.1 Bad faith as self-deception

One way of defending such a view would be to see it as a consequence of the fact that self-deception always detracts from wellbeing. Since Sartrian bad faith is a species of self-deception, it would follow that Sartrian bad faith always detracts from wellbeing.[9] This is suggested, perhaps, when Sartre writes (in "Existentialism is a Humanism") that:

> [W]e may judge (and this may be a logical rather than a value judgment) that certain choices are based on error and others on truth. We may also judge a man when we assert that he is acting in bad faith. [I] do not pass moral judgment against him, but I call his bad faith an error. Here, we cannot avoid making a judgment of truth. (2007, p. 47)

This would place a Sartrian WELLBEING ACCOUNT within the scope of the traditional view in modern psychology and sociology that, as Taylor puts it, "the ideally strong character … would be able to face unflinchingly the truth about himself or herself." (1989, p. 33) As David Jopling notes, Sartre not only criticizes self-deception about our human condition, but also praises self-knowledge about "the whole of our way of life, individual life history, or basic moral framework." (1992, p. 121) On Jopling's reading, a basic ethical project, for Sartre, is to bring to light "the always presupposed background or horizon of our life experience," and to become aware of "what we are really up to, and who we are." (Ibid.)

---

[8] We leave open the question of whether Sartre or Beauvoir endorses the WELLBEING ACCOUNT. Their overwhelmingly negative descriptions of people in bad faith suggest that they take bad faith to detract from wellbeing. Consider the fact that our facticity and our transcendence, which are not validly coordinated in bad faith, "*ought* to be capable of a valid coordination" (1956, p. 98, our emphasis), Beauvoir's description of the inauthentic person as a "sub-man" that "arouses contempt" (Beauvoir 1948, p. 43 and passim), and Sartre's censure of those in bad faith as either "cowards" or "bastards." (2007, p. 49) However, this negative attitude towards bad faith is to be contrasted with Sartre's apparent contempt for sincerity and "good faith" (as forms of bad faith) and the suggestion that bad faith is an inescapable aspect of the human condition. (1956, pp. 112-6).

[9] This, too, if Catalano is right that "[f]or Satre, self-deception and bad faith are the same." (1993, p. 81) This suggestion would seem to apply only to "weak bad faith," in Catalano's terminology.

Thus perhaps the badness of bad faith derives simply from the fact that it involves a self-deceptive failure of the agent to know herself.[10]

This account of the badness of bad faith will not work, for there are cases in which self-deception does not detract from wellbeing. As Jopling notes, "[c]ertain kinds of self-ignorance and self-deception have a strong adaptive function." (Ibid.) Consider some examples from Marcia Baron:

> … the AIDS victim who deceives himself into believing that his chances of recovery are at least fifty-fifty; the Salvadoran who convinces himself that his sister escaped into the mountains or fled to Mexico, and was not, despite the evidence, tortured and killed by the army's death squads. It is difficult to find these instances of self-deception objectionable. (1988, p. 433)

Contemporary research by social psychologists strongly suggests both the ubiquity and the advantage of a "self-enhancement bias." Numerous studies have confirmed non-depressed subjects' tendencies to rate themselves as "above average" or "better than most," across a wide range of positive features and abilities: intelligence, leadership ability, competence at one's job, athletic prowess, and so on (Abrahamson and Alloy 1981, Campbell 1986). Subsequent studies found that non-depressed subjects reliably tended to overestimate their positive qualities, by comparing subject's self-assessments with assessments of third-party observers (Lewishohn et al. 1980, Conway and Ross 1984, John and Robins 1994). Similarly, non-depressed subjects tend to overestimate their degree of control over events in the world (Alloy and Abrahmson 1979) and to have unrealistically optimistic views of their futures (Weinstein 1980, Alloy and Ahrens 1987). (See Taylor and Brown 1988 and Brown and Dutton 1995 for reviews of this literature.) Self-deception seems to be involved in at least some of these cases. But if so, there is a reliable connection between self-deception and not being depressed. If that's right, then it is not plausible that self-deception always detracts from wellbeing.

Baron describes a plausible case of self-enhancement bias at work:

> [A] historian denied tenure at a second-rate school gets a job at a considerably inferior school. Imagine that he is good but not first-rate, and that he was judged by the tenure committee to be a far worse scholar and teacher than he is. In reaction to the tenure decision he exaggerates the gap between the assessment of him and what a fairer assessment would have yielded. He compares his work to that of major contemporary historians and judges his to be as good or better than most of their work. (1988, pp. 433-4)

On our view, the historian's self-deception does not necessarily detract from the goodness of his life. Furthermore, his irrational belief about himself may persist for much of this life, and it may concern deep or fundamental aspects of himself (i.e. the quality of his academic work may be a deep or fundamental aspect of himself), but nevertheless his life is not in virtue of his self-deception worse than it would otherwise be. The psychological research cited above

[10] Cf. Catalano's distinctly epistemic characterizations of bad faith as a refusal to be critical of one's beliefs (1993, p. 87) and as a failure to respond rationally to evidence (1990, p. 687-8).

suggests that self-deception of a certain kind is *systematic and long-lasting*, for most non-depressed subjects, and that the self-deception in question concerns our beliefs *about deep and fundamental aspects of ourselves* (our character traits, our abilities, our control over events in the world, our futures).[11]

Baron argues that self-deception is dangerous because it threatens to undermine our agency, just as deception of others, because it requires manipulation, threatens to undermine the agency of others. And since "self-deception can become a habit, a strategy one falls back on too often" (1988, p. 437), and since "self-deception often requires, for its efficacy, further self-deception" (Ibid.), it is particularly dangerous. If self-deception undermines agency, and bad faith is a species of self-deception, then might the WELLBEING ACCOUNT be correct after all? Two replies to this.

First, it's not obvious that enjoying a lower degree of agency, as a result of one's having undermined one's own agency, necessarily detracts from the goodness of a person's life. It indeed seems bad to be manipulated, when we consider cases in which someone is manipulated by another person. But it is bad to be manipulated by yourself? That seems to be the question we started with; so the appeal to the idea that self-deception undermines agency hasn't added anything.

Second, there are reasons to be skeptical of the idea that self-deception always threatens to "spread" in the way that Baron suggests. The thought seems to be that if a person believes p as a result of self-deception, she will be rationally required to believe various other things (e.g. things implied by p) or to revise beliefs of hers that conflict or fail to cohere with the belief that p, and thus the self-deception will "spread." But this assumes that the self-deceiver will do what is rationally required, once she acquires the belief that p. Why assume that someone who engages in an irrational strategy of belief formation (the initial self-deception) will subsequently transform into a rational believer? Indeed, two familiar aspects of beliefs based on this kind of irrationality are (i) that believers are remarkably resistant to evidence against the irrational belief, and (ii) that irrational beliefs of this kind become "compartmentalized," so that believers fail to draw conclusions from them and to incorporate them into other parts of their system of beliefs. The amateur basketball player who deceives himself into thinking that he could be competitive in the NBA doesn't conclude from this that he should go to NBA tryouts; he may even know that he wouldn't be competitive in the NBA, and thus have conflicting beliefs, but his irrationality may prevent him from noticing this or giving it any attention, and thus from its "spreading" in such a way as to undermine his agency or otherwise detract from his wellbeing.

[11] Are most people falling short of some ideal, however difficult to attain? Perhaps self-deception (and bad faith along with it) are inescapable features of the human condition, as Sartre suggests (1956, pp. 112-6), or perhaps escapable only for a heroic few, or with immense effort. (See Detmer 1986, pp. 102-31, and Santoni 1995, pp. 89-109.) Recall that we do not mean to attack the idea that there might be possible worlds in which authenticity is always better than inauthenticity, but just that the actual world is not like that. We leave the possibility open that escaping self-deception is some kind of ideal, but we find it implausible that something avoidable only by the heroic few could be something that makes the lives of the rest of us "wasted or unfulfilled."

3.2 Bad faith as a species of self-deception

So we conclude that Sartrian bad faith isn't bad because self-deception, in general, is bad. We must then try to understand the badness of bad faith as flowing from some feature specific to the species of self-deception Sartre describes (self-deception about one's human condition, e.g. one's status as a transcendence), and not some feature of self-deception in general. Is it plausible that deceiving myself about my status as transcendence, in particular, always detracts from my wellbeing?

There are two reasons to be suspicious of this claim. First, there are cases in which Sartrian bad faith doesn't intuitively detract from a person's wellbeing. Imagine that Katie's overbearing parents have planned for her to be a violinist since she was a small child, and have persistently pressured her into taking violin lessons, going to music camp, practicing four hours every day, and so on. Katie resents her parents, but at this point can't stand up to them: if she were to quit playing violin, they'd refuse to pay for college, disown her, or simply make things more unpleasant than she is willing to bear. Her only rebellion consists in this: in her own mind she refuses to accept responsibility for her violin playing, maintaining that she has no choice about the matter, and that her being a violin player is "just the way things are." Katie acquiesces to continuing being a violin player, until she's out of her parent's sphere of influence, but she does this by abdicating personal responsibility for her actions. Katie is in bad faith: she flees from her transcendence, reducing herself to a facticity – an unfree object, controlled by her parents, not capable of choosing otherwise, and not responsible for her situation.

But, we submit, Katie's bad faith doesn't detract from her wellbeing. It is not the case that she would be better off, all things considered, were she to move in the direction of authenticity – if, for example, she were to refuse to play violin (by her own lights, that alternative is worse than "putting up with it") or (as Sartre might counsel) if she were to accept responsibility for her situation, and admit that she freely chooses to play violin. Both of these options seem no better than the option Katie has chosen: to continue to play, but to ignore her status as a transcendence. For Katie's bad faith, given her situation, is her only way of manifesting her resentment of being pressured by her parents and of coping with their treatment of her.

Consider a variant on a case from Joseph Catalano (1993, p. 85): Marcus studies moderately hard and does exceptionally well on his first math exam, and comes to think of himself as a "math whiz." His newfound confidence in his own innate mathematical abilities leads him to believe that he can be the top student in his class. He studies even harder, convinced that he is "born to be a mathematician." When people compliment his efforts, he deflects their praise by saying, "I didn't choose my abilities, they chose me." Marcus is in bad faith: he flees from his transcendence, reducing himself to a facticity – a "math whiz" with "innate abilities." But does Marcus' life go worse, as a result of this? Suppose the alternative is that Marcus has no confidence that he can do well in mathematics; if he were not to believe that he is a natural "math whiz," he would do poorly, because he would lack the motivation to study and to approach his work with confidence. Marcus is, we may imagine, self-deceived: he thinks he is a natural "math whiz," when really his good performance is his own doing. But he is better off self-deceived, and in Sartrian bad faith, than he would otherwise be.

In Catalano's example (Ibid.), the student does poorly on his first exam, and develops a neurotic conviction that he is naturally bad at mathematics. This, of course, seems bad. In both Catalano's example and in the case of Marcus, the agent thinks of himself as (in one respect) a mere facticity – as "naturally" or "innately" being bad or good at mathematics. In Catalano's example, this leads to a bad outcome (the student's negative view of himself is a "self-fulfilling prophecy"), but in Marcus' case, it leads to a good outcome (perhaps because his self-conception is a virtuous form of "self-enhancement"). Bad faith, in other words, does not always detract from wellbeing.

The second reason we are suspicious of the claim that bad faith always detracts from wellbeing is that, while we are sympathetic with at least the spirit of Sartre's "double property" claim about the human condition, it seems implausible that refusing to accept this claim would always detract from one's wellbeing, since failure to recognize one's status as a transcendence (for example) is a philosophical mistake, at best. Katie and Marcus are wrong, we concede, about their freedom: she chooses to play the violin and is (at least in some sense) responsible for that fact about herself; he chooses to work hard at mathematics and is (at least in part) responsible for his good performance. But how could making this philosophical mistake necessarily detract from the goodness of one's life?

This question is relevant only because we have rejected the view, above, that self-deception in general always detracts from wellbeing. Since it does not, we now must wonder why this particular species of self-deception is problematic – since it is not the self-deception, as such, that presents the problem. The person in bad faith (in some sense) knows that she is a transcendence, since otherwise her bad faith would be mere self-ignorance, and not self-deception. But her failure (in some other sense) to know – or to appreciate, or to accept, or to be aware of – her status as a transcendence is said to be problematic bad faith. Our point is that, given that the error involved here a philosophical one, it would be surprising that anyone who is wrong about this fact, in particular, is always going to be worse off.

Is bad faith a mere philosophical mistake? Is it merely a philosophical mistake to think of oneself as a mere facticity, and to not appreciate one's transcendence? Some argue that bad faith is a more basic kind of mistake, based on Sartre's thesis of the "translucency" of consciousness (see, e.g., 1956, p. 11, p. 89, and p. 125). Thomas Anderson writes:

> In general, the individual in bad faith denies one of these dimensions of his or her reality and identifies his or her self with the other. But that is not the whole story, for what makes such denial bad faith, instead of ignorance or a simple mistake, is that the individual is lying to him- or herself. The person is lying precisely because *she cannot be unaware, even without reflection, of the twofold nature of his or her being*. [T]he person in bad faith lies to himself about the twofold dimension of human reality as both freedom and facticity at the same time that he is *prereflectively aware of being both*. (1993, p. 15, our emphasis)

We have conceded that bad faith is not mere ignorance – that it is a species of self-deception. But we seek an explanation of why this particular species of self-deception would be necessarily inimical to wellbeing. And we have objected to that idea on the grounds that bad faith seems to involve a mere philosophical mistake. But if Anderson is right, the mistake is not philosophical – for our transcendence (and our facticity) is something

obvious, something that no one can fail to be aware of. It is not something we come to know on reflection; it is something we are necessarily pre-reflectively aware of.

Is it really plausible, that our transcendence (i.e. the fact that we are free and undetermined beings) is something we are necessarily pre-reflectively aware of? If our freedom is so obvious, why have many seemingly honest philosophers denied it? Were they all cowards, fleeing from the burdens of responsibility? If not, then perhaps our freedom is not something so obvious but rather something that it takes some effort to see, and therefore something that one could honestly be mistaken about.

The matter is complicated by the fact that, as Catalano argues:

> Sartre's claim that consciousness is translucent does not imply that we always have a correct understanding of that of which we are aware, whether this be our own internal states or external objects. Translucency does not guarantee that I will always correctly conceptualize that of which I am aware. (1990, p. 680)

So we concede that the Sartrian can safely say that bad faith does not require that one accept any obscure metaphysical view (e.g. the ontology of *Being and Nothingness*), and that it does not require thinking of oneself under the description "mere facticity, and not transcendence." But if to be in bad faith is to deny something about one's human condition that is evident or obvious, then the fact of translucency (if it is a fact) does not obviate our objection, since (as Catalano points out) translucency does not guarantee that I will properly appreciate or understand everything of which I am pre-reflectively aware. One reason we have focused on the case of Michelle is that cases like hers seem to be those in which our radical freedom (if such we have) is least obvious to us. Cases like Michelle's – in which one's own desires are unclear, in which the social situation is complex, in which the other's expectations create various interpersonal pressures – are precisely the kinds of situations in which one's transcendence is not obvious to one – even if one is, in some sense, always aware of it. If Michelle is failing to recognize an important fact about herself, she is doing so in a context where that fact is not epistemically accessible to her. To hold her culpable, in these circumstances, for failing to recognize herself as a transcendence *does* seem to require that she hold a certain philosophical theory – that she have a certain way of responding to her epistemically opaque awareness of her own transcendence. Failure to meet these demanding requirements seems not to be the kind of thing that could always detract from the goodness of a person's life.

Given all this, it is safe to conclude that Sartrian bad faith may not detract from wellbeing even when it is prolonged or substantial. As argued above, self-enhancement bias, the illusion of control, and irrational optimism seem to be long-lasting and to concern deep or fundamental aspects of ourselves; and yet these species of self-deception do not plausibly detract from the goodness of our lives. Why think Sartrian bad faith is any different? Again, while perhaps avoiding bad faith – by recognizing and accepting one's status as both a facticity and a transcendence – is some kind of ideal (and certainly, if Sartre's ontology is right, this would be a philosophical ideal, in as much as it would be to believe a correct philosophical theory of the human condition), the WELLBEING ACCOUNT, if inauthenticity is Sartrian bad faith, is not defensible.

4. HARMFUL INAUTHENTICITY?

There seem to be cases in which inauthenticity detracts from the goodness of the inauthentic person's life.[12] Consider Mary, a woman who is ashamed of her sexual desires because of their orientation, which she believes, on religious grounds, to be sinful. As a result, Mary misses out on possibilities for sexual fulfillment. Or consider Joe, who has been raised to have a very particular sense of what it is to be a "real man," on which a "real man" is strong and unemotional. Joe has young children whom he loves dearly, but because of his machismo he is cold and distant towards them, and never cultivates the kind of warm and close relationship he would "really" like to have. Mary and Joe seem to be inauthentic, in some sense, and their inauthenticity, so the objection goes, is bad for them – it detracts from the goodness of their lives. Doesn't this suggest that that some version of the WELLBEING ACCOUNT must be correct?

We shall argue that what detracts from Mary and Joe's wellbeing is not their inauthenticity, as such (whatever precisely their inauthenticity amounts to, in these cases). This is either because (i) the feature that makes their lives go worse is not something that always detracts from wellbeing (even if it is plausibly called "inauthenticity"), or (ii) the feature that makes their lives go worse is not plausibly called "inauthenticity," but is rather some other problematic feature of their situations. As above, we do not claim that inauthenticity (of the sort involved in these cases) never detracts, as such, from the goodness of a person's life.

One possibility is that Mary and Joe are *emotionally repressed*. They do not let their emotions play their proper role in influencing their behavior. Mary avoids or passes up on opportunities for sexual fulfillment; Joe avoids or passes up on opportunities to connect with his kids. But consider the case (structurally analogous to Mary's) of a pedophile who feels ashamed of his sexual desires, and on this basis does not allow them to influence his behavior. There is a fine sense, perhaps, in which the pedophile's pain (at not following through on his sexual impulses) detracts from his wellbeing, but his repression of these desires is, if anything, a good thing – his is a better life, given his virtuous repression of his morally problematic desires (even if they are part of his sexual "nature").[13] We do not just mean that such repression is likely to have good consequences for others, although plausibly it will. We mean that such repression is better *for the pedophile*. This is at least plausible on the assumption that the pedophile himself rejects his sexual desires on the ground that they are immoral. He might reject them in virtue of their likely consequences for him; such desires move him in the direction of immoral conduct which he abhors. Or, equally plausibly, he might reject them in virtue of their content; such desires embody an attitude towards children that he recognizes as morally inappropriate, e.g. as a failure of owed respect, or as attitudes constitutive of vice.

---

[12] Thanks to an anonymous reviewer for suggesting these cases.

[13] We are not denying here that it might be best if the pedophile lost his pedophilic urges and was no longer alienated from his sexual feelings. The point is that, given the way the world is, the pedophile's emotional repression does not detract from the goodness of his life. The point is that what is wrong with Mary and the pedophile isn't merely that they are emotionally repressed; it is rather something else bad about their respective predicaments.

What this shows is that emotional repression does not always detract from wellbeing. So this way of understanding the cases does not support a WELLBEING ACCOUNT.

Even if emotional repression is prolonged or substantial, it may not always be worse to be repressed, than not. Again, the pedophile's immoral sexual desires may not go away (and thus prolonged repression will be preferable) and may be quite deep and fundamental vis-à-vis his identity (and thus substantial repression will be preferable). What this shows is that emotional repression, as such, does not always detract from wellbeing – whether repression is bad may depend on the individual, the content of the emotions being repressed (see the "third possibility," below), the historical origin of the repression (see the "fourth possibility"), and various other features of the situation.

A second possibility is that Mary and Joe are *alienated from their desires*. Mary experiences her sexuality as alien; Joe experiences his affection for his children as alien. But alienation is not always a bad thing. Consider Nomy Arpaly's account of Le Carré's Oliver Single (1999, pp. 3-8), who is alienated from his desire to betray his criminal father and defect to help the police in their efforts against the Single firm. At the crucial moment, "his arm reaches out for [the phone] unbidden, his hand grasps it and puts it to his ear, leaving him with the responsibility of what to say." (Le Carré, quoted in Arpaly 1999, p. 4. Note the similarity between Single's alienation and Michelle's bad faith.) Single is alienated from his desire to defect. Is this a bad thing? In other words, does it detract from his wellbeing?

We think not. What is going on, in Single's case, is that his values have come into conflict with one another: one the one side, he loves his father and is naturally loyal to his family, but, on the other side, he loathes and morally objects to the family's firm's deep affiliation with organized crime. On a natural reading, Single's alienation is a device he uses (as it were) to negotiate between these two sets of conflicting values. As Arpaly points out, Single's drinking prior to his defection is "something that he does to ease himself into the difficult task of defecting, which he unconsciously anticipates." (1999, p. 5) The same should be said about his feeling of alienation itself. Single just does not have the strength of will to wholeheartedly defect, sans alienation. If so, then he is better off alienated then he would otherwise be – free of alienation, but unable to do the right thing.

Moreover, there may even be a sense in which alienation is *ideal* for Single – for to defect wholeheartedly seems to be an even worse kind of betrayal of his family: a betrayal in thought, in addition to the betrayal in action. Someone whose values come into conflict and who must make a difficult decision *ought* to be conflicted. Alienation from one's eventual decision is, we submit, one way of giving the losing side of the argument its due. Single is alienated from his decision precisely because he loves his family – and that is not a bad thing.

Even if one's alienation is prolonged or substantial, it may not be a bad thing. Consider, again, the case of the pedophile. His immoral sexual desires are really his, they are as much a part of him as anything is. But his strategy for coping with them, for keeping them from influencing his behavior, could easily be to become alienated from them. Indeed, this is a familiar strategy from addiction recovery programs that treat addiction as a disease: the patient is to think of her desire for drugs, for example, as alien and external, as a

manifestation of "the Disease," but not of her true self.[14] If such strategies work, then alienation does not always detract from wellbeing.

We must keep in mind the distinction between something's being bad for one *as such* and something's being bad for one in virtue of its consequences. The pain that Mary and Joe feel, as a result of their alienation, surely detracts from their wellbeing. But this doesn't show that their inauthenticity itself (whatever it amounts to) is bad for them.

A third possibility is that that there is something objectionable about the *contents* of Mary's attitude towards sexuality and of Joe's attitude towards emotional sensitivity. This is plausible in Joe's case: the idea that a man ought to be unemotional is a morally pernicious attitude to have. Mary's sexual shame may be morally problematic (if, e.g., her view is that her desires are sinful because of their orientation) or perhaps it is "unhealthy" to have any negative attitude towards sex, as we are sometimes told by contemporary psychologists.

But if the problem with Mary and Joe is the content of their attitudes, then it is not their inauthenticity that detracts from the goodness of their lives. If the problem is the content of their attitudes, then the problem is not the conflict between attitude and feelings (which seems to be what constitutes inauthenticity in these cases). We can appreciate this point by noting that, if there is something wrong with the content of Mary and Joe's attitudes, then it would be just as bad for them to feel about *others* as they do about themselves. If it is bad for Mary to think ill of her own sexuality, isn't it also bad for her to think ill of the sexuality of others? If it is bad for Joe to reject his own sensitive inclinations on the grounds that they are effeminate, isn't it also bad for him to do the same with respect of to sensitive inclinations of others?

Thus what may be going on in these cases is that we are responding to the fact that Mary and Joe have immoral or unhealthy attitudes towards sexuality or emotional sensitivity. If so, our intuition that there is something bad about Mary and Joe's situations may just be a reaction to their bad attitudes towards sex and sensitivity, and not a reaction to their inauthenticity. (Consider, for example, our different reaction to cases of inauthenticity in which the attitude manifested is not immoral or unhealthy: the case of the pedophile, the case of Oliver Single, the case of the addict who loathes his desire for the drug, and so on.)

A fourth possibility is that Mary and Joe have been *coerced* (or manipulated or brainwashed) into betraying their true selves. If Mary's sexual shame, or Joe's machismo, in fact are the result of some kind of objectionable coercion, then we think that it is their having been coerced that has detracted from the goodness of their lives, and not their being inauthentic, as such. (Compare the case of the passing Jews, §2.) If, on the other hand, Mary's shame and Joe's machismo are not the result of coercion, but rather simply have an "external" origin (e.g. their respective upbringings), we do not think this is sufficient to render their situations problematic. As we argued in §1, our values always have "external" origins.

[14] A similar point applies when it comes to Sartrian bad faith (§3): the first step in a twelve-step program is to admit that you are powerless over your addiction – a case of fleeing from transcendence, if ever there was one. But it seems to us that such bad faith could obviously be the best course of action for a person, given relevant assumptions about that person, her addiction, the reliability of the twelve-step method, etc.

As we argued in the case of Smith (§2), our intuitions will shift depending on whether we describe Mary's as a case of coercion. If we say that her religious views are the result of harsh and oppressive indoctrination at the hands of a dogmatic and illiberal community of "fundamentalists," then we intuit that her shame detracts from the goodness of her life. But in this case, what is detracting from the goodness of Mary's life is her oppression by her community, not her inauthenticity, as such. On the other hand, if we say that her religious views are the result of years of careful and free inquiry, resulting eventually in a close reading of Augustine on which her sexuality is shameful, then the "problem" with Mary is that she has made a philosophical mistake (at least, we would argue that it is a mistake). There is certainly no problem of "inauthenticity" in such a case.[15]

It is not clear how we should understand the metaphysics of the "true self" in such cases. Even supposing her religious beliefs to be inherited from her community, by what principle should we treat Mary's sexual desires as reflecting her "true self," and her admittedly socially formed sexual values, and her corresponding shame, as alien? If the "external" origin of her shame amounts to its being explained by the fact that it corresponds to the values of her community, this seems insufficient to render it alien, and not a reflection of her "true self." Likewise with Joe. Why should we think that his sense of his own masculinity is "less him" than are his more sensitive inclinations?

Perhaps Mary and Joe are experiencing what W.E.B. Du Bois calls "double consciousness" (1907, p. 3), or what Sandra Bartky calls "psychological oppression." (1990, pp. 22-31) These are conditions experienced by those who have internalized morally problematic beliefs and attitudes about their own inferiority or moral worth. What is especially bad about double consciousness is not simply that it involves painful alienation, but that these feelings are manifestations and continuing expressions of pervasive and pernicious social inequalities. Additionally, the attitudes that those with double consciousness take toward themselves are themselves morally problematic: a woman who thinks that women are worthless adopts a morally problematic attitude towards women. These facts about double consciousness – that it is painful, that it manifests continuing oppression, and that its content is objectionable – explain why we judge that something terrible is happening in such cases. The inauthenticity of the individuals involved (whatever exactly it amounts to) is not the relevant factor.

So, to sum up: (i) if what goes wrong in Mary and Joe's cases is that they are emotionally repressed or alienated from their desires, then we reply that repression and alienation do not always detract from the goodness of a person's life. So the WELLBEING ACCOUNT is not plausible. On the other hand, (ii) if what goes wrong in Mary and Joe's cases is that their attitudes towards sex and sensitivity are immoral or unhealthy, or that their attitudes are the result of coercion or oppression, then we reply that the badness in these cases is not the result of inauthenticity, as such.

---

[15] Imagine, alternatively, that Mary is ashamed of her sexual orientation (she is attracted to aggressive, powerful men) because she considers it a manifestation of ongoing historical gender domination. Perhaps what prompts our intuition that Mary's life has gone badly, in the original version of the story, is the religious nature of her shame. But in that case, inauthenticity is not the issue.

We think there is a good explanation of what is wrong with Mary and Joe that does not appeal to what these they do *to themselves*. It is far more plausible, we think, to explain the badness of Mary's and Joe's situations in straightforwardly moral terms, i.e. in terms of what Mary and Joe do *to others*. If there is something wrong with Mary, it is that she is complicit in the propagation of sexual mores that harm people. And if there is something wrong with Joe, it is that he is not treating his children as well as he should. In the next section, we develop a similar account of the badness of inauthenticity, in general.

5. INAUTHENTICITY AS LYING

Recall Michelle, who lets her date hold her hand while, in some sense, "not noticing" that she's letting him. Note that the story requires that Michelle "does not quite know what she wants" (Sartre 1956, p. 97); this is what makes her reluctance to make a decision psychologically plausible. She can "not know what she wants" in either of two senses. On the one hand, she might simply be ignorant of certain facts which would settle things for her one way or the other: if he's a Laker fan, then that's a dealbreaker; she wholeheartedly wants to be romantically involved with someone who isn't a Laker fan, and her "not knowing what she wants" just consists in her not knowing whether this particular man is a Laker fan or not. On the other hand, she might have all the relevant information, but remain conflicted or ambivalent about whether she wants to be involved with the man she's on a date with. This will be the case if she has desires that conflict, either necessarily or because of contingent facts about this particular man: she's physically attracted to him, but finds him dim-witted, for example. Her "not knowing what she wants" here consists in her having desires that conflict.

"Not knowing what you want," in either of these senses, isn't intuitively a form of inauthenticity.[16] However, there *is* a sense in which Michelle is inauthentic in Sartre's story, but it's not a matter of her beliefs (e.g. her engaging in self-deception) or her desires (e.g. her "not knowing what she wants"). She is inauthentic when she doesn't withdraw her hand, for in doing this she continues her flirtation, and (so Sartre seems to think, anyway) thereby gives her companion the impression that she *does* "know what she wants," namely, to become romantically involved with him. In not withdrawing her hand, she gives the impression that she *has* made a decision, when in fact she has not. In acting as she does, she runs the risk of misleading or deceiving him.

There is something strange about Sartre's description of this case, which may get in the way of appreciating the point we're trying to make. Michelle is described as being under pressure to "make a decision" about something, a "necessary" decision, but about what, exactly? Whether to have sex with her companion? Whether to make some kind of avowal or confession of attraction? Whether to let him hold her hand? The latter seems the only decision Michelle is forced to make in the course of the story as Sartre tells it, and thus the only decision that seems in any obvious way "necessary." And what are the "intentions which the man … cherishes regarding her"? The intention to have sex with Michelle? The intention to take her out again? The intention to marry her? How does Michelle "know very well" what these intentions are? If she does know very well what the intentions of her

[16] There are, however, critics of ambivalence, e.g. Frankfurt 1992.

date are, she'd be the first person to ever enjoy such certain knowledge of the mind of a prospective romantic partner. Why are her companion's intentions presumed to be determinate and clear, when Michelle's own intentions are indeterminate and unclear?

However, our point is that Michelle does seem to run the risk of misleading or deceiving the man she's on a date with, if only because he unjustifiably expects some kind of assurance or "sign" that she feels a certain determinate way about him. Let's grant the dubious and problematic assumption that holding hands with someone *does* constitute an avowal of non-ambivalent romantic interest; there is, then, an incongruence between Michelle's avowal and her actual feelings, to borrow Trilling's phrase. And it is in virtue of this incongruence that Michelle runs the risk of misleading or deceiving her companion.

This suggests an alternative to the WELLBEING ACCOUNT of the badness of authenticity. On this alternative account, which we favor, inauthenticity is bad because the inauthentic person always runs the risk of misleading or deceiving other people. To the extent that this is morally problematic, being inauthentic is morally problematic as well. On this view, inauthenticity isn't bad because of how it affects the authentic person herself, it's bad in virtue of how it affects (or threatens to affect) other people.

So we propose to revive the idea, considered in §2, that inauthenticity essentially requires acting in a way not representative of one's actual beliefs, desires, and other attitudes. But rather than combining this with a WELLBEING ACCOUNT of the badness of inauthenticity, we propose a *moral* account of the badness of inauthenticity. There are two plausible versions of this account. According to the first:

> MORAL ACCOUNT, VERSION 1: Being inauthentic is morally problematic when, and only when, it results in someone being misled or deceived, and it is problematic to the extent, and only to the extent, that someone is misled or deceived.

Someone deserves censure for being inauthentic only when said inauthenticity is morally problematic. Thus Smith's phoniness, her posing as a basketball fan, can warrant censure only if her colleagues are deceived into thinking that she is a basketball fan. If not (e.g. if she tells them about her indifference to basketball), then, on our view, there is nothing wrong with her inauthenticity.

There are reasons to prefer an alternative version of the MORAL ACCOUNT, on which:

> MORAL ACCOUNT, VERSION 2: Being inauthentic is morally problematic when, and only when, the inauthentic person runs a foreseeable risk of misleading or deceiving someone, and on which it is problematic just to the extent that said foreseeable risk is high.

Thus even if Smith's colleagues aren't actually misled, her inauthenticity will still be morally problematic if she runs some substantial risk of misleading or deceiving them (or others). There are two reasons to favor the second version of the MORAL ACCOUNT. First, given our conception of inauthentic behavior (as incongruence between action and real attitudes), inauthentic behavior is a kind of behavior that, in general, has the potential to mislead or

deceive people. Knowingly engaging in behavior of that kind is morally problematic, even when nothing bad happens as a result, because such behavior is morally risky. Just as negligent behavior that puts others at risk of physical harm is morally problematic, even if no one is actually harmed, behavior that is known to have significant potential to mislead is morally problematic, even if no one is actually misled. Second, someone who engages in inauthentic behavior may become disposed to engage in it in the future. Thus, given the first reason, just stated, anyone who engages in inauthentic behavior runs the risk of inculcating a morally problematic habit.[17]

On this version of the MORAL ACCOUNT, a person might deserve censure even if her inauthenticty neither misleads nor deceives anyone: she deserves censure for engaging in the kind of behavior that runs a substantial risk of misleading and deceiving people, and for engaging in behavior that might make her disposed to mislead or deceive people in the future.[18]

Two caveats. First, even if someone's inauthentic behavior is morally problematic (as it sometimes will be, according to both versions of the MORAL ACCOUNT), it may not be morally impermissible: consider the Jews who denied their Jewishness to escape the concentration camps. To say that some behavior is morally problematic is to say that one always has a prima facie (or perhaps even pro tanto) moral reason to avoid that behavior. Second, that someone's inauthentic behavior is morally problematic, even morally impermissible, does not entail that it would not best for that person, all things considered, to be inauthentic. Consider Michelle. Perhaps misleading her date is morally problematic, perhaps nothing else speaks morally in favor of her misleading him, so that from the moral point of view, she ought not hold his hand. She might still be better off holding his hand: if she thinks withdrawing her hand will make him angry, or create a scene, or perhaps just because she enjoys ambiguously flirting with him, despite the fact that this involves misleading him. The MORAL ACCOUNT says that being inauthentic is sometimes morally bad; this doesn't mean that being inauthentic is so morally bad that good people don't sometimes indulge in it.

The second version of the MORAL ACCOUNT captures the BASIC INTUITION about authenticity, that there is something bad about being inauthentic, and something good about authenticity. In this respect, the MORAL ACCOUNT succeeds where the versions of the WELLBEING ACCOUNT that we considered above fail. But the MORAL ACCOUNT also succeeds in vindicating the CULPABILITY INTUITION: because being inauthentic is often morally problematic, the inauthentic person will often deserve censure. (But, given the two caveats, above, there might be cases of morally problematic inauthenticity in which it is morally best, or all-things-considered best, not to censure the inauthentic person.) A person

---

[17] The success of this argument for the second version of the MORAL ACCOUNT depends on an empirical assumption, common to many virtue theoretic accounts of morality: that engaging in a certain kind of behavior tends to habituate one to that kind of behavior. To the extent that this is false (and in particular, to the extent that it is false when it comes to inauthentic behavior), the first version of the MORAL ACCOUNT may be preferable to the second version.

[18] This defense of the MORAL ACCOUNT mirrors some remarks Baron makes about the self-deceiver's capacity to mislead others; see her 1988, pp. 434-5.

can be accused of doing something morally problematic, in a way that she cannot be accused of not living well. For this reason, the CULPABILITY INTUITION does not sit easily with the WELLBEING ACCOUNT; but it is a natural corollary of the MORAL ACCOUNT.

Bibliography:

Abrahamson, L.Y., and Alloy, L.B. (1981), "Depression, non-depression, and cognitive illusions: A reply to Schwartz," *Journal of Experimental Psychology* 110, pp. 436-47.

Alloy and Abramson (1979), "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?," *Journal of Experimental Psychology: General* 108, pp. 441-85.

Alloy, L.B., and Ahrens, A.H. (1987), "Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others," *Journal of Personality and Social Psychology* 52, pp. 366-78.

Anderson, T. (1993), *Sartre's Two Ethics: From Authenticity to Integral Humanity* (Open Court).

Arpaly, N. (2003), *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford University Press).

Baron, M. (1988), "What is Wrong with Self-Deception?," in McLaughlin and Rorty (eds.), *Perspectives on Self-Deception* (University of California Press), pp. 431-49

Bartky, S. (1990), *Femininity and Domination: Studies in the Phenomenology of Oppression* (Routledge, 1990).

Beauvoir, S. de (1948), *The Ethics of Ambiguity,* trans. B. Frechtman (Citadel Press).

Bernstein, R. (1971), *Praxis and Action: Contemporary Philosophies of Human Activity* (University of Pennsylvania Press).

Brown, J.D., and Dutton, K.A. (1995), "Truth and Consequences: The Costs and Benefits of Accurate Self-Knowledge," *Personality and Social Psychology Bulletin* 21, 1288-1296

Campbell, J.D. (1986), "Similarity and uniqueness: The effects of attribute type, relevance, and individual differences in self-esteem and depression," *Journal of Personality and Social Psychology* 50, pp. 281-94.

Catalano, J. (1990), "Successfully Lying to Oneself: A Sartrian Perspective," *Philosophy and Phenomenological Research* 38, pp. 384-98.

----- (1993), "Good and Bad Faith: Strong and Weak Notions," in Hoeller, K. (ed.), *Sartre and Psychology* (Humanities Press), pp. 79-90.

Conway, M., and Ross, M. (1984), "Getting what you want by revising what you had," *Journal of Personality and Social Psychology* 47, pp. 738-48.

Detmer, D. (1986), *Freedom as Value: A Critique of the Ethical Theory of Jean-Paul Sartre* (Open Court).

Du Bois, W.E.B. (1907), *The Souls of Black Folk* (A.C. McClurg).

Flynn, T. (2004), "Jean-Paul Sartre," *Stanford Encyclopedia of Philosophy*.

Frankfurt, H. (1992), "The Faintest Passion," *Proceedings and Addresses of the American Philosophical Association* 66, reprinted in his *Necessity, Volition, and Love* (Cambridge University Press, 1999), pp. 95-107.

Gardner, S. (1993), *Irrationality and the Problem of Psychoanalysis* (Cambridge University Press).

John, O.P., and Robins, R.W. (1994), "Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism," *Journal of Personality and Social Psychology* 66, pp. 206-19

Jopling, D. (1992), "Sartre's Moral Psychology," in Howells, C. (ed.), *The Cambridge Companion to Sartre* (Cambridge University Press), pp. 103-39.

Kruks, S. (2001), *Retrieving Experience: Subjectivity and Recognition in Feminist Politics* (Cornell University Press).

Lewishohn, P.M., Mischel, W., Chaplin, W., and Barton, R. (1980), "Social competence and depression: The role of illusory self-perceptions," *Journal of Abnormal Psychology* 89, pp. 203-12.

Parfit, D. (1984), *Reasons and Persons* (Oxford University Press).

Santoni, R. (1995), *Good Faith, Bad Faith, and Authenticity in Sartre's Early Philosophy* (Temple University Press).

Sartre, J.P. (1956), *Being and Nothingness*, trans. H. Barnes (Washington Square Press).

---- (2007), *Existentialism is a Humanism*, trans. C. Macomber (Yale University Press).

Taylor, C. (1989), *Sources of the Self* (Harvard University Press).

----- (1991), *The Ethics of Authenticity* (Harvard University Press).

Taylor, S., and Brown, J.D. (1988), "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103, 193-210

Trilling, L. (1972), *Sincerity and Authenticity* (Harvard University Press).

Velleman, J.D. (2002), "Identification and Identity," in Buss, S., and Overton, L. (eds.), *Contours of Agency* (MIT Press), pp. 91-123.

Weinstein, N.D. (1980), "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology* 39, pp. 806-20.

Wood, A. (1988), "Self-Deception and Bad Faith," in McLaughlin, B.P., and Rorty, A.O., *Perspectives on Self-Deception* (University of California Press), pp. 207-27.