

Inauthenticity and the Happiness-Welfare Link[†]

Dale Dorsey

Department of Philosophy
University of Kansas
1445 Jayhawk Boulevard
Wescoe Hall, rm. 3090
Lawrence, KS 66045
ddorsey@ku.edu

DRAFT of 10/29/2012

Pretheoretically, it's hard to escape the thought that a good life is, at bottom, a happy one. We take happiness to be an important feature of her well-being. Some believe that happiness should be the most significant determinant of a life's quality.¹ Offhand, none of these suggestions sound at all implausible.

However, not all states of happiness are created equal. Happiness can be *sadistic* or *shallow*. I might take pleasure in causing pain; I might take pleasure, say, in knowing the number of grains of sand in a particular hourglass. This sort of happiness, arguably, doesn't contribute to our own prudential interest in the same way that beneficent, or sophisticated, happiness might.² In addition, happiness can be *inauthentic*. Some instances of happiness are the product, say, of *sour grapes*, *adaptation*, or *brainwashing*. My happiness might be a learned response to negative stimuli; someone, for instance, might "learn to be happy" in circumstances they themselves would regard as degrading or otherwise undesirable. Most reject the claim that inauthentic happiness does much to improve my life.

For the sake of this investigation, I leave aside the problem of malicious and shallow happiness and instead focus on the problem of inauthenticity. I argue that an analysis of the nature of inauthentic happiness, i.e., happiness

[†]I would like to thank Anna Alexandrova, Donald Bruckner, Dan Haybron, Antti Kauppinen, Valerie Tiberius, Sam Wren-Lewis, and the various participants of the 2012 University of Leeds conference on measures of subjective well-being, from whom I learned a lot.

¹See, for instance, Fred Feldman, *Pleasure and the Good Life* (Oxford: Oxford University Press, 2004).

²For a contrary view, see Fred Feldman, *What is this Thing Called Happiness?* (Oxford: Oxford University Press, 2010).

that is not “in tune” with the person whose happiness it is—reveals that we should think twice about maintaining any strong link between happiness and welfare. The argument of this paper takes the form of a dilemma. After introducing the conceptual terrain in §§1-2, in §3 I argue that there is only one plausible account of the nature of inauthentic happiness, which I call the “correspondence account”. In §4 I introduce a problem for the correspondence account, for which there are only two strategies of response. Both strategies, or so I claim, result in the demise of a strong link between happiness and welfare.

1. *Happiness and Welfare*

The term “happiness” is about as vague as a term can get in moral theory. So it would do to say a little about how I use it. Some use it as synonymous with the term “well-being”. In other words, someone is happy to the extent that someone’s life is going well.³

This is not what I mean. When I say “happiness” I mean to pick out a particular sort of mental state or feeling that is distinguished from other sorts of mental states or feelings (i.e., sadness, anxiety, *ennui*, etc.). Different thinkers will identify happiness with different mental states. For instance, Fred Feldman holds that happiness simply refers to a state of *attitudinal pleasure*:

[T]o be happy at a time is to have a positive net balance of intrinsic occurrent attitudinal pleasure over intrinsic occurrent attitudinal displeasure at that time. More simply, to be happy at a moment is to be taking more pleasure than displeasure in things at that moment. Happiness in an interval is the integral of happiness at moments within the interval. Happiness in a domain is happiness taken in objects suitably associated with the domain. Happiness in life as a whole is happiness in the interval that is your whole life.⁴

On the other hand, Dan Haybron characterizes happiness in the following way:

[H]appiness... consists in a person’s overall emotional condition. This in turn has two elements: the individual’s various central af-

³Richard Kraut, “Two Conceptions of Happiness” in *The Philosophical Review* 88 (1979); John Stuart Mill, *Utilitarianism* II.2.

⁴Feldman, 137.

fective states (roughly, moods and emotions) and mood propensities. . . [T]o be happy is for one's emotional condition to embody a stance of *psychic affirmation* in response to one's life: emotionally responding to it as a favorable life—a life that is going broadly well for one, with only minor problems at most. . . *To be happy, then, is for one's emotional condition to be broadly positive—involving stances of attunement, engagement, and endorsement—with negative central affective states and mood propensities only to a minor extent.*⁵

For the moment, I want to focus not on the differences in Feldman and Haybron's accounts, but rather their similarities. Though Haybron's account is broader than Feldman's (i.e., Feldman's account of happiness is narrowly identified as attitudinal pleasure, Haybron's is a cocktail of various affective and mood states), both Feldman and Haybron embrace what I shall call an "affect-based account" of happiness.⁶ For such accounts, happiness is an affective state: a state of pleasurable feeling, a positive emotional state, or some combination thereof. But happiness *qua* affective state has unique properties that are not possessed by all affective states. In particular, happiness has a component of directedness: one will be happy *about* some state of affairs, one's life, a particular object or event, etc. For Feldman, happiness is characterized as being pleased *by* some particular object, state, or event. For Haybron, happiness is a broadly positive emotional condition *in response* to certain objects, events, or states of affairs (including one's life).

Some reject affect-based accounts. L. W. Sumner, for instance, understands happiness as not just an affective state, but a conjunction of affective and cognitive/evaluative states. He writes:

Being happy. . . means having a certain kind of positive attitude toward your life, which in its fullest form has both a cognitive and an affective component. The cognitive aspect of happiness consists in a positive evaluation of the conditions of your life, a judgment that, at least on balance, it measures up favourably against your standards or expectations. . . The affective side of happiness consists in what we commonly call a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it. Because it is less cognitively demanding than a

⁵Haybron, 147.

⁶Cf. Dan Haybron, "Philosophy and the Science of Subjective Well-Being" in *The Science of Subjective Well-Being*, ed. Eid and Larsen (New York, NY: Guilford Press, 2008).

judgment about how one's life is going as a whole, it is what we have in mind when we say that a child or an animal is happy, or is leading a happy life.⁷

The key distinction between affect-based accounts and, as I shall call them, "Sumnerian" accounts, is that affect-based accounts do not, whereas Sumnerian accounts do, insist that states of happiness are constituted not just by affective or emotional states, but also by *sui generis* evaluative states, directed toward that which one is happy about. For Sumner, to be happy about one's life is to maintain a positive affect toward that life, and also to evaluate one's life in a positive way. As with the relevant affective states, I will remain ecumenical concerning what mental state will characterize the non-affective evaluative state. Sumner holds that this state is cognitive,⁸ but one could, in principle, identify a conative state, i.e., a first- or second-order desire,⁹ etc. In any event, the distinction between these broad views should be clear enough for my purposes.

For now I'm going to set aside the Sumnerian account and focus on affect-based accounts. This is not because I think they're more plausible, but rather because doing so is sufficient to explore the central problem of this paper for *both* such accounts. I'll return more explicitly to Sumnerian accounts in §3.3 below.

2. *Inauthentic Happiness*

With these preliminaries in mind, many theorists believe that the following thesis is plausible:

The Happiness-Welfare Link (HWL): facts about the welfare or goodness of a life (or at a time within a life) are fully determined by facts about happiness within, or about, that life (or at that time).

This view is endorsed by each of Feldman, Haybron, and Sumner, and it's not hard to see why. There is a great deal of plausibility in claiming that facts about happiness determine facts about well-being.¹⁰ When we try to

⁷Sumner, 145-6.

⁸Incidentally, I agree. See Dale Dorsey, "Subjectivism without Desire" in *The Philosophical Review* 121 (2012).

⁹Cf. David Lewis, "Dispositional Accounts of Value" in *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 1988).

¹⁰Haybron, "Philosophy and the Science of Subjective Well-Being", 33.

figure out how well a person is doing at a time, we often seek to determine how happy he or she is at that time. When we wonder whether a particular life was good for the person who lived or lives it, we ask: was this a happy life?

But HWL is threatened. Take the following examples:

Nigel: Nigel lives in horrible poverty and squalor. However, given that many very powerful people have an interest in Nigel's happiness (and also his poverty) Nigel has been brainwashed into being happy about his life.

Now:

Sally: Sally is a talented painter, and makes a living at it. But she has always longed to be a professional architect. Sally, however, has found that her architecture has been rejected by the artistic and commercial communities. This had caused Sally a great deal of frustration and anger. However, she has tried hard to adapt, and has been successful at adapting, to her circumstances. As a result, she is *happy* about her life and career, despite the fact that she regards it as undesirable overall.

And one more:

Rusty: Rusty is a smoker, but he's trying to quit. Most days, however, he cracks. At a certain point, he just can't take it any more and lights up despite himself. As he is smoking, Rusty is happy about doing so, but comes to regret it.

There is something *off* about these cases of happiness. I'm going to characterize this "offness" in the following, very rough, way: these cases of happiness are not really *in tune* with the person who maintains them. For Nigel, his happiness is the result of brainwashing; for Sally, the result of strategic psychological engineering; for Rusty, a bare addiction. Insofar as these states of happiness are not in tune, plausibly, such states might rightly be called "inauthentic"; they do not, in Sumner's terms, "reflect the subject's own point of view".¹¹

The phenomenon of inauthentic happiness is a problem for HWL. Nigel, Sally, and Rusty's happiness cannot adequately account for the (dis)value of their lives. Surely we wouldn't say, of Nigel, that he's living a good life despite being happy about it. Inauthentic happiness can also affect our

¹¹Sumner, 172.

judgments about a person's welfare at a time. For instance, we wouldn't say, of Rusty, that *while he is smoking* he is well-off, despite the fact that he's very happy about it. Furthermore, we wouldn't say that merely because Sally has come to be happy *about her career choice* that she is living just as good a life as she would have had she made it as an architect.

It seems right to say, then, that the normative and evaluative importance of happiness is threatened by the possibility that such happiness can be inauthentic. One might put the threat this way. To fully determine the life quality of Sally, Rusty, and Nigel, we need *additional information*: information that goes beyond simply the fact of how happy they are. And some of this information will be *quite significant*. Nigel is not living a good life, though he may be very happy about it. As Sumner writes: “[Inauthentic happiness] does not accurately reflect the subject's own point of view. Welfare therefore consists in *authentic* happiness.”¹²

Of course, we should not *overstate* the extent to which the evaluative significance of happiness is threatened by its inauthenticity. Take, for instance, Rusty. Imagine that Rusty is not *happy* about taking the illicit smoke break, though he succumbs anyway. Surely this would be worse! Surely, if the only thing we could change about Rusty in this case is to provide him some modicum of happiness, even if inauthentic, there is reason to do so. Take also Nigel. If we assume that nothing else can be done for Nigel (i.e., he will never escape poverty, there are no prospects for states of authentic happiness, etc.), it seems right to say that there is strong reason to promote a state of *inauthentic* happiness. Indeed, doing so may be prudentially *optimal* for Nigel. As long as Nigel is going to be in poverty, one might say, he may as well be happy about it. And that holds *even if* such happiness is inauthentic.

Such reflections yield, I think, the following thought. Though there is reason to promote states of happiness generally, it also seems right to say that there is substantially *less* reason to promote inauthentic states of happiness than authentic states of happiness. Rusty, e.g., is made better-off to a far lesser extent insofar as he is inauthentically happy rather than authentically happy. Take Nigel *prior* to his brainwashing. Imagine now that though you cannot do anything with regard to Nigel's poverty (in other words, you cannot improve his life in *non-happiness* ways), you can either promote a state of inauthentic happiness by brainwashing him into being happy about his very poor conditions, or you could promote a state of *authentic* happiness by altering the world in a way about which Nigel will be

¹²Sumner, 172. My emphasis.

(authentically) happy. For instance, you might plant a number of beautiful willow trees that would generate a perfectly in-tune state of happiness for Nigel. When it comes to our reasons to promote happiness *qua* happiness, there is *far* more reason to engage in the latter course of action rather than the former. The latter is an instance of authentic happiness rather than inauthentic happiness. It seems worth marking this claim in the following way:

Principle: authentic happiness is better, all other things equal, than inauthentic happiness; there is stronger reason, all other things equal, to promote authentic happiness than inauthentic happiness.

Principle is of striking importance for HWL. If we are able to distinguish the evaluative significance of inauthentic and authentic happiness then it would appear that no more information is required to evaluate the lives of Nigel, Sally, and Rusty beyond facts about their states of happiness. All we need to know is that the happiness that they in fact maintain is inauthentic. Hence if we accept *Principle*, we seem to have a wide enough base of information to adequately explain the poor, or comparatively poor, quality of the lives lived by the characters named here.

If *Principle* is false, however, then it would appear that the lower-quality lives of Rusty, Sally, and Nigel cannot be accounted for *strictly* in terms of happiness. Given that Rusty, Sally, and Nigel maintain a high level of happiness, presumably just as high as people who live high quality lives, if we reject *Principle* we must reject the claim that the facts about their lack of welfare can be adequately explained simply by the axiological character of their happiness, which would just be to reject HWL.

Principle is of striking importance to HWL. But notice that *Principle* is not simply *analytic*. It is a substantive claim about the nature of happiness that must be defended. But *Principle* itself makes reference to an imprecise distinction, viz., that between inauthentic and authentic happiness. And hence any defense of *Principle* must rest, at the very least, on a more precise account of this distinction. I hope to show, however, that a proper conception of the inauthentic/authentic distinction either fails to be compatible with *Principle*, or can remain compatible with *Principle* only by yielding a conception of authentic happiness that is far too thin to capture sufficient facts about welfare to vindicate HWL.

3. What is Inauthentic Happiness?

What does it mean for a state of happiness to be “out of tune” with a person, or fail to capture a subject’s own point of view? I discuss three possibilities here, and eventually settle on what I call “the correspondence account”: a state of happiness properly reflects a person’s point of view if and only if that state of happiness is taken in an event, state of affairs, etc., that the person in question values.

3.1. *Authenticity as Autonomy*

Sumner, among others, suggests that authentic happiness is, most importantly, *autonomous*.¹³ Of course, autonomy is itself a concept that requires a specifying conception; I consider two potential conceptions here.

First, some hold that autonomous happiness is itself the subject of a form of *endorsement*.¹⁴ Leaving aside whether this is the right account of autonomy, one could offer an account of authentic happiness that takes the endorsement view as its model: a state of happiness is authentic if and only if some further evaluative mental state (such as a desire, a judgment, or, perhaps, an instance of “second-order” happiness, i.e., happiness about being happy) endorses that state of happiness.

Second, some have held that an autonomous mental state is one that is caused by the right sort of *process*. This account seems to attract Sumner:

However the details of a fully adequate view are worked out in the end, the implications for our theory of welfare are clear. Self-assessments of happiness or life satisfaction are suspect... when there is good reason to suspect that they have been influenced by autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, role scripting, and the like. Since these are all socialization processes, and since we are all historically embedded selves, the practical question becomes how much emancipation from her background and social conditions a subject must exhibit in order for her self-assessment to be taken at face value.¹⁵

¹³See, most importantly, Sumner, 166-70. Sumner also lists the requirement of full information. I’ll leave that aside for now.

¹⁴Sumner, 168-9.

¹⁵Sumner, 170-1. Worth noting here, of course, is that Sumner is referring to the autonomy of a Sumnerian account of happiness. But one could easily translate this proposal into an account of the autonomy of an affect-based conception of happiness: happiness (affect-based) is non-autonomous to the extent that it was caused by the problematic processes noted here.

Of course, whether mental state ϕ , created by the right sort of process, is necessarily autonomous or not is beyond my purview here. But one might say that a state of happiness is *authentic* when it is *not* the product of various problematic processes. Whether or not that is the right account of autonomy, it may very well be an acceptable account of *authenticity*.

I think that both the process-based and endorsement-based accounts of authentic happiness are mistaken. Start with the former. It is difficult to assess the proper process account unless more is said about just what constitutes proper processes. But let's leave this aside for a moment. Take Rusty. Rusty comes to be happy about a specific state (smoking) because *he's addicted*. But now compare Rusty his smoke-buddy, Ricky. Ricky is just like Rusty—he's just as addicted to smoking—but Ricky isn't trying to quit, and doesn't feel bad about himself afterwards. Ricky positively *revels* in smoking, and shows no sign of regret or remorse for doing so. In this case, it seems quite right to say that Ricky's happiness in smoking is authentic, Rusty's isn't.

This example sheds light on a very general problem with the proper process model. However one understands the right or proper processes and how to distinguish these from the wrong or improper processes, it *must* be able to distinguish between the happiness of Rusty and Ricky. After all, the former's happiness is inauthentic, the latter's happiness is authentic. But it would seem that there is no *process* that distinguishes the development of Rusty and Ricky's happiness. They're both just as addicted; they both came to their addictions in precisely the same way, or so we can assume. The difference between Rusty and Ricky is not a distinction in processes by which their happiness arose rather a distinction in *attitudes*: a distinction in whether they *prefer to be smoking* or not. And hence it cannot be that the proper process account provides a reasonable approach to the nature of inauthentic or authentic happiness. (Even though this may very well be a plausible account of the nature of autonomous happiness.)

Well, then, how about it? Why not simply reject the proper process view in favor of the endorsement view? This proposal would appear to allow a distinction between Rusty and Ricky: the latter endorses his own happiness in smoking, the former doesn't. But the endorsement view fails in other ways. Take Sally. Sally wants to be an architect, but realizes that she has no talent for it. So she attempts a process of psychological engineering: through therapy, or various other mechanisms, she is able to make herself happy about her circumstances, despite the fact that she finds them oppressive and undesirable. Question: does Sally endorse the state of happiness in the circumstances she is in? Absolutely! After all, she has

a strong, perhaps even decisive, prudential reason to develop it. Insofar as she could not obtain that which she desired, it makes prudential sense to be happy about the state of affairs she's in.¹⁶ And so she will of course endorse this state, despite the fact that it is inauthentic. And hence it seems to me that though the endorsement view adequately accommodates a distinction between Ricky and Rusty, it cannot capture the nature of inauthentic happiness: sometimes we endorse states of inauthentic happiness because, after all, they're good (perhaps even *optimal*) for us.

Of course, I have considered only two conceptions of autonomy here; there are potentially many more. Insofar as two primary accounts of autonomy seem to have failed as an account of authenticity, however, it is perhaps more productive to begin to look elsewhere.

3.2. *Authenticity and Richness*

Two central accounts of autonomous happiness fail to provide an adequate account of authenticity. But rather than suggesting that the problem for authentic happiness is a problem of lack of endorsement or a fact about the way it develops, one might say that the problem arises as a result of what one's happiness is taken *in*. Perhaps there is something about, say, Nigel's *life* and Rusty's *smoking* that renders any form of happiness taken in these things inauthentic.

One such proposal might be to treat authentic happiness just as happiness that is directed toward states that are in some way or other objectively good. Authentic happiness is *taken in, or directed toward, good things*.¹⁷ One such suggestion is made by Haybron. Haybron writes that authentic happiness requires:

richness. Briefly, the authenticity of one's happiness increases, other things being equal, to the extent that it is grounded in richer, more complex ways of living. For such ways of living more fully express one's nature. Someone might conceivably be happy, for example, leading the impoverished life of Rawls's grass counter. The choice to lead such a life could well be autonomous, say as a means of making happiness easier to come by. But there is not much of *him* in such a way of life, for he isn't really doing

¹⁶As Stephen Stills advises, if you can't be with the one you love, love the one you're with.

¹⁷Cf. Robert Adams, *Finite and Infinite Goods* (Oxford: Oxford University Press, 2002), 93-101.

much of anything—indeed, his happiness reflects a stunted vision of himself. . . A more authentic life—a life more fully expressing *his* nature, his individuality—would have him fully engaged in the business of living, with all the richness of an ordinary human life. And the resulting happiness would, it seems, be more authentic as well.¹⁸

This proposal fails for a number of reasons. First, it cannot distinguish Ricky and Rusty: Ricky and Rusty are happy about precisely the same thing, and hence one cannot distinguish the authenticity of their happiness simply on the basis of their being happy about good things. Second, and perhaps more seriously, this proposal seems to mix up two problems, viz., the problem of *shallowness*, already noted and set aside, and the problem of inauthenticity. To see this, recall Sally. Sally’s happiness is inauthentic. But it would seem manifestly odd to say that Sally’s inauthentic happiness isn’t *taken* in an objectively good state—i.e., the state of being a successful painter—or that her life somehow lacks in the dimension of complexity or richness described by Haybron.¹⁹ But that doesn’t mean that whenever one is happy about such a career one’s happiness is authentic.

Of course, this may just mean that richness is a necessary, not a sufficient, condition for authentic happiness. But this (a) still requires us to identify the sufficient condition(s)²⁰ and (b) seems false anyhow. Take the following case:

Cynthia: Cynthia has an odd number of hairs on her head. She takes great happiness in this state of affairs, committed as she is that the proper arrangement of human hairs is in an odd number.

Cynthia’s happiness—odd though it is—is surely not inauthentic or out of tune with Cynthia. But this doesn’t say anything about the objective value of the oddness or lack thereof of the number of hairs on Cynthia’s head or

¹⁸Haybron, 186.

¹⁹Haybron offers a very complex account of authentic happiness. As I read him, there are five conditions. First, one’s happiness mustn’t be based on false beliefs. Second, one’s happiness mustn’t be based on non-autonomous values. Third, authenticity requires “proper functioning, at least within broad limits”. Fourth, richness, as defined here. (Haybron, 185-6.) But even if we include all of these factors, they still seem to get Sally’s case wrong. Her happiness is based neither on non-autonomous values nor false information. Plausibly, she maintains “proper functioning” and it seems to me entirely unclear that we should regard her value as non-autonomous; certainly it is perfectly possible to regard this form of (inauthentic) happiness as a prudential value autonomously.

²⁰See note 21.

whether Cynthia’s way of living is rich or complex. Even if this state lacks all evidence of objective goodness (whatever the rubric), this doesn’t mean we can’t say that Cynthia takes authentic happiness in this state.

The current account is a non-starter, and so the reader might justifiably wonder why it receives treatment here. Answer: the problems with this view are generally illustrative of an important conceptual point. Different people can use the term “authenticity” in different ways. One way to understand the nature of authentic happiness is just to claim that authentic happiness is *valuable* happiness. And if that’s correct, and if we believe that Cynthia’s happiness is itself not particularly valuable, there may be pressure to characterize Cynthia’s happiness as inauthentic. If using the term in this way, there may be general pressure to require that any conception of authentic happiness *also* solve the problem of shallowness. But this is not the way I use this term. I understand authenticity, as Sumner does, as a state of happiness that represents or reflects a person’s point of view or is “in tune” with that person. As I use the term, the concept of authenticity simply refers to the extent to which a state of happiness reflects a person’s own point of view. And if this is right, it is clear that shallow happiness can be authentic: it can represent a person’s point of view just as much as rich happiness, or happiness taken in objectively valuable states.²¹

This, of course, does not mean that we must say that Cynthia’s happiness is just as good as any other kind of happiness. Shallow happiness might also be worth less, just as inauthentic happiness is. But if we regard shallowness as a problem for the value of specific states of happiness this doesn’t mean that richness must be an aspect of *authentic* happiness; we can simply treat the problem of shallowness as a separate problem and evaluate happiness along a further dimension. To claim, however, that rich happiness is necessarily authentic or in tune with a person in question seems a category error.

3.3. *The Correspondence Account*

Recall again the endorsement view. This view did not succeed because it is perfectly possible to see substantial reasons for particular states of happiness (including the reason that such a state of happiness would be a

²¹Haybron seems to believe that this kind of richness more fully captures a person’s nature or “individuality”. (Haybron, 186.) But this claim is implausible. The richness of a person’s life (or lack thereof) says nothing about the fulfillment of an *individual’s* nature, or that which is *in tune* with a particular individual. Given a person’s mental states, it could very well be that to take happiness in a “rich” versus comparatively simple life is itself inauthentic.

prudentially optimal benefit) even though those states are inauthentic. But one might find the problem with the endorsement view not in the suggestion that endorsement *per se* is relevant—which seems to capture the essential difference between Ricky and Rusty—but rather in the proposed *locus* of endorsement. Rather than holding that authentic happiness is characterized by an endorsed *state of happiness*, why not say that authentic happiness is characterized by a state of happiness taken in *an endorsed or valued state*? For instance, rather than saying that authentic happiness is Ricky’s endorsement of his *state of happiness in smoking*, but rather his endorsement of *smoking*?

Note, again, that I’m taking for granted (for the moment) affect-based accounts of happiness. This sort of an account allows us to make a distinction between a state of happiness (which is itself directed toward a state, object, or event) and the attitude of endorsement, i.e., an *evaluative* attitude (again, I will remain neutral on precisely the content of such attitudes). But once we allow such a distinction, we might account for authentic happiness by saying that authentic happiness is in tune with the person in question insofar as that to which one’s happiness is *directed* is identical to that to which one’s *positive evaluations* are directed. If I value, for instance, the state in which I am a grass counter, and am happy about it, this state of happiness is authentic because it is taken in a state that I value. When I am happy about something I *don’t value*—for instance, having a career as a painter rather than an architect—my happiness is inauthentic. Call this the “correspondence” view. Obviously the correspondence view of authentic happiness more-or-less coincides with the Sumnerian account of happiness *qua* a combination of an affective attitude toward ϕ and an evaluative attitude toward ϕ . But if this view is correct, this may be a happy result for the Sumnerian account of happiness. For the Sumnerian view, there may be no problem of inauthentic happiness at all. (More on this later.)

As stated, the correspondence view appears to respond adequately to cases of inauthentic happiness we have already seen. Take, for instance, Nigel. In this case, it would appear that any happiness Nigel takes in his life or circumstances will not match a similar state of valuing (occurrent or dispositional). In this case, his state of happiness (which, as we have been assuming so far, is strictly affective) is simply a product of brainwashing. On a Sumnerian view, Nigel simply *won’t be happy*, and hence the Sumnerian view also can appear to capture the relatively poor quality of Nigel’s life (*mutatis mutandis* for other cases). Take also Sally. Sally clearly does not *value* the state in which she is a painter rather than a successful architect. And hence it seems correct to say that insofar as Sally’s happiness is taken

in a state that is contrary to that which she has a disposition to value, this state of happiness is inauthentic: it is a state of positive affect directed toward a state of affairs that, on reflection, she would not endorse as valuable. This view also appears to properly account for the distinction between Ricky and Rusty: Rusty does not, where Ricky does, endorse or value the state in which he smokes. These are the right answers and hence such cases lend support, broadly speaking, to the correspondence account. When accompanied by *Principle*, the correspondence account can adequately explain the comparatively poor lives of Nigel, Sally, and Rusty.

One might object. Surely states of authentic happiness about such-and-such can arise without necessarily taking a specific evaluative attitude toward such-and-such. I may be happy, for instance, about a display of fireworks that surprises and delights; but I don't take any *per se* evaluative attitude toward that display. Sometimes we may just be caught, as it were, "off guard", happy without even considering whether the state at which our happiness is directed is good or bad. And why should we say that those states of happiness are necessarily inauthentic?

I don't think the correspondence view needs to claim that such states are not authentic. I here mark a distinction between *occurrent* and *dispositional* valuing. Someone occurrently values ϕ at t to the extent that this person actually takes the relevant valuing attitude toward ϕ at t . A person *dispositionally* values ϕ at t to the extent that this person has the disposition, given his or her actual mental states, to value ϕ at t . Dispositional valuing does not entail occurrent valuing. And hence in the case of someone who takes a state of happiness without also taking the evaluative attitude, this does not entail that the person in question fails to *dispositionally* value. Rather, he or she simply fails to *occurrently* value the state in which happiness was taken. However, I think it is quite right to say that whenever a person fails to have at least the disposition to value some state, any happiness taken in that state is inauthentic. If I lack even the *disposition* to judge the state in which I view the fireworks as a good one, it seems right to say that such happiness is not in tune with my own point of view. And hence a sensible refinement of the correspondence account is to treat happiness as authentic on the assumption that this particular state of happiness is taken in an object or state of affairs the person in question has at least the disposition to value.

The correspondence view gets the right answers to all cases discussed so far. This is good evidence for it. But I offer here a further argument to show that, broadly speaking, any proper account of the nature of authenticity must at least be compatible with the correspondence view. Recall first

that authentic happiness is happiness that is supposed to fail to represent or reflect the person’s point of view, fail to be “in tune” with the person in question. This was the crucial conceptual point illustrated by the failure of the “richness” account. But this entails that to assess the authenticity of a state of happiness one must compare this state to some *other* mental attitude or disposition (actual or counterfactual) possessed by the person in question—the person’s “point of view”, as Sumner calls it. But to understand this, we must posit *some other mental state*, possessed by the person in question, that the state of happiness “reflects” or does not, or “is in tune with” or is not. But what is this? What is a person’s “point of view” with which a state of happiness will or will not be in tune? It cannot simply be some other affective state, as this alternative affective state, whatever it is, would present the very same potential for inauthenticity. Rather, it must be some sort of evaluative state (whether cognitive or conative, or whatever): something that evaluates some state of the world as good, or desirable, or worth wanting, or, at the very least, *wanted*. But there are only two plausible ways that a state of happiness could be in tune with such a state. First, it could be that the state of happiness itself is evaluated positively. But this is a version of the endorsement view, and has already been rejected. And hence it must be that the state in which happiness is taken is itself evaluated positively. And this just is the correspondence view.²²

4. *Iterated Inauthenticity, Trade-off, and the Demise of HWL*

This is a key moment in my argument of this paper. As I shall argue, the truth of the correspondence view is the demise of HWL. To see this, I’m going to introduce a challenge to the correspondence view to which an acceptable account of the authenticity of happiness must respond. Following this, I investigate two responses, both of which cause serious problems (though in different ways) for HWL.

To begin, notice that the correspondence account as stated entails that there is no problem for inauthentic happiness on a Sumnerian account, insofar as on that view instances of happiness correspond to valued states. But Sumner *himself* believes that his account of happiness is subject to the phe-

²²It is worth noting also that Sumner and Haybron seem to endorse something like the correspondence view as providing *at least* a necessary condition for authentic happiness. Sumner, obviously, insofar as he believes that happiness that does not correspond isn’t a genuine state of happiness. But Haybron also insists that a necessary condition of authentic happiness is that it is based on autonomous “values”; this point will be explored in more depth in the next section.

nomenon of inauthenticity. Sumner asks, of classic examples of adaptation from Sen:

Why are we reluctant to take at face value the life satisfaction reported by ‘the hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie’? Presumably because we suspect that the standards which their self-assessments reflect have been artificially lowered or distorted by processes of indoctrination or exploitation. In that case, the obvious remedy is to correct for the conditions under which their expectations about themselves came to be formed. The problem is not that their values are objectively mistaken but that they have never had the opportunity to form their own values at all.²³

According to Sumner, “life satisfaction” (or a state of Sumnerian happiness taken in one’s life) should not be taken at face value insofar as the evaluative attitudes or “standards” with which states of happiness correspond can themselves be “distorted by processes of indoctrination or exploitation”. For the correspondence view, the problem here should be obvious. Evaluative states, however they are conceived, can *themselves* be inauthentic, for reasons explored by Sumner and for many others. Hence it would appear that the inauthenticity of evaluative states “bleeds into” the inauthenticity of happiness. After all, why believe that authentic happiness should be in tune with (or be constituted by) evaluative states that are themselves inauthentic? Call this the problem of “iterated inauthenticity”.

Insofar as a correspondence view is *broadly* correct, any acceptable account of the authenticity of happiness must respond to the problem of iterated inauthenticity. There are two strategies for doing so. First, one might try to supplement the correspondence view with additional constraints on one’s evaluative attitudes, designed to ferret out only those evaluations that are genuine or authentic. One might then offer a theory of authentic happiness that requires evaluative states be genuine or authentic in these ways. Indeed, there may be many such accounts. According to Haybron, a crucial necessary condition for authentic happiness is that authentic happiness not correspond to or be “grounded” by evaluative states that are not autonomous.²⁴ In addition, he lists condition of adequate information.²⁵ According to Sumner, evaluative states that help to constitute happiness must

²³Sumner, 166.

²⁴Haybron, 186.

²⁵Haybron, 185.

pass tests of full information and autonomy.²⁶ Elsewhere, I have offered a somewhat different account: evaluative states are authentic when they are coherent and complete, which carries with it a requirement of full information.²⁷ Call this the “amendment strategy”.

Alternatively, one might simply deny that the problem of iterated inauthenticity is genuinely a problem for the correspondence view. (Call this, obviously enough, the “denial strategy”.) I’m not going to hang my hat on the denial strategy or the amendment strategy. No matter which strategy you adopt, HWL is deeply threatened. The amendment strategy results in the rejection of *Principle*, the denial strategy, though compatible with *Principle*, entails that facts about happiness and its authenticity fail to capture important facts about life quality for persons.

4.1. The Amendment Strategy

To see the consequences of the amendment strategy for *Principle*, I want to discuss the possibility of *trade-offs* between states of happiness that correspond to authentic evaluative states and states of happiness that correspond to inauthentic evaluative states. (I should note that this way of understanding the proper trade-offs is not kosher according to the Sumnerian account of happiness. On that account, we are not trading off states of happiness that *correspond* to various authentic or inauthentic evaluative states. But instead we are trading off states of happiness that are in part *constituted* by various authentic or inauthentic evaluative states. I flag this to lay it aside; it won’t affect the argument and fans of the Sumnerian account are free to make the proper substitutions *mutatis mutandis*.) Here are three cases. First,

Rusty, again: Rusty is a smoker, but he’s trying to quit. Most days, however, he cracks. At a certain point, he just can’t take it any more and lights up despite himself. As he is smoking, Rusty is happy about doing so, but comes to regret it. However, Rusty’s evaluation of his smoking is, in fact, inauthentic, the product of numbing advertisements from the American Lung Association. Rusty’s authentic evaluations evaluate his smoking positively. Imagine that we could render Rusty happy about smoking, or could render it the case that Rusty quits smoking, and is happy

²⁶Sumner, 139.

²⁷Dale Dorsey, “Preferences, Welfare, and the Status-Quo Bias” in *Australasian Journal of Philosophy* 88 (2010).

about that state.

Second:

A and B: Imagine two people, A and B. Neither is happy about their lives. A maintains inauthentic evaluations, but simply could not be brought to live a life that conforms to her authentic evaluations. B maintains authentic evaluations and, in fact, her life conforms to such evaluations. Imagine that we could promote a state of happiness that conforms to either person's actual evaluations (but not both).

Third:

C and D: Imagine two people, C and D. To promote a state of happiness for C would be to promote a state of happiness that corresponds to an authentic evaluative state. To promote a state of happiness for D would be to promote a state of happiness that corresponds to an inauthentic evaluative state. The states of happiness are equal in every respect save for the fact that D will experience *slightly* more happiness than C.

These cases are distinct in several ways, but each tells against the suggestion that we should trade happiness that corresponds to (or is in part constituted by) inauthentic evaluative states for happiness that corresponds to (or is in part constituted by) authentic evaluative states. Take the first case. Suppose that Rusty's authentic evaluations—were he to possess them—would endorse his state of smoking. But given that Rusty *himself* does not value that state, would we say that Rusty's happiness about not smoking, or happiness about actually smoking (despite the fact that he does not actually value that state) makes him better-off? The former answer, to me anyway, seems obvious. Take now A and B. I find it very difficult to believe that one should prefer B's happiness to A's. Indeed, I must confess a slight preference for the happiness of the former individual, as we might presume that she is worse-off given the case as stated. But if we presume that both individuals are equally well-off, I feel no pull whatsoever to the claim that one ought to promote the state of happiness that conforms to authentic evaluations rather than the one that conforms to inauthentic evaluations. Finally, consider C and D. In this case, it appears that everything about the states of happiness in question are identical save for the fact that C's happiness corresponds to authentic evaluations, *vice versa* for D, and the fact that D's happiness will be of slightly greater quantity: he will be slightly

happier. Which should we prefer? Again, this is a considered judgment that perhaps not all will share, but I find it very clear that D's is to be preferred, if for no other reason than simply the fact that promoting a state of happiness for D will promote *more happiness*. And if this is right, it is hard to accommodate the general claim that happiness that corresponds to authentic evaluations rather than inauthentic evaluations is more valuable.

This verdict is made all the more striking when we compare trade-offs that involve states of happiness that do not correspond to actual evaluations and those that do. Take:

E and F: Imagine two people, E and F. To promote a state of happiness in E would be to promote a state of happiness that is taken in an object or state of the world that E does not actually value, that does not correspond to E's evaluative states. To promote a state of happiness in F, however, would be to promote a state of happiness that is taken in an object or state of the world that F does, actually, value.

It seems extremely plausible that one would prefer F's happiness to E's, all other things equal. Corresponding happiness (even if this state of happiness corresponds to inauthentic evaluations) is more valuable than a state of happiness that does not similarly correspond. But this verdict does not seem as plausible when considering trade-offs between states of happiness that correspond to inauthentic, and those that correspond to authentic, evaluations.

What is the point of all this? I have established, I think, the following claim:

Trade-off: it is not the case that states of happiness that correspond to authentic evaluative states are to be preferred to states of happiness that correspond to inauthentic evaluative states.

If we accept the amendment strategy, *Trade-off* wrecks *Principle*. *Principle* holds that states of authentic happiness are better than states of inauthentic happiness, which is precisely what is rejected by the combination of *Trade-off* and the amendment strategy. But given that *Principle* is central to a defense of HWL, the amendment strategy results in HWL's rejection.

4.2. *The Denial Strategy*

Rather than amending one's conception of authentic happiness to avoid

the problem of iterated inauthenticity, one could adopt the denial strategy and simply reject the claim that iterated inauthenticity is a problem. (You might be tempted to do this if you want to preserve *Principle* in light of the cases discussed in the previous section.) But the denial strategy, though it remains compatible with *Principle* (in other words, *Principle* is not threatened by our reluctance to make the trade-offs indicated in the previous section, simply because the inauthenticity of an evaluation says little about the inauthenticity of a corresponding state of happiness), nevertheless demonstrates quite clearly that HWL cannot be correct. The denial strategy yields a theory of authentic happiness that is too thin to represent all judgments about the relative quality of lives. Recall the characters noted by Sen. It seems right to say, for instance, of the “overexhausted coolie” and others, that these individuals do not maintain a particularly high degree of welfare. Their lives are not going particularly well, and this is surely part because *their own evaluate states are inauthentic*. But if it’s the case that the problem with these individuals’ lives cannot be accounted for in terms of inauthentic *happiness* (which is implied by the denial strategy), it would appear that there is significant information about the quality of life that cannot be accounted for simply by considering the extent to which that person is happy and whether that person’s happiness is authentic or inauthentic. HWL, it would appear, simply fails.

Indeed, something like this very point is noted by partisans of HWL. Sumner notes that the problem of inauthentic happiness (which, for a Sumnerian account of happiness, just is the problem of iterated inauthenticity) affects those “who think there is *some* interesting connection between welfare and happiness.”²⁸ Of the problem of inauthentic evaluations, Sumner later writes: “Its importance cannot be overstated; it is surely the main reason for questioning the adequacy of any subjective theory of welfare, whatever its constituent ingredients, and for favouring more objective accounts.”²⁹ Even if the authenticity of evaluative states should not determine whether a person’s states of happiness are authentic, the fact that a person values particular states inauthentically is surely relevant to a consideration of his or her *well-being*. Whether our lives conform to authentic evaluative states surely helps to determine how well our lives are going.

The last claim calls out for explanation. Though it is beyond the purview of this paper to sort through all such possibilities, one significant thought is that conformity of one’s life to one’s authentic evaluations *itself* makes one

²⁸Sumner, 157.

²⁹Sumner, 162.

better-off.³⁰ The fact that I authentically value the achievement of some goal seems to entail that this is, of itself, reason to believe that its achievement is an intrinsic good for me. And hence while inauthentic evaluations do not give rise to inauthentic happiness, they do have a general tendency to give rise to *worse lives* (at least when compared to lives that correspond to authentic evaluative states). Notice that this is an *explanation* of this phenomenon, not an argument that the phenomenon is real. That the authenticity of evaluations at least plays some role in the notion of life quality is not delivered by the question-begging assumption that the fulfillment of that which one values is itself a contributor to a life well-lived. I come to this conclusion, rather, by a simple consideration of cases, like Sen's, in which the fact of authentic happiness—on the denial strategy—leave facts about welfare unaccounted for.

5. Three Responses

Three responses to the above argument are worth consideration here. The first two are responses on behalf of the amendment strategy, the last a response on behalf of the denial strategy.

5.1. Revising Principle

I claimed that the most significant problem for the amendment strategy was that it seemed implausible to believe that states of happiness that correspond (or are constituted by, in Sumerian terms) authentic evaluations are worth more than states of happiness that correspond to actual evaluations, which may be inauthentic. This fact entails, on the amendment strategy, a rejection of *Principle*. Insofar as *Principle* is central to any defense of HWL, however, this strategy entails a rejection of HWL.

But is *Principle* really as central to HWL as I think? A partisan of HWL might complain that though *Principle* fails on the amendment strategy, he or she can simply be more fine-grained about the evaluative relationship between states of happiness. Authentic states of happiness, so it may be claimed, are better than inauthentic states of happiness, except in cases of inauthentic happiness that correspond to an actual, but perhaps inauthentic, evaluative state. This is a perfectly open option, but this view has the same problems as the denial strategy. If the general axiological attitude toward

³⁰Of course, this does not entail that our evaluations are the sole determinant of that which is good for us.

the relative value of happiness that corresponds to inauthentic evaluations is roughly equivalent to the value of happiness that corresponds to authentic evaluations, a happiness-based account of welfare, one that conforms to HWL, will not be able to account for the real difference in the quality of lives between individuals whose lives conform, and those whose lives do not conform, to their authentic evaluations. And as explored in §4.2, this difference is real.

5.2. *Rejecting Trade-off*

The cases I offer in §4.1 seek to show that we should not prefer states of happiness that correspond to authentic evaluations to states of happiness that correspond to actual, but inauthentic, evaluations.

But one might reflect on a slight variation on one of Sen's cases. Take an overexhausted coolie who chooses to live in this way as a result of genuine desires to live close to the soil, or finds a life of struggle and destitution a way of penitence before the almighty, or for whatever other reason. Call this person, "G". Now compare this person to an overexhausted coolie who chooses to do so for the same reasons, but that the salience of these reasons was the result of "brainwashing, role-scripting" etc. Call this person "H". Though we may say that both lives are *bad*, all things considered, it would seem that G is living a *better* life. But if this is correct, it would seem that it *is* better to maintain states of happiness that correspond to *authentic*, rather than inauthentic evaluations.

This verdict, however, is too fast. We could, in principle, explain the difference (which I agree is real) in the quality of these lives simply in terms of the value of their happiness. *Or* we could explain the difference in the quality of these lives in terms of other factors, including the extent to which their lives conform to their authentic evaluations. Only if the former is the better explanation should we reject *Trade-off*. But take again the cases discussed in §4.1. These cases exposed a reasonably robust judgment that *when it comes strictly to the axiological contribution of states of happiness*, we should not prefer states of happiness that correspond to authentic evaluations to states of happiness that correspond to actual, but inauthentic, evaluations. To then explain the relative value of the cases offered in the previous paragraph simply in terms of happiness is to overturn these judgments, it seems to me, solely to preserve HWL. As such it is *ad hoc* and anti-accommodationist. It is better, it seems to me, to preserve *Trade-off*, and explain the relative value of the lives of G and H in a way that does not simply to their states of happiness. Here's another way to put this point.

There could be many explanations for the relative value of the two variations of the overexhausted coolie noted in the previous paragraph. But the cases of §4.1 can be interpreted as tests of the HWL-friendly explanation, viz., that the relative value of these lives should be explained in terms of happiness. But this explanation appears to come up short. We should, therefore, explore an alternative explanation, and with it the rejection of HWL.

5.3. *Shallowess, Sadism, Etc.*

The final response—this time on behalf of the denial strategy—runs as follows. Throughout this paper, I’ve been more or less friendly to the suggestion that a fan of HWL may wish to make further axiological distinctions in the value of happiness beyond the distinction between authentic and inauthentic happiness. Surely, he or she will say, the mere fact that a state of happiness is authentic is not sufficient to render it axiologically superior or sufficient to render a life good. After all, this happiness could be sadistic or shallow. But if this is correct, one might think that this information is something the partisan of HWL will bring to bear in evaluating the lives of, e.g., the landless laborer. After all, it might be said, this happiness, though not inauthentic, is *shallow*. And if we allow, in a *Principle*-like way, that shallow happiness is also less valuable than rich happiness, we can accommodate additional facts about the disvalue of lives within the broad “happiness” metric that a mere emphasis on authentic over inauthentic happiness cannot. (Notice that many advocates of HWL, including Sumner, will reject this view. As noted above, for Sumner, welfare just *is* “authentic happiness”.)

However, even if we accept this amendment, it doesn’t succeed. For it to do so, it must be the case that any instance of the relevance of inauthentic evaluations to welfare can be captured by other independently motivated evaluative distinctions among states of happiness. But even *a priori* this is very implausible. Note again the two versions of an overexhausted coolie discussed in the previous section. Their lives are of differing quality, and on the denial strategy the differing quality of their lives cannot be explained by the extent to which their lives maintain inauthentic or authentic happiness. But it *also*, it seems to me, cannot be explained by the relative shallowness of (or any other significant fact about) their happiness: G and H’s happiness is taken *in the same state*. The only difference between these two lives is the authenticity of their evaluations. And if this is correct, the denial strategy cannot succeed.

6. *Conclusion*

By way of conclusion, seems worthwhile to recap the general argument of the paper. It takes the form of a dilemma. Point one: non-correspondence accounts of the authenticity of happiness fail. Point two: to vindicate the correspondence account against the problem of iterated inauthenticity, one must adopt either the amendment strategy or the denial strategy. Point three: in light of *Trade-off*, the amendment strategy is incompatible with *Principle*, and hence HWL fails. Point four: the denial strategy offers a theory of authentic happiness that is too thin to vindicate the claim that facts about happiness could plausibly capture our judgments of life quality. And hence, point five: HWL fails. The strong link between happiness and welfare cannot be sustained.