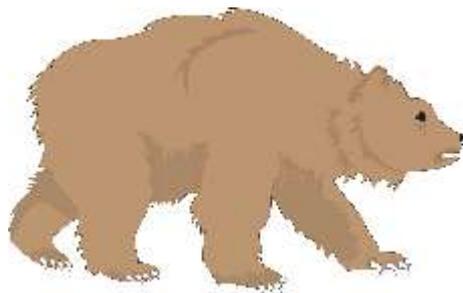


Ursus Philosophicus

Essays dedicated to Björn Haglund on his sixtieth birthday



Happiness and the good life

Bengt Brülde

Abstract

The paper starts with a presentation of the pure happiness theory, i.e. the idea that the quality a person's life is dependent on one thing only, viz. how happy that person is. To find out whether this type of theory is plausible or not, I examine the standard arguments for and against this theory, including Nozick's experience machine argument. I then investigate how the theory can be modified in order to avoid the most serious objections. I first examine different types of epistemic modifications of the theory (e.g. the idea that a person's happiness is more valuable for her if it is based on a correct perception of her own life), and then turn to a number of modifications which all make the value of a person's happiness depend on whether the evaluative standard on which her happiness is based satisfies certain requirements. In connection with this, I present and defend my own modified version of the happiness theory.

Introduction

Many people tend to believe that it is ultimately happiness that makes a life worth living, i.e. that the good life is identical with the happy life. On the "pure version" of this view, happiness does not just play a very central role in the good life (quality of life, or well-being): it is also assumed that there are no other final prudential values besides happiness. On this view, things like love, friendship, meaningful activity, freedom, human development, or the appreciation of true beauty are "merely" instrumentally valuable for us, i.e. they are not good as ends but "merely" as means to the only thing that is good as an end, namely happiness.

This view gives rise to several questions, three of which will be thoroughly discussed in this paper. The first question is what version of the pure happiness theory that

is most plausible, i.e. what conception of happiness that has most moral and rational significance. Here, it is argued that the hybrid view of happiness is the conception which makes happiness most valuable for us. The second question is whether (the most plausible version of) the pure happiness theory is a reasonable theory of well-being. To find out if this is the case, some of the most common arguments that have been given for and against this position are examined, and it is concluded that it is not a very plausible theory. The third question is how the pure theory can and should be modified in order to avoid the most serious objections, e.g. whether we should accept Sumner's modified happiness theory, according to which a life can't be really good for the person who lives it unless the cognitive part of a person's happiness (the value judgement) is informed and autonomous. In connection with this, I present and defend my own modified version of the happiness theory. In short, the central purpose of this paper is to find out what exactly the role of happiness in the good life is, e.g. whether it is the only thing that has final value (is good as an end) for a person, or whether there are also other final values besides happiness.

The pure happiness theory and its different versions

According to the pure happiness theory, the quality of a person's life (how good a life is for the person who is living it) is dependent on one thing only, namely how happy that person is. The happier she is, the better her life is, and the unhappier she is, the worse. On this view, happiness is the only thing that has positive final value (value as an end) for a person, and unhappiness is the only thing that is bad as an end. Moreover, the value that a certain state of happiness has for the experiencing subject is wholly dependent on how happy this state is (how "intense" the happiness is), e.g. the prudential value of a happy state is not at all dependent on how the state originated, or on whether it is based on illusory beliefs, sadistic preferences, or the like. (This is not to deny that happiness also has instrumental value, however, e.g. that it is also good because it has positive effects on e.g. the person's health and social behaviour, which in turn may give rise to more happiness.)

The precise meaning of these claims is of course dependent on what we take a person's happiness to consist in. So, how might the concept of happiness be understood, and how is it best understood in this particular context?

Four conceptions of happiness

There are at least four conceptions of happiness that may be of relevance in this context, namely (1) the cognitive (or attitudinal) view, (2) the hedonistic view, (3) the mood view, and (4) the hybrid view, according to which happiness is a complex mental state consisting both of an affective and a cognitive component.

1. The cognitive view. On the pure cognitive (or attitudinal) view, happiness is regarded as a cognitive state, or more specifically, as a positive attitude (in the philosophical sense of this term) towards one's life as a whole. On this view, happiness has no affective component, i.e. the attitude in question need not (at least not by definition) be accompanied by any pleasant feeling. To be happy is simply to evaluate one's own life in a positive manner, to approve of it, or to regard it favourably. Or more specifically, happiness is conceived of as consisting in "a positive evaluation of the conditions of your life, a judgement that, at least on balance, it measures up favourably against your standards or expectations. This evaluation [...] represents an affirmation or endorsement of (some or all of) the conditions or circumstances of your life, a judgement that, on balance and taking everything into account, your life is going well for you." (Sumner 1996, p. 145) That is, the satisfaction involved in happiness is global, what one is happy with (the object of appraisal) is one's life as a whole. Moreover, it seems pretty clear that this global attitude cannot be reduced to the aggregate of one's particular (perceived) satisfactions and dissatisfactions. Haybron (2001) refers to this view as the life satisfaction view, and Veenhoven (1984) refers to this type of happiness as contentment.

How much a person endorses her life (or how "enmeshed she is with her world"; cf. Barrow 1980) seems to be a matter of to what degree she perceives her conscious aims (or important goals) to be achieved. This is mainly a matter of "completeness" (cf. Nordenfelt 1991): To be maximally happy in this regard is to be satisfied in all respects, to have a favourable attitude towards one's total life-situation (and not merely some elements or aspects of it). There are no standing dissatisfactions, and there is no desire to

change anything; the world as one perceives it is exactly what one would like it to be. A person's level of satisfaction may also be a function of how "strong" the satisfactions are, however, where strength should not be understood in terms of pleasantness.

The most serious objection to this view can be found in Haybron (2001). On Haybron's view, we rarely have the global attitudes to which the theory appeals; for most people most of the time, there simply is no fact of the matter whether and to what extent they are satisfied with their lives as wholes. (If there is no lasting, stable attitude underlying people's judgements of life satisfaction, this is of course bad news for the hybrid view as well; see below).

On *the pure affective view*, happiness is some kind of affective state. To *be* happy is (roughly) to *feel* happy. On this view, happiness has no cognitive component, e.g. it doesn't essentially involve any positive evaluation of one's life as a whole. There are at least two different versions of this view, versions that differ with regard to what kind of subjective well-being one thinks of as constitutive of happiness.

2. On *the hedonistic theory*, happiness is best regarded as a favourable balance of pleasure over displeasure (or "pain"). The idea that happiness consists in a positive or favourable affect balance can be regarded as a special version of this idea.

3. According to the mood theory, happiness is a certain kind of positive mood state (or "thymic state"), a state that need not be about anything in particular. Sumner (1996) characterizes this mood as a feeling of energy, vitality, and buoyancy of spirit, as "a mood of optimism or cheer which colours your outlook on your life and on the world in general" (p. 144). He also points out that this feeling can range from bare contentment to deep fulfilment.

4. On *the hybrid view*, happiness is regarded as a complex mental state, in part cognitive and in part affective. To be happy is, roughly, a matter of (a) cognitively evaluating one's life as a whole in a positive manner, and (b) to feel good as a result. That is, a person's level of happiness is a function of two things, viz. (a) how satisfied he is with his life (as he himself perceives it or conceives of it), how positive his attitude toward his own life is, and (b) how good he feels about this. (This type of view is sometimes called the life satisfaction view, and it has been shared by many thinkers, e.g. Tatarkiewicz 1976, Veenhoven 1984, Nordenfelt 1991, Sumner 1996, Brülde 1998.)

The cognitive component has already been characterized above (cf. the cognitive view). As far as the affective component is concerned, most hybrid theorists seem to prefer the account offered by the mood theory to the account offered by the hedonistic theory. How happy a person is in the affective dimension is (in any case) not merely a matter of how intense and frequent his pleasure is, i.e. the degree to which a person is affectively happy does not necessarily vary with the degree to which he feels pleasure. Here are some examples of the alternative views that have been offered. On Barrow's (1980) view, we should give more weight to "important pleasure", e.g. to the pleasant states of mind that accompany the perceived achievement of important goals: the more important pleasure, the more happiness. On Sumner's (1996) view, "[t]he 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" (p. 146). According to Tatarkiewicz (1976), the degree of affective happiness is (rather) a matter of to what extent one's consciousness is dominated by pleasantness. He writes: "If satisfaction is to be considered complete it must pervade the whole of our consciousness and not just flow over its surface" (p. 11), it must reach "the depth of a man's consciousness and touch [...] its innermost fibre" (ibid., p. 21). "Surface experiences, however pleasurable, will not provide complete satisfaction" (ibid., p. 11).

Regardless of how exactly the two components are characterized, it is worth pointing out that they do not always co-vary, e.g. it may well be the case that one person A is more cognitively satisfied with his life than another person B is with his life, while B's hedonic level of affect is higher. It is also worth noting that it is not quite clear how these two dimensions should be combined into a single scale of happiness. For example, suppose (again) that A has a more favourable attitude towards his own life than B has, whereas B feels better. In this case, it is not clear who is happier on the hybrid view, A or B. It is not even clear if this question has a determinate answer.

Four possible versions of the pure happiness theory

Depending on which conception of happiness that is incorporated into the pure happiness theory, we end up with four different versions of this theory.

If happiness is understood as a favourable balance of pleasure over pain, the pure happiness theory is identical with pure hedonism.

If happiness is instead regarded as a certain kind of mood state, the pure happiness theory has to be regarded as a *restricted version* of hedonism. The reason why it is not a pure form of hedonism is that it conceives of certain kinds of pleasant experiences (viz. pleasant moods) as more valuable (and more conducive to happiness) than others, e.g. transient pleasant sensations.

A happiness theory that incorporates the cognitive conception of happiness is clearly not a version of hedonism, however, but rather a kind of mental state theory that attributes final value to other mental states besides pleasure. The difference between the cognitive version of the happiness theory and the desire theory is that the former theory does not make a person's well-being depend on the state of the world. It is sufficient that the subject *believes* that his life is going the way he wants it to go. That his life is *actually* going the way he wants it to go is not essential for well-being.

If we incorporate the hybrid view into a pure happiness theory, we get a mental state theory that is in part hedonistic, but that also makes a person's level of well-being depend directly (and not just causally) on how satisfied she is with her life.

To conclude, no matter which version of the pure happiness theory we have in mind, it is obviously a mental state theory, i.e. it claims that the quality of a person's life is wholly dependent on the person's mental states, and not at all on the state of the world (except in a causal sense).

Which version of the pure happiness theory is most plausible?

Before we ask ourselves whether any of these four theories is a plausible theory of well-being, let us first consider which of the four theories that is most reasonable.

A case against hedonism

In my view, pure (classical) hedonism can hardly be the most reasonable version of the pure happiness theory. The pure hedonist claims that the quality of a certain life is a function of how much pleasure and displeasure this life contains. The more pleasure and the less "pain" a life contains, the better this life is for the person who is living it, period. Here, it does not matter at all to what category a person's pleasures belong, e.g. whether they are sensations, emotions, or moods. All that matters is how pleasant these

pleasures are. This suggests that a large amount of pleasant sensation is sufficient to make a human life good, no matter what other features this life has. On this view, a life that is totally devoid of emotional content might well be good for the person who lives it, and so might an extremely fragmentary life, e.g. a life that consists in a sequence of disconnected pleasant sensations. The theory also implies that “the life of a pig” may well be better for a person than “the life of a human being”, i.e. that it does not really matter (from an evaluative point of view) whether a life is recognizably human or not. These implications are not acceptable, however, and pure hedonism therefore has to be rejected.

The idea is that pleasant sensations are too episodic, too tied to experiences of specific activities or conditions, to be identifiable with the most valuable kind of happiness. Here is a more detailed version of this argument, given by Nagel (1970):

Suppose an intelligent person receives a brain injury that reduces him to the mental condition of a contented infant, and that such desires as remain to him can be satisfied by a custodian, so that he is free from care. Such a development would be widely regarded as a severe misfortune, not only for his friends and relations, or for society, but also, and primarily, for the person himself. (p. 5)

The pure hedonist would not necessarily regard the injury as a “misfortune for the person himself”, however. On his view, it is not necessarily bad for a person to be “reduced to the mental condition of a contented infant”; it all depends on whether there happens to be more pleasure in his life after the injury or not. This is what a pure hedonist might say:

If we did not pity him then [when he was three months old], why pity him now; in any case, who is there to pity? The intelligent adult has disappeared, and for a creature like the one before us, happiness consists in a full stomach and a dry diaper. (ibid., p. 6)¹

¹ It is worth noting that Mill (1863) gives a similar objection to pure hedonism, viz. the following one: Most of us would not consent “to be changed into any of the lower animals, for a promise of the fullest allowance of a beast’s pleasures”, or “to be a fool”, or “an ignoramus”, or “selfish and base”, even if we would be “persuaded that the fool, the dunce, or the rascal is better satisfied with his lot” than we are with ours (Mill, 1863, ch. II, p. 8). In

All this suggests that pure hedonism should be rejected, and that we should (instead) shift our attention to the more global attitude of life satisfaction.

A case against the cognitive version of the happiness theory

The big problem with the cognitive version of the happiness theory is that it doesn't attribute any final value to *feeling* happy. Imagine a person who is perfectly satisfied with most or all the conditions of her life, but who does not feel happy at all. Does this person have a good life? Is her life as good as the life she would have had if she felt happy as well? I think not. This does not in any way suggest that the value of evaluating one's life in a positive way must lie exclusively in the pleasure it produces, however, only that a situation cannot be really good for a person unless it includes pleasure as a component.

The cognitive version of the happiness theory also suggests that a person's life can get better even if there is a decrease in his hedonic level, viz. if he himself comes to evaluate his life in a more positive manner. This is another reason why we should reject the cognitive version.

A case for the hybrid version of the happiness theory

This leaves us with two versions of the happiness theory, namely the mood version and the hybrid version. In my view, we should prefer the latter version. The main reason for this is that by taking the subject's own evaluation of her own life into account, it is more consistent with "the idea of the sovereign subject".

One of the main "reasons" why the desire theory seems intuitively plausible to many people (at least to many "modern westerners") is that it appeals to "the idea of the Sovereign Subject" (the phrase is from an earlier version of Scanlon 1993), or what

short, "[i]t is better to be a human being dissatisfied than a pig satisfied; better to be a Socrates dissatisfied than a fool satisfied" (ibid., p. 9). However, this argument can only be regarded as a version of the argument just given if it can be assumed that a pig's (etc.) level of satisfaction is dependent on one thing only, viz. on what how much pleasant *sensation* there is in his life. What the argument shows us (if successful) is that a life which contains nothing but lower pleasures cannot be a good life (or more specifically, a good *human* life) for the person who is living it, or alternatively, that a human life is not a good human life unless it also contains higher pleasures, e.g. pleasures stemming from successful use of our "higher", characteristically human, capacities (cf. Tranøy 1973).

Harsanyi (1982) calls “the Principle of Preference Autonomy”, i.e. the idea that “in deciding what is good and bad for a given individual, the ultimate criterion can only be his own wants and his own preferences” (p. 55).

But *why* should we accept this idea? Well, consider an opposing view, such as the idea that it has final value for us to feel pleasure or to be engaged in creative activity. Now, ask yourself: Do you really think it would be good for a person to feel pleasure, or to engage in creative activity, if he himself did not evaluate these things in a positive manner, e.g. if he strongly preferred to perform routine tasks? I think not. In any case, we can safely conclude that it is (at least under normal circumstances) better to feel good *and* to evaluate one’s life in a positive way than to merely feel good. So from now on, the hybrid conception of happiness will be taken for granted. This is clearly the view which has most moral and rational significance.

Is the pure happiness theory a plausible theory?

It has now been argued that the pure happiness theory is most plausible if it incorporates the hybrid view. So the question arises: Is the resulting version of the pure happiness theory a reasonable theory of the good life?

Arguments for the pure happiness theory

What the pure happiness theorist needs to establish is (roughly) that happiness and unhappiness have a very special evaluative status. The challenge is not to show that it is good for us to be happy, but to convince us that there is nothing else besides happiness and unhappiness that has positive or negative final value for a person, i.e. that all alternative theories of prudential value should be refuted. So, how can the pure happiness theorist argue for this claim?

1. Arguments against alternative theories of well-being are sometimes of the following kind: The happiness theorist takes some alternative theory (e.g. the belief that it has final value for us to have intimate relationships) and tries to explain its occurrence in terms of happiness. The reason why some people tend to regard it as intrinsically good for us to have intimate relationships is (on this type of view) that it tends to make creatures like us happier; if it would not make us happier, we would probably not regard

it as good at all.² Arguments of this type are far from conclusive, but this does not mean that they are bad. In any case, it is clear that such arguments might be highly effective, or more specifically, if a person would come to believe that all his intrinsic evaluations can be explained in this way, then it is likely that he would convert to the pure happiness theory.

2. Arguments of type (1) are often supplemented by arguments of the following type: The happiness theorist considers an opposing view, such as the idea that it is intrinsically good for people to have intimate relationships. He then asks (rhetorically): “Would it be good for a person to have intimate relationships to others, even if this would not make him any happier at all?” He gives a negative answer to this question, and then concludes that the value of having intimate relationships must lie in the happiness it produces, i.e. that its value is merely instrumental.

As Parfit (1984) points out, “[t]his reasoning assumes that the value of a whole is just the sum of the values of its parts” (p. 501). If we do not make this atomistic assumption, the alleged fact that a certain whole (e.g. a-person-P-is-happy-because-he-wants-to-have-an-intimate-relationship-and-now-he-has-it) cannot have final value for P unless it contains happiness as a part does not give any support to the idea that the value of the whole *resides* in this part. This assumption is not plausible, however, and we therefore have to conclude that (2) is, at least *qua* argument for the pure happiness theory, a bad argument.

The other premise of the argument is not implausible, however, since it may well be the case that a situation cannot have final value for a person unless it includes happiness as a component. If we reject atomism but accept this premise, we might end up with the view that “what is best for people is a composite”. It is not just their being

²Cf. Smart’s (1973) arguments for pure hedonism and against modified hedonism: (i) The reason why some of us (e.g. Mill) believe that complex and intellectual pleasures are intrinsically better than sensual and simple pleasures is that the former are often fecund while the latter do not only lack fecundity, but are actually (often) the reverse of fecund (cf. pp. 17-18). (ii) The reason why some of us (e.g. Moore) tend to think that sadistic pleasure has “no intrinsic value at all, or perhaps even a *negative* intrinsic value” (ibid., p. 25) is that we feel a distaste for the consequences of sadism, and not an immediate distaste for sadism as such: “Our repugnance to the sadist arises, naturally enough, because in our universe sadists invariably do harm” (ibid., p. 25). In both cases, the modified hedonist is (somehow) characterized as confused, conceptually or psychologically: The reason why he is not a pure hedonist is either that he blurs the distinction between final and instrumental values, or that he is bad at introspection.

in a happy state, “[n]or is it just their having knowledge, engaging in rational activity, being aware of true beauty, and the like. [...] We might believe that if we had *either* of these, *without the other*, what we had would have little or no value” (cf. *ibid.*, p. 502), i.e. that the good life consists in being happy “for the right reason”.

3. The pure happiness theorist can also argue as follows: “Let us (again) consider the idea that it has final value for us to have intimate relationships. This view implies that there can be an increase in a person’s well-being even if there is a decrease in his happiness level, e.g. if there is a large enough increase in the level of intimacy in his relationships. This is implausible, however, and the alternative view should therefore be rejected. Moreover, it seems that whatever alleged value we replace intimate relationships with, we get the same result: There simply is no thing such that a decrease in a person’s level of happiness can be compensated by an increase in this thing”.

This argument might strike some people as convincing, but it is not as solid as it may seem. First, it is not really an argument for the pure happiness theory. It is true that the idea that every decrease in happiness is also a decrease in well-being implies that the pure versions of the opposing views must be wrong, but the idea is not inconsistent with all the opposing views, e.g. it is perfectly consistent with the idea that what is best for people is to be happy “for the right reason”. Moreover, it is doubtful whether the argument can convince anyone who isn’t already a pure happiness theorist. Consider someone who has a strong intrinsic desire to learn more about himself, and suppose he learns something about himself that makes him a little less happy. This does not necessarily make his life worse for him, especially not if he himself intrinsically prefers unhappy knowing to happy ignorance.

Arguments against the pure happiness theory

What the arguments *against* the pure happiness theory purport to show is not that happiness is a bad thing, but that happiness is not all that matters, or more specifically, that the value of a life for the person who lives it does not just depend on what its experiential content is like, i.e. on how much happiness and unhappiness this life contains. Most of these objections aim to establish a more general point as well, viz. that the prudential value of a life does not just depend on how it feels from the inside, i.e. that all kinds of mental state theories leave something important out. This is not all that

these arguments purport to establish, however. Some of them can also be viewed as arguments for the stronger thesis that a person's well-being can be affected by things he does not know anything about.

Arguments from delusion and deception

The most common type of objection that has been given against the pure happiness theory is the objection from deception. The pure happiness theory claims that a person's well-being consists solely in having the right kind of mental states, i.e. that we can assess a subject's well-being entirely "from the inside", with no reference at all to the actual conditions of her life (cf. Sumner 1996, p. 157). This implies that a totally deluded person, whose happiness is based on misperceptions of the conditions of her life, has a perfectly good life. This is quite counter-intuitive, however, and the pure happiness theory must therefore be rejected.

There are many different versions of this argument in the literature. Here are some of the more well-known examples:

According to [...] [the pure happiness theory and other mental state theories], what affects well-being can only be what enters experience, and the trouble is that some of the things that persons value greatly do not. My truly having close and authentic personal relations is not the kind of thing that can enter my experience; all that can enter is what is common to both my truly having such relations and my merely believing that I do. And this seems to distort the nature of these values. If I want to accomplish something with my life, it is not that I want to have a *sense* of accomplishment. This is also desirable, but it is different from, and less important than, the first desire. [...] If either I could accomplish something with my life but not know it, or believe that I had but not really have, I should prefer the first. That would be, for me, the more valuable life. (Griffin 1986, p. 19)

Imagine a man who dies contented, thinking he has achieved everything he wanted in life: his wife and family love him, he is a respected member of the community, and he has founded a successful business. Or so he thinks. In reality, however, he has been completely deceived: his wife cheated on him, his daughter and son were only nice to him so that they would be able to borrow the car, the other members of

the community only pretended to respect him for the sake of the charitable contributions he sometimes made, and his business partner has been embezzling funds from the company, which will soon go bankrupt. [...] In thinking about this man's life, it is difficult to believe that it is all a life could be, that his life has gone about as well as a life could go. Yet this seems to be the very conclusion mental state theories must reach! For from the "inside" – looking only at the man's experiences – everything was perfect. [...] So if mental states are all that matter, then [...] there is nothing missing from this man's life at all. It is a picture of a life that has gone well. But this seems quite an unacceptable thing to say about his life; it is surely not the kind of life we would want for ourselves. So mental state theories must be wrong. (Kagan 1998, pp. 34-35)

Happiness (or unhappiness) is a response by a subject to her life conditions *as she sees them*. It is a matter of whether she is finding the *perceived* conditions of her life satisfying or fulfilling. But what if her perceptions of important sectors of her life is a misperception? What if she is deceived (by others and by herself) about them? Suppose, for instance, that her happiness depends in part on the loyalty and affection of a partner who is in fact merely using her for his own purposes. When she discovers the truth she will, of course, be miserable. But what are we to say of those months or years during which she was deluded? She was certainly happy then, but was her life going well for her? (Sumner 1996, p. 156)

The same kind of deception is vividly captured in Rachels' (1986) story of *Wonmug*, and in the movie *Truman Show*. Here, Truman's wife and his best friend are both actors, and Truman himself is totally ignorant about this circumstance. However, the most famous objection of this type is probably Robert Nozick's "Experience Machine" (Nozick 1974, 1989). Nozick asks us to imagine a machine that could give us any experience that we desired. Being plugged into the machine, one would all the time be floating in a tank, with electrodes attached to one's brain, and the machine would stimulate the brain in any way pre-programmed. While in the tank, one would not know that one is there; one would think that it is all actually happening. Nozick now asks us (rhetorically) whether we would plug into this machine for life. For most of us, the answer is "no", and the reason for this is (of course) that "something matters to us in

addition to experience” (Nozick 1974, p. 44), or alternatively, that we “care about things in addition to how our lives feel to us from the inside” (Nozick 1989, p. 104).

We care about more than just how things feel to us from the inside; there is more to life than feeling happy. We care about what is actually the case. We want certain situations we value, prize, and think important to actually hold and be so. [...] We want to be importantly connected to reality, not to live in a delusion. [...] What we want and value is an actual connection with reality. [...] To focus on external reality, with your beliefs, evaluations, and emotions, is valuable *in itself*, not just as a means to more pleasure or happiness. [...] We do not, of course, simply want contact with reality; we want contact of certain kinds: exploring reality and responding, altering it and creating new actuality ourselves. [...] No doubt, too, we want a connection to actuality that we also share with other people. One of the distressing things about the experience machine, as described, is that you are alone in your particular illusion. (Ibid., pp. 106-107)

Some arguments from deception do not just purport to show that there is more to well-being than happiness, however, they also aim to establish an even stronger claim, viz. that a person’s well-being can be directly (and not just causally) affected by things he doesn’t know anything about, e.g. that one can be directly (“intrinsically”) harmed by what one does not know.

[On the happiness theory] [...] what you don’t know can’t hurt you. It means that even if a man is betrayed by his friends, ridiculed behind his back, and despised by people who treat him politely to his face, none of it can be counted as a misfortune for him as long as he does not suffer as a result. (Nagel 1970, p. 4)

Loss, betrayal, deception, and ridicule are on this view bad because people suffer when they learn of them. But it should be asked how our ideas of human value would have to be constituted to accommodate these cases directly instead. One advantage of such an account might be that it would enable us to explain *why* the discovery of these misfortunes causes suffering [makes us unhappy] – in a way that makes it reasonable. For the natural view is that the discovery of betrayal makes us

unhappy because it is bad to be betrayed – not that betrayal is bad because its discovery makes us unhappy. (ibid., p. 5)

What makes Nagel's claim stronger than the claims in the arguments above is the idea that it is positively bad to be e.g. ridiculed behind one's back. Above, the claim was merely that happiness based on deception is not as good as happiness based on accurate perception. It is important to see that this argument does not presuppose any particular view on *why* it is intrinsically bad for a person to be betrayed, deceived, or ridiculed behind his back, however, e.g. whether it is bad because the person has an intrinsic aversion against it (as the desire theorist would say), or because it is simply bad for him (as an objective list theorist might say). That is, as the argument stands, we only know that it presents a serious threat to the pure happiness theory, i.e. we do not know what alternative theory of prudential value that is supported by this argument.

Arguments from defective preferences or standards of evaluation

We have now looked at a number of arguments that purport to show that experience is not all that matters. Most or all of these arguments refer to delusional beliefs, and they make use of the circumstance that a person's happiness might be based on false beliefs about the conditions of his life, in such a way that if the person came to know what his life is really like, he himself would be devastated. Let us now turn to a different category of arguments, a category that might be labelled "happy slave arguments". Here, there is no reference what so ever to false or delusional *beliefs*. The problem is rather connected to the *preferences* or *evaluative standards* which make the happy slave or subjugated housewife content with respective lots.

On the hybrid version of the pure happiness theory, happiness is, in part, a matter of perceived desire satisfaction, i.e. to be happy is, in part, to endorse one's life as one perceives it. Above, we saw that a happy life might not be a good life if this endorsement is based on grave misperceptions. Here, the idea is rather that a happy life might not be prudentially good if the endorsement in question is based on false or irrational standards of evaluation, e.g. that it is acceptable or even desirable to be enslaved, ridiculed, betrayed, despised, or slandered. Let us now take a closer look at three different arguments of this type.

1. The following well-known argument purports to show that our adaptive preferences should be regarded as irrelevant (or less relevant) in this context. That is, if a person is happy because his positive evaluation of his own life is based on adaptive preferences, then his happiness is not as valuable as it would otherwise have been, it might even have no value for him at all.

Let us first consider the idea that autonomous and other non-adaptive desires should (*in the context of interpersonal comparisons*, e.g. in the context of social justice) have more weight than adaptive desires. Both Nussbaum (1990) and Sen (1985) have argued that actual desire is (in this area) “a malleable and unreliable guide to the human good” (Nussbaum 1990, p. 213). Their main reason seems to be that our actual desires are often adaptive desires, and that these desires are of little or no relevance in this context. This criticism is very similar to Elster’s (1983) criticism of the “actual preference utilitarian” (who regards actual desire satisfaction as “the criterion of justice and social choice”). Elster writes:

[W]hy should the choice between feasible options only take account of individual preferences if people tend to adjust their aspirations to their possibilities? For the [preference] utilitarian, there would be no welfare loss if the fox were excluded from consumption of the grapes, since he thought them sour anyway. But of course the cause of his holding them to be sour was his conviction that he would be excluded from consuming them, and then it is difficult to justify the allocation by invoking his preferences. (p. 109)

This is Nussbaum’s (1990) objection to the idea that what is good for people is (in this interpersonal context) “a function of the satisfaction of desires or preferences that people happen, as things are, to have”:

The central difficulty with this proposal is [...] that desire is a malleable and unreliable guide to the human good, on almost any seriously defensible conception of good. Desires are formed in relation to habits and ways of life. At one extreme, people who have lived in opulence feel dissatisfied when they are deprived of the goods of opulence. At the other extreme, people who have lived in severe

deprivation frequently do not feel desire for a different way, or dissatisfaction with their way. Human beings adapt to what they have. In some cases, they come to believe that it is right that things should be so with them [they may even be *happy* with their lot]; in other cases, they are not even aware of alternatives. Circumstances have confined their imaginations. So if we aim at satisfaction of the desires and preferences that they happen, as things are, to have, our distributions will frequently succeed only in shoring up the status quo. (p. 213)

This is how Sen (1985) formulates the same argument (notice that he explicitly has the context of interpersonal comparisons in mind):

Comparative intensities of desire [...] are influenced by many contingent circumstances that are arbitrary for well-being comparisons. Our reading of what is feasible in our situation and station may be crucial to the intensities of our desires, and may even affect what we dare to desire. Desires [and standards of evaluation] reflect compromises with reality, and reality is harsher to some than to others. The hopeless destitute desiring merely to survive, the landless laborer concentrating his efforts on securing the next meal, the round-the-clock domestic servant seeking a few hours of respite, the subjugated housewife struggling for a little individuality, may all have learned to keep their desires [and standards of evaluation] in line with their respective predicaments. Their deprivations are gagged and muffled in the interpersonal metric of desire fulfillment [or happiness]. In some lives small mercies have to count big. (pp. 190-191; the square brackets are all my own additions)

This is quite a good argument in the context in which it really belongs, i.e. in the context of interpersonal comparisons. How happy a person is with his lot depends on his standard of evaluation, which is (in turn) causally dependent on his beliefs about his own possibilities, and this suggests that we should reject the idea that it is happiness that should be maximized or justly distributed. However, the alleged fact that the pure happiness theory is (for the reason just cited) no good in the context of *interpersonal* comparisons does not allow us to conclude (as Nussbaum seems to think) that adaptive happiness is less valuable than autonomous happiness in the *intrapersonal* context. If an unhappy slave could be turned into a happy slave, this might well constitute an

improvement in well-being. Or would it? Consider the following two arguments (both from Brülde 1998).

2. *HAPPY*. Think of some external or relational circumstance that you strongly want (intrinsically) to be a part of your life. It may be to be with a loved person, to have an intimate and reciprocal relationship, to engage in some creative activity, to act morally, to develop as a person, to achieve something important, or the like. For simplicity's sake, call this intrinsically desired situation X.

Now, imagine that you accept the pure happiness theory. You then believe that X can only be good for you in the instrumental sense, i.e. because it contributes to your happiness (this is so, even if you happen to desire it intrinsically). Let us also assume that you are now reasonably happy, partly because X is the case. Now, imagine that you would be offered (for free) a harmless drug called HAPPY. If you take HAPPY, you will be happier than you are now, but you will also lose interest in X. X will not concern you anymore, you will simply stop wanting that X obtains, and this is also the reason why you will become happier. Would you have a better life if you take HAPPY than if you don't?

The same story can be repeated for external conditions that you think are bad for you. This could be being totally deceived, being enslaved, being totally passive or utterly lonely. Call this bad thing Y. Here, taking HAPPY will change your attitude toward Y in a special way. It will make you start liking Y, and as a result of this, you will become happier than you would otherwise have been. In this case, would your well-being increase if you take HAPPY?

If you say no to HAPPY³, you are not really a pure happiness theorist: the experiential dimension is not the only dimension you really care about. Most of us want more than happiness, however, we also want to be happy “for the right reason” or “on the right grounds”. We want, I think, certain kinds of wholes, where happiness is only one (let alone essential) component. And the reason why we want this is that we have realized that the pure happiness theory is an implausible theory.

3. Suppose that someone intrinsically desires to be betrayed, manipulated, slandered, deceived, and to suffer great pain, and that it makes him happier to have

³*Qua* self-interested, that is; it is of course possible to say no to HAPPY for other (e.g. moral) reasons; just as it is possible to refuse to plug into the experience machine for moral reasons.

these desires fulfilled. Does it really have final value for him to be happy for these reasons? It can be argued that none of these things are good for him, either because the preferences on which his happiness is based is deeply irrational, or because the objects of these desires are objectively bad for him.

Conclusion

To conclude, there are a number of strong arguments against the pure happiness theory, and this theory should therefore be rejected. This is not to deny that happiness is important, however. What the arguments show is “merely” that there is more to well-being than happiness. That is, it seems that we need to modify the happiness theory in such a way that it can avoid the objections above, e.g. by adding other final values (besides happiness) or by claiming that happiness only counts as well-being if its cognitive component satisfies certain epistemic or other requirements.

Possible modifications of the pure happiness theory

Redefinitions of happiness no solution

Before we look at how this might be done, let us first note that there is another possible way to respond to the objections above, namely to redefine the concept of happiness in a way that preserves the identity between happiness and well-being. On this type of view, the arguments above do not really show that there is more to well-being than happiness, but rather that there is more to happiness than having the right sort of mental states (cf. Kagan 1998, p. 35).

Some philosophers do not rest content with a purely psychological concept of happiness, they have also claimed that there are objective criteria for happiness. The most common example of such an objective criterion is the idea that a person cannot be really or truly happy unless her positive evaluation of her life is based on beliefs about the world which are true or at least justified (cf. Sumner 1996, p. 157). Tatarkiewicz (1976) is an example of a philosopher who claims that a person is not really happy unless the satisfaction involved is (objectively) justified.

Another objective criterion that is sometimes included in the very concept of happiness is the idea that something can contribute to a subject's (true) happiness only if

it is *independently valuable* (cf. Sumner 1996, p. 163), and that it is (for this reason) not conceptually possible to be happy in any circumstances. Nozick (1989) adopts a weaker form of this requirement. On his view, “[w]e would be reluctant to term someone happy at a particular moment or in life in general if we thought the evaluations upon which his emotion was based were wildly wrong. Yet it would be too stringent to require that the evaluations be correct” (p. 111).

To include the idea that a person’s positive evaluation of her own life is based on true (or justified) beliefs in the very analysis of “happiness” is quite implausible, however. This is simply not how we use the term. Instead, we should stick to a purely psychological notion of happiness, i.e. we should regard it as a kind of mental state, a state that is dependent on how we see our lives and not necessarily on how these lives really are. This subjective notion of happiness is not merely descriptively adequate, it also allows us to distinguish justified from unjustified happiness. Moreover, there is good reason to believe that the different attempts to define happiness in objective terms are really based on evaluative considerations. What these theorists really try to say is most probably that psychological happiness is not sufficient for well-being, e.g. that it is not as valuable for a person to be happy if his happiness is based on false or illusory beliefs.

To conclude, we should simply accept that happiness is a state of mind and give up the simple identity between happiness and well-being (cf. Sumner 1996, p. 158). What the objections above show is not that there is more to happiness than a mental state, but that there is more to well-being than happiness. Kagan’s business man or the person in the experience machine might not have very good lives, but they are indeed happy.

Other possible modifications of the pure happiness theory

The possible modifications of the happiness theory all concern the cognitive component of happiness, not the affective component. To repeat, the cognitive aspect of person’s happiness consists in a positive evaluation of the conditions of her life as she herself perceives them, i.e. a judgement that her life (as she sees it) measures up favourably against her own standards or expectations. In a similar manner, a person is cognitively unhappy if she evaluates her life as a whole, as she herself conceives of it, in a negative

way, i.e. if she, on balance and taking everything into account, thinks her life is going badly for her (cf. *ibid.*, p. 145).

There are at least two kinds of possible modifications of the happiness theory. The first possibility is to make the value of a person's happiness depend on whether or not his *perception* (or conception) of her own life is epistemically appropriate or acceptable, i.e. whether the *beliefs* on which her positive evaluation is based are true, justified, informed, or the like. On this type of view, a state of happiness that is based on true or justified beliefs is (other things being equal) more valuable for the subject than a happy state that is based on false or unjustified beliefs. The purpose of this move is to meet the objection from delusion or deception.

The second possible modification is to make the value of a person's happiness depend on whether *the evaluative standard* on which her positive evaluation is based satisfies certain requirements, i.e. whether the standard which she uses to assess her life is true, justified, autonomous, or the like. On this type of view, a happy state based on acceptable preferences makes a larger contribution to the subject's well-being than a happy state that is based on a defective evaluative standard. The purpose of this move is to avoid the "happy slave" type of objections.

Epistemic requirements

There are at least three possible modifications of the first type, modifications that differ with regard to what epistemic condition a person's *beliefs* about her own life must satisfy if her happiness is to count as maximally valuable.

The first possible modification consists in requiring that a person's positive evaluation of her own life is based on true beliefs about her own life, i.e. that she is in touch with reality, that she perceives her life as it actually is. According to this *truth* or *reality requirement*, "happiness counts as well-being only when it is based on a view of the conditions of our lives which is free from factual error" (*ibid.*, p. 158). On the strongest version of this view, happiness has no prudential value at all unless fully informed. There is also a weaker (and more plausible) version, however, according to which happiness based on reality makes a larger contribution to well-being than happiness based on factual error, i.e. a view which discounts the value of happiness at a steady rate as it becomes less informed (cf. *ibid.*, p. 161). It is worth noting that this

requirement is heavily inspired by the desire theory, which claims that what has prudential value for us is that our desires are in fact satisfied, not just that we believe that they are. That is, a happiness theory which incorporates this condition is really a mixture between the pure happiness theory and the desire theory.

The second possibility is to require that a person's positive evaluation of her own life is based on rational, justified or reasonable beliefs about her own life. According to this "*justifiability requirement*", happiness counts as well-being only when it is based on a view of the conditions of our lives which is reasonable under the circumstances (given the information available), or the like. On the strongest version of this view, a person's happiness has no value at all unless it is based on rational beliefs about her life, and on the weaker version of the same view, a happy state based on rational beliefs makes a larger contribution to well-being than a happy state based on irrational beliefs.

A third possibility is formulated and adopted by Sumner (1996), who thinks that the two requirements just mentioned are both too strong. On Sumner's view, all we should require is that the affirmation or endorsement of one's life should be "informed" or "based on a clear view" (p. 160). But what is meant by "a clear view" here, i.e. how well informed must a subject be if her happiness is to count also as her well-being (given that the two requirements above are both too strong)? At this point, Sumner appeals to the subject's own epistemic preferences. As he himself puts it, "[t]he relevance of information for a person's well-being is a personal matter to be decided by personal priorities" (ibid., p. 161). "The extent to which the illusoriness of the experiences *matters* for an individual's well-being [...] depends on the extent to which she decides (or would decide) to *make* it matter" (ibid., p. 161, footnote 25). But what is this supposed to mean? Suppose that someone's endorsement of his life is in fact uninformed or illusory, e.g. that his happiness depends in part on the loyalty and affection of a partner who is in fact merely using him for her own purposes. How would Sumner assess this person's well-being? In Sumner's own words, this depends on "whether the endorsement will, or would, survive the acquisition of the missing information" (ibid., p. 161), i.e. he suggests that the crucial test consists in how the person would respond in retrospect, if he would recognize the illusion. As far as I can tell, this implies that if the deluded man would not in any way regret what happened

when he discovers the truth, we have to admit that his life was really good during the time he was deluded.

This is but one reason why Sumner's view leaves a lot to be desired: It is doubtful whether his theory can really avoid the objection from deception. The reason for this is that it is unclear whether Sumner's information requirement makes the resulting theory of well-being a state-of-the-world theory, i.e. a theory that doesn't just makes well-being dependent on states of mind, but also on states of the world (as he himself claims; cf. *ibid.*, pp. 174-175). The fact that most endorsements that are informed in Sumner's sense are also "based on fact" does not make Sumner's theory a state-of-the-world theory. Moreover, the state of the world only plays a very indirect role in Sumner's reasoning, viz. as a part of his "retrospective test". For him, the crucial test is how one would respond in retrospect, if one would learn the truth. To introduce the idea of truth in this roundabout and hypothetical manner does not make the resulting theory a state-of-the-world-theory. It can also be questioned whether the retrospective perspective to which Sumner refers is as privileged as he thinks it is. It can even be doubted whether Sumner's appeal to the subject's own epistemic preferences is an epistemic requirement at all.

In my view, it is not really possible to avoid the argument from delusion unless one accepts the truth requirement, or alternatively put, unless one adopts the desire theoretical idea that a person's quality of life is dependent on whether his life is *actually* going the way he wants it to go. Sumner conceives of this view as "arrogant", "unreasonably puritanical" and "presumptuously dogmatic", and he offers two arguments for this position. First, he appeals to the fact that "[w]e do not invariably reassess earlier periods of happiness in this austere manner once we realize the extent to which they depended on false beliefs about states of the world" (*ibid.*, p. 158). But again, why is this retrospective perspective so privileged? Second, he thinks the reality requirement should be rejected because it suggests that e.g. comforting fantasies do not make any contribution at all to our well-being (*ibid.*, p. 159). If successful, this argument only hits the strong version of the truth requirement, however, i.e. it is not as unreasonable to suggest that happiness based on true beliefs makes a larger contribution to well-being than comforting fantasies.

It is true that Sumner's appeal to the subject's own epistemic preferences is (in a way) an appeal to individual sovereignty, but his idea that the reality requirement is inconsistent with the sovereignty of the individual is not very plausible. On the contrary, it might be argued that the reality requirement is actually *more* in line with the idea of the sovereign subject (cf. above) than Sumner's requirement. The reason for this is that reality requirement suggests that we match a person's actual existence against *her own* actual standards and preferences, i.e. that we take the person's own desire to live a certain kind of life seriously when we assess the prudential value of this person's life. After all, what the subject herself values is a certain kind of life, so why shouldn't we make use the person's own evaluative standard when we try to determine to what extent her happiness contributes to her well-being? In my view, this is true subjectivism.

I agree with Sumner that the justifiability requirement is implausible, but we reject this requirement for somewhat different reasons. In Sumner's view, this requirement is invalid because it "presumes to dictate to individuals how much their deviations from an ideal epistemic standpoint should matter to them. But that is for them to decide" (ibid., p. 159). In my view, the justifiability requirement is simply irrelevant. The crucial issue is whether a person's life is actually going the way she wants it to go, not whether her endorsement of her life is based on rational beliefs about what her life is like, especially if we consider the circumstance that our rational beliefs may well be false.

Another important question in this context is whether we should view the negative value of unhappiness in an analogous way. Sumner seems to think so, e.g. when he claims that "[a] life is going badly for someone when she (authentically) experiences its conditions as unsatisfying or unfulfilling, or disclaims or disowns them" (ibid., p. 177). But can we really take this kind of symmetry for granted? For example, is it less bad for me if my unhappiness is inauthentic (uninformed) than if it is authentic (informed)? And does it really matter whether unhappiness is based on true or false beliefs? It might be argued that it is less reasonable to impose epistemic requirements on the negative side of the spectrum, e.g. that uninformed unhappiness is just as bad for the subject as informed unhappiness.

This is how I suggest that the happiness theory can and should be modified in order to deal with the objection from delusion. Let us now look at how the theory can

and should be modified in order to deal with the happy slave arguments, e.g. the problem of the malleability of people's standards of self-assessment.

Do our standards of evaluation have to meet any requirements?

There are several possible modifications of this type as well, modifications that differ with regard to what requirement a person's evaluative standard (or global preferences) must satisfy if her happiness is to count as maximally valuable.

The perhaps most common modification of this kind consists in requiring that a person's positive evaluation of her own life is based on an evaluative standard (or global desire) that is, in some sense, rational. On the strongest version of this *rationality requirement*, a person's happiness has no prudential value at all unless it is based on a rational evaluative standard, i.e. the perceived satisfaction of an irrational global desire does not contribute to the person's well-being at all.⁴ On the weaker (and more reasonable) version of the same requirement, it is *ceteris paribus* better for a person to be happy if his happiness is based on a perceived satisfaction of rational desire than if it is based on the perceived satisfaction of a desire that is in fact irrational. According to this weaker view, the prudential value of a subject's life would be discounted by the extent to which the subject's belief in its value was irrational.

Now, there are several different conceptions of rational desire (cf. Brülde 1998, pp. 236-260), conceptions which give rise to different "senses" in which a desire might be rational or irrational. For example, *the informed theory* claims that a desire is rational if and only if it is informed, where a desire is informed if and only if it is "formed by appreciation of the nature of its object" (cf. Griffin 1986, p. 14). On *the intrinsic theory*, there are desires that are intrinsically rational and irrational, where an intrinsically rational desire is a desire that is rational because of its content, e.g. in virtue of the fact that its object is (in some objective sense) worth desiring. On *the genetic theory*, a desire is rational if and only if it has the right kind of causal history, e.g. if it has not been shaped by "irrelevant causal processes" (e.g. by manipulation). *The deliberative theory* claims that a person's actual desire is rational if and only if it would survive a process of ideal deliberation, e.g. if he would still desire it "if he knew the relevant

⁴ It is important to distinguish this rationality requirement (which concerns our evaluative standards) from the justifiability requirement (which concerns our beliefs about, or our perception of, our own lives).

facts, was thinking clearly, and was free from distorting influences” (cf. Parfit 1984, p. 118).

The fact that there are several different conceptions of rational desire means that we must ask ourselves which of these conceptions that is most relevant in this particular context. Which conception of rational desire gives, if incorporated into a rationality-oriented happiness theory, rise to the most plausible theory?

The informed theory and the intrinsic theory both seem to assume that there are objective prudential values, i.e. that there are situations that are (in some objective sense) worth desiring or worth avoiding, situations that a person would desire or want to avoid if he came to appreciate their “true nature”. This suggests that if any of these two theories are incorporated into the rationality requirement, what we get is really a requirement that a person’s positive evaluation of her own life is based on a *correct* or *true* evaluative standard, i.e. that those conditions of her life that she herself regards as valuable (for her) are objectively valuable. According to the strongest version of this “*value requirement*”, something can contribute to a subject’s well-being (directly and intrinsically) only if “(1) the subject finds it satisfying or fulfilling, or endorses it as an ingredient in her life, and (2) it is independently valuable” (Sumner 1996, p. 163). There is also a weaker (and more plausible) version, however, viz. the idea that it is more valuable for a person when his happiness is based on correct evaluations (when he endorses something that has positive objective value) than when it is based on mistaken evaluations (when he endorses something that has no value or negative objective value).

If we incorporate the genetic theory into the rationality requirement, we get a requirement that a person’s positive evaluation of her own life is based on an evaluative standard that has the right kind of causal history, e.g. a standard that has not been shaped by manipulation. This requirement is quite similar to Sumner’s *autonomy requirement*, according to which a person’s happiness counts as well-being only when the evaluative standard on which her positive evaluation is based is *autonomous*. The reason why this is so is that the autonomy of a desire is often regarded as a matter of how the desire was formed (cf. e.g. Elster 1983, Sumner 1996, pp. 169-170). If this “historical approach” is adopted, we have to conclude that if a person’s happiness is based on expectations or values that have been artificially lowered or distorted by processes of indoctrination or exploitation, then it makes (on this view) less of a

contribution to her well-being than if it is based on values or desires that have been formed under acceptable conditions.

The historical approach suffers from certain weaknesses, however, e.g. it does not allow for the possibility that people sometimes autonomously come to accept values which were initially formed by non-autonomous means (cf. *ibid.*, p. 170). This suggests that the autonomy requirement should not just include the idea that we should disregard values that have been engineered or manipulated by others, but also the idea that a person's happiness has a higher value for her if her endorsement of her own life is based on values and desires that are, in some important sense, her own, i.e. she has identified with these values, or acknowledged them as her own, or endorsed them as her standards (*ibid.*, p. 168). It is worth noting that this additional idea is purely about autonomy, however, i.e. no one has ever suggested that a desire can be rational in virtue of the fact that the desiring subject acknowledges it as her own.

Finally, we might choose to incorporate the deliberative theory into the rationality requirement. To do this is to require that a person's positive evaluation of her own life is based on an evaluative standard that would survive a process of ideal deliberation, e.g. a standard that would not extinguish if the subject "knew the relevant facts, was thinking clearly, and was free from distorting influences". On this view, a person's happiness makes a larger contribution to her well-being if it is based on values that are rational in this sense than if it is based on values that would, in fact, not survive confrontation with clear and repeated judgements about established facts.

Let us now try to find out if any of these views are plausible. For example, is it plausible to require that a person's evaluative standard (or global preferences) is correct if her happiness is to count as maximally valuable? The appeal to objective values is of course based on the assumption that there are such values, i.e. that there are situations that have final value for us, whether or not we want to have the good things, or to avoid the bad things (cf. Parfit 1984, p. 493). This might seem like a rather questionable assumption, but let us (for the time being) assume that it is correct. Would this make the value requirement reasonable? Sumner (1996) thinks not. Even if there were objective values, "a value requirement still seems objectionably dogmatic in imposing a standard discount rate on people's self-assessed happiness" (p. 165). In my view, this is not a reasonable position. Suppose that it is objectively good for all of us to live in intimate

relationships, and that it is objectively bad for all of us to be enslaved. Doesn't this strongly suggest that being satisfied with an intimate relationship is prudentially better than being satisfied with being enslaved? It might even suggest that the happiness of the happy slave makes no contribution to his well-being at all.

It can be doubted whether there are any objective prudential values, however, and that we should therefore refrain from stipulating what standards people should use to assess their lives. There is another way in which "objective prudential values" like love, creativity, or freedom can be taken into account, however. The objectivist view can be taken to suggest that there are a number of objectively important dimensions (or domains) in a person's life, and that the traditional list of relevant objective values can be used to specify these domains, e.g. the interpersonal dimension or the dimension of achievement. A rather weak version of the value requirement can now be formulated (cf. Brülde 1998, p. 377): A person's level of well-being is (roughly) a function of how satisfied he is with his life as a whole, *but only as long as he takes all the objectively important dimensions into account*. Or more specifically, the endorsement which determines how good a person's life is (on the whole) must (so to speak) include how satisfied he is in a number of "objectively pre-determined" areas, and a person's level of satisfaction in the relevant areas must (roughly speaking) be "in line with" the traditional objective values. It is doubtful whether this type of modification can avoid the happy slave objection, however, i.e. it most probably needs to be supplemented by some other criterion.

So, are there good reasons to accept the view that a person's happiness has higher prudential value if the evaluative standard on which her positive evaluation is based is *autonomous*? Can Sumner's autonomy requirement handle the happy slave objection? Well, on this theory, we cannot assume *a priori* that the happy slave has a bad life, or that a life of servility and subservience (a life which we regard as trivial or demeaning or depraved) cannot be prudentially valuable for its subject. However, "we can have pretty good empirical reasons for thinking that [...] [such a life] is rarely embraced under conditions of full [...] autonomy", and therefore not likely to be a good life (Sumner 1996, p. 182).

However, to determine whether or not the autonomy requirement is reasonable, e.g. whether it is too weak or too strong, we first have to know what autonomy is. As we

have already seen, there are two ways to conceive of autonomy in this context. On the first view (the genetic approach), a desire or standard of evaluation is autonomous only if it has been formed under acceptable conditions, e.g. if it has not been artificially shaped by processes of manipulation, indoctrination or exploitation. On the second view, a desire or standard of evaluation is autonomous only if the subject identifies with it, or acknowledges it as her own.

The first view is clearly too strong, partly because it cannot account for the fact that people sometimes come to accept autonomously values which were initially formed by non-autonomous means (cf. above). The genetic theory contains valuable insights that ought to be preserved, however. In my view, these insights are better expressed in terms of the deliberative theory of rational desire. On this theory, a desire or evaluative standard can only be rational if it would survive a confrontation with clear and repeated judgements about established facts. This implies that if a certain evaluative standard is causally dependent on (maintained by) false beliefs or “ignorances”, then we have to regard it as irrational. If we incorporate this view into the rationality requirement, we get the idea that a person’s happiness makes less of a contribution to her well-being if it is based on values that are maintained by false beliefs (etc.) than if it is based on values that would survive a process of ideal deliberation. This view accounts for the circumstance that we often want to give less weight to values which have been formed by indoctrination or exploitation, but it also allows for the possibility that we should sometimes give weight to certain values even though they were initially formed by non-autonomous means, namely if the subject has come to accept these values autonomously *after* they were formed.

The problem with the second view is that it seems superfluous as a component in a theory of well-being. It seems reasonable to assume that a person’s happiness has higher prudential value if her endorsement of her own life is based on values and desires with which she identifies, or acknowledges as her own. But this may simply be due to the fact that she tends to attribute more value to a certain situation if this is the case. It is not very easy to find examples of endorsements based on evaluative standards which the subject herself does not want to have, but here is possible example. Suppose that a certain male can only regard a relationship with a woman as good for him if the woman in question is beautiful in a traditional or conventional sense. Suppose also that he does

not acknowledge this evaluative standard as his own, that he even wishes he could get rid of it. Now suppose that he actually has a conventionally beautiful girlfriend, and that he is happy as a result. In this case, his endorsement is based on a standard that is not “his own”, but I see no reason why we should regard his happiness as less prudentially valuable because of this very circumstance. The circumstance that he does not identify with his own standard is only relevant in an indirect manner, viz. if it makes him appreciate his relationship less.

To conclude, my suggestion is that it is intrinsically better for a person to be happy if the evaluative standard on which her happiness is based satisfies two different requirements. First, the evaluative standard should have a certain scope, it should include a number of “objectively pre-determined” areas. On this view, a person’s happiness makes a larger contribution to her well-being if her endorsement of her own life is based on an evaluative standard which takes all the objectively important life domains into account. Second, we should also require that the evaluative standard is *not* causally maintained by false beliefs or “ignorances”. On this view, a person’s happiness makes a larger contribution to her well-being if her endorsement is based on an evaluative standard that would survive confrontation with all the relevant factual information. In my view, no further requirements are necessary, e.g. we must not assume (with Sumner) that it has less prudential value for a person to be happy if her endorsement is based on an evaluative standard that she does not acknowledge as her own.

Finally, let us just note that it is far from obvious that we should view the negative value of unhappiness in an analogous way. For example, we cannot automatically assume that it is worse for a person to be unhappy if his unhappiness is based on a rational evaluative standard than if it is based on an irrational evaluative standard (or vice versa). It seems less reasonable to impose any requirements on people’s evaluative standards if we have the negative side of the spectrum in mind. In short, the pure (unmodified) happiness theory seems to be far more plausible if it is viewed as a theory of the bad life than if it is regarded as a theory of the good life.

Conclusion

Three major questions have been addressed in this paper. I first asked what version of the pure happiness theory that is most plausible, and argued that happiness has most prudential value if we accept the hybrid conception of happiness. I then asked whether this (i.e. any) version of the pure happiness theory is a reasonable theory of well-being, and argued that it is not. The third question concerned how the pure happiness theory can and should be modified in order to avoid the most serious objections directed against it. I first examined different types of epistemic modifications, and arrived at the conclusion that we should accept the reality requirement, i.e. that a person's happiness is more valuable for her if it is based on a correct perception of her own life. I then examined a number of modifications which make the value of a person's happiness depend on whether the evaluative standard on which her positive evaluation is based satisfies certain requirements. Here, I argued that a person's happiness has higher prudential value if the evaluative standard on which it is based takes all the objectively important life domains into account, and if it (moreover) is *not* causally maintained by false beliefs or "ignorances". I also pointed out that it is far from obvious that we need any of these modifications in order to determine how prudentially bad it is to be unhappy. The pure (unmodified) happiness theory is not a plausible theory of the good life, but it may well be a plausible theory of the bad life.

Bengt Brülde

Department of Philosophy, Göteborg University

Box 200, 405 30 Göteborg

Sweden

bengt.brylde@phil.gu.se

References

Barrow, R. (1980). *Happiness and Schooling*. New York: St. Martin's Press.

Brülde, B. (1998). *The Human Good*. Göteborg: Acta Universitatis Gothoburgensis.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford:

Clarendon Press.

- Harsanyi, J. (1982). Morality and the Theory of Rational Behaviour. In: A. Sen and B. Williams (eds.), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
- Haybron, D.M. (2001). *Happiness and Ethical Inquiry: An Essay in the Psychology of Well-Being*. New Brunswick, NJ: Brunswick Rutgers, the State University of New Jersey.
- Kagan, S. (1998). *Normative Ethics*. Boulder, Col.: Westview Press.
- Mill, J. S. (1863/1980). Utilitarianism. In: *Utilitarianism, On Liberty and Considerations on Representative Government*. London: Dent & Sons.
- Nagel, T. (1970/1979). Death. In: *Mortal Questions*. Cambridge: Cambridge University Press.
- Nordenfelt, L. (1991). *Livskvalitet och Hälsa: Teori och kritik*. Falköping: Almqvist & Wiksell.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Nozick, R. (1989). *The Examined Life: Philosophical Meditations*. New York: Simon & Schuster.
- Nussbaum, M. (1990). Aristotelian Social Democracy. In: R.B. Douglass, G. Mara and H. Richardson (eds.), *Liberalism and the Good*. New York: Routledge.
- Parfit, D. (1984). *Reasons and Persons*. New York: Oxford University Press.
- Rachels, J. (1986). *The End of Life: Euthanasia and Morality*. Oxford: Oxford University Press.
- Scanlon, T. (1993). Value, Desire, and Quality of Life. In: M. Nussbaum and A. Sen (eds.), *The Quality of Life*. Oxford: Clarendon Press.
- Sen, A. (1985). Well-being and Freedom. *The Journal of Philosophy* 82, pp. 185-203.
- Smart, J.J.C. (1973). An Outline of a System of Utilitarian Ethics. In: J.J.C. Smart and B. Williams (eds.), *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Sumner, L.W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- Tatarkiewicz, W. (1976). *Analysis of Happiness*. The Hague: Martinus Nijhoff.
- Tranøy, K. E. (1973/1988). Mill's Pig and the Quality of Pleasure. In: *The Moral Import of Science: Essays on Normative Theory, Scientific Activity and Wittgenstein*. Sørensgrend: Sigma.
- Veenhoven, R. (1984). *Conditions of Happiness*. Dordrecht: D. Reidel.