

## The Possibility of Modified Hedonism

by

SANDY BERKOVSKI

Bilkent University

---

*Abstract:* A popular objection to hedonist accounts of personal welfare has been the experience machine argument. Several modifications of traditional hedonism have been proposed in response. In this article I examine two such responses, recently expounded by Feldman and Sumner respectively. I argue that both modifications make hedonism indistinguishable from anti-hedonism. Sumner's account, I claim, also fails to satisfy the demands of theoretical unity.

*Keywords:* welfare, hedonism, pleasure, experience machine

### 1. Introduction

IT IS EASY TO SEE why hedonism about personal welfare, in its standard form, held sway over so many theorists. Consider the way we think about health. In the world  $w$ , Silvio has developed cancer not long before his death. Tumours grew, but no symptoms were detectable. Silvio's physical abilities were the same as in the world  $w'$  where his cancer did not develop. In both worlds he dies in a car accident, being totally ignorant of his illness. Now, health, being an important good to have, must bear on the level of welfare. The presence of tumours should make a person less healthy. Therefore, in the world  $w$  Silvio was less healthy than in  $w'$ . One might then expect Silvio to have less welfare in  $w$  than in  $w'$ . But this seems false. As long as there are no symptoms of an illness revealed at the phenomenal level – that is, as long as they did not yield any states of pleasure or pain – the mere presence of the disease in the organism makes no difference for the individual's welfare. Therefore, our conception of health relevant to the estimation of welfare ought to be a hedonist one.

The reasoning, according to the hedonist, should generalize: every good must be interpreted in the same way as we have interpreted health. Silvio's life in the world  $w$  is good exactly to the same extent as his life in the world  $w'$ . Two individuals that have exactly the same pleasure-states are assigned the same amount of welfare. Contrapositively, a change in the subject's welfare will imply a change in the subject's pleasure-states. Thus welfare strongly supervenes on pleasure:

*Standard hedonism (S-hedonism).* (1) Pleasure is a class of psychological states.  
(2) Necessarily, individuals indiscernible in terms of their pleasures are necessarily indiscernible in terms of their welfare: For every world  $u$  and

every world  $v$ , if an individual  $a$  in the world  $u$  has exactly the same pleasures as the individual  $b$  in the world  $v$ , then  $a$  has the same amount of welfare as  $b$ .<sup>1</sup>

In offering this formulation of hedonism we have not said anything about the nature of pleasure. In that we are following a venerable tradition. It is notable how coy earlier hedonists were on developing an analysis of pleasure. Thus Mill says only:

The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure. (Mill, 2003, p. 99)

Mill then goes on to expound his theory of the quality of pleasure. Though he gives some clues about how to measure quality, there is still preciously little detail about the nature of pleasure, now characterized by quantity and quality. Paley is more explicit in affirming the psychological nature of pleasure. He similarly claims that ‘happiness’ is the domination of pleasure over pain, but speculates at some length what happiness may be, and how it is distinguished from individual pleasures:<sup>2</sup>

If any positive signification, distinct from what we mean by pleasure, can be affixed to the term “happiness”, I should take it to denote a certain state of the nervous system in that part of the human frame in which we feel joy and grief, passions and affections. Whether this part be the heart, which the turn of most languages would lead us to believe, or the diaphragm, as Buffon, or the upper orifice of the stomach, as Van Helmont thought; or rather be a kind of fine net-work, lining the whole region of the praecordia, as others have imagined; it is possible, not only that each painful sensation may violently shake and disturb the fibres at the time, but that a series of such may at length so derange the very texture of the system, as to produce a perpetual irritation, which will show itself by fretfulness, impatience, and restlessness. It is possible also, on the other hand, that a succession of pleasurable sensations may have such an effect upon this subtle organization, as to cause the fibres to relax, and return into their place and order, and thereby to recover, or, if not lost, to preserve, that harmonious conformation which gives to the mind its sense of complacency and satisfaction. This state may be denominated happiness, and is so far distinguishable from pleasure, that it does not refer to any particular object of enjoyment, or consist, like pleasure, in the gratification of one or more of the senses, but is rather the secondary effect which such objects and gratifications produce upon the nervous system, or the state in which they leave it. These conjectures belong not, however, to our province. (Paley, 2002, p. 13)

---

1 Here we assume the equivalence of modal and possible-worlds formulations. However, see Rønnow-Rasmussen (2011, pp. 10–11) for some doubts about the adequacy of supervenience in formulating theories of value.

2 While Paley was certainly opposed to some variants of hedonism, he was equally certainly a hedonist. Many passages would confirm this view, such as: “In the conduct of life, the great matter is, to know beforehand, what will please us, and what pleasure will hold out. So far as we know this, our choice will be justified by the event” (Paley, 2002, p. 19).

This illuminating passage, hidden in a footnote, suggests a more accurate picture of utilitarian hedonism, and of hedonism in general, than often supposed.<sup>3</sup> Hedonism, according to it, is committed to the supervenience of welfare over a certain class of psychological states. The verdict on whether this class is to consist of sensations, or emotions, or some altogether different states, is deliberately postponed.<sup>4</sup> And it may even turn out that there is no one psychological property shared by all these states, other than them being designated as ‘pleasant’. Earlier hedonists were primarily interested in affirming the phenomenal character of welfare. The precise analysis of this phenomenal character was to be left for future empirical research. The theory of pleasure embedded in S-hedonism should, therefore, leave enough conceptual leeway for such empirical research. It affirms the phenomenal (psychological) nature of ‘pleasure’, but remains silent on its specific characteristics.

Now, opponents of hedonism have long insisted that at least in certain situations facts about the world, rather than the individual’s own experiences, should influence this individual’s welfare. My task below will be to examine three main modifications of hedonism designed to accommodate these problematic cases. The plan is straightforward. In section 2 I begin by reconstructing the main argument against hedonism. I discuss the first two modifications of hedonism in sections 3–5. The third modification is evaluated in sections 6–7.

## 2. The Argument from Illusion

Hedonism got challenged in thought experiments involving various deception scenarios, ironically having the same form as the health scenario above. The reasoning goes as follows. What contributes to your welfare is ultimately the enjoyment you take in the course of different episodes of your life. Certainly you can derive enjoyment, for example, from achieving various goals. But welfare, as we have already noted, should not depend on the actual fact of achieving your goals. It should rather depend on your perception that your various goals are achieved. These perceptions, which shape your attitude of enjoyment or suffering, can be simulated. And your simulated perceptions (or ‘experiences’) will not be different – will not feel different – from your genuine perceptions. Their subjective quality, and therefore the associated pleasures and pains, are supposed to remain constant.

3 See, e.g., Feldman (2004, 2010), Haybron (2008).

4 Thus we should be able to circumvent the issue whether pleasure is a unique sensation. See Sidgwick (1907, p. 127) and also Sidgwick (2000). For a recent discussion see Persson (2006, pp. 30–36) and Brax (2009, pp. 23–32). We similarly postpone the resolution of an issue such as how stable pleasures must be to contribute to welfare, as well as other problems in the links between pleasure and ‘happiness’. For an interesting recent discussion see Feldman (2008).

What the last condition amounts to is best clarified by mundane examples. In one case you have a glass of kefir, in another case you have a glass of liquid yoghurt. The substances are different, but they may feel the same to you, and the pleasure you have in both cases may be the same. Or in one case you look at the Parthenon, in another case you look at its plastic replica. The objects are very different, but they again may feel the same to you, with the same pleasure derived in both cases.

Thus, if S-hedonism were true, then spending your life in an experience machine is as good, or even much better, than living outside the machine. But of course living outside the machine is better. Therefore, hedonism is false. Similarly, you can be deceived in your beliefs about your family life. So, if hedonism were true, to be cheated on by your wife and never discovering her infidelity is at least as good as having a genuinely faithful wife. But it is not as good – it is in fact much worse. Therefore, hedonism is false.

This reasoning by a *reductio* generalizes. We consider two lives similarly endowed with pleasure, but radically differing in terms of a non-hedonic property  $\phi$ . We contend, by an appeal to our intuitions, that the lives rich in  $\phi$  should be better than the lives lacking in  $\phi$ . Consequently,  $\phi$  should be a factor of welfare. Candidates for  $\phi$  included justice and beauty. Here, however, our reference point will be the argument from illusion, by far the most popular argument in the recent debate. The *reductio* takes the following form:

1. If hedonism is true, then welfare supervenes on pleasure.
2. If welfare supervenes on pleasure, then welfare supervenes on certain facts of the subject's psychological life.
3. If welfare supervenes on certain facts of the subject's psychological life, then a life involving a perfect deception or simulation is not better than a life involving no such deception or simulation.
4. However, a life involving no perfect simulation or deception *is* better.
5. Therefore, hedonism is false.

Modifications of hedonism I want to examine here attempt to resist the *reductio* by challenging the key premisses 1 and 2. Before turning to them, let me quickly comment on other premisses. To question premiss 3, the hedonist may argue that 'facts of psychological life' should be interpreted analogously to wide contents. If one individual sees the White House and another individual sees its replica, then they do not really see the same thing. Similarly, if one individual takes pleasure in seeing the White House, while another has pleasure of the same intensity and duration in seeing its replica, then their pleasures are still distinct. Such a move necessitates a revision of the concept of pleasure. As it must involve rejection of premiss 2, it will be covered below.

With regard to 4, the hedonist can raise doubts about the persuasiveness of the proposed thought experiment. He can bite the bullet and deny flat out that the

thought experiments used to justify the premiss are conclusive.<sup>5</sup> Any thought experiment, he might reply, relies on our intuitions in some counterfactual scenario. Some scenarios bear more resemblance to the familiar circumstances of our lives, some bear less. Those bearing less resemblance run the risk of obscuring the relevant details of the situation and thereby delivering a wrong verdict. Consider the case of a deceived spouse. The case is that of a perfect lie, a lie that goes undetected for an indefinitely long time and leaves no marks on the personal relationships of the protagonists. This is fanciful. In reality the lie is likely to be discovered – if not immediately, then some time after. And if it is not discovered, it is still liable to produce changes in the agent's inner world and eventually to influence his behaviour. The thought experimenter is supposed to ignore this possibility and to imagine an adultery having no effect on the behaviour of the deceiving wife.

The appeal of this response is, I think, limited. The thought experiments used in the *reductio* required us to agree to no more than this: our lives being too short, some illusions may remain forever undetected. And certain illusions among these pertain to aspects of life so significant as to detract some value from this life. Granted, thought experiments are conducted in idealized circumstances, but here at least we need not resort to fanciful science fiction and unduly stretch our imagination to recognize the point they purport to establish. Secondly, it would be welcome to pass a verdict on the argument above based on a theoretical principle. If the current response is accepted, however, we are engaging in a fight of intuitions impossible to arbitrate. For a more conclusive defence of hedonism we need to turn to other premisses.

### 3. Hedonism Qualified

Let us look at the premiss 2 of the *reductio*. We have interpreted hedonism as a view that grounds welfare in psychological facts about the agent. The intent is to take the net balance of positive and negative attitudes experienced by the agent. It is also to bypass any possible extra-psychological aspects of these attitudes. If I enjoy eating a frog and you similarly enjoy eating ice-cream, then these experiences contribute exactly the same amount to welfare. Or if I take pleasure in donating money to charity and you take pleasure in torture, then again exactly the same amount is contributed.

Analogously, if I take a certain amount of pleasure in being led to believe that I am  $F$  while being in fact  $\sim F$ , and you take the same amount of pleasure in believing that you are  $G$  while indeed being  $G$ , then our respective pleasures make the same

---

<sup>5</sup> See, e.g., Bronsteen *et al.* (2010).

contributions to our respective welfare. Whether or not a certain state of affairs obtained is immaterial to the level of our welfare. What matters is the belief that it obtained, or failed to obtain, and the relevant psychological states this belief generated.

The most natural suggestion for resisting such an idea is to alter the way we calculate the contribution a particular pleasure makes to overall welfare. We could devise the ‘Qualifying Rule’ as follows.<sup>6</sup> First gauge the intensity of that particular pleasure. Its value  $i$  will reflect an internal psychological state of the agent. Then check whether the pleasure is directed at an actual state of affairs. If it is, then the value  $W_P$  contributed by the given pleasure to the overall welfare is  $i$ . If it is not, then the value  $W_P$  is  $i \cdot n$ , where  $0 < n < 1$  and  $n$  is substantially less than one. In general, let the enjoyment  $P$  be directed at the state of affairs  $S$  and let its intensity be  $I$ . We get the following:

$$W_P = \begin{cases} I & \text{if } S \text{ is actual} \\ I \cdot n & \text{if } S \text{ is non-actual.} \end{cases}$$

On the face of it, the Qualifying Rule does not look terribly plausible. Suppose I am offered a dessert. It looks and feels like a cheesecake. Since I love cheesecakes, I take great pleasure in eating the dish. In reality, however, it is cleverly made of chicken. I cannot stand chicken. Sometime later I am told the truth. I may of course experience revulsion, at that later moment, at the thought of having eaten chicken. Still there should, it seems, be no reason to adjust my earlier pleasure. It makes no sense to say that, given my current information, the pleasure I experienced earlier was in fact lower. Pleasure has a date. My present disgust cannot reach into the past.

One might reply by distinguishing between two cases. In one of them I take pleasure in experiencing the taste of a cheesecake. This taste is perhaps not unique to cheesecakes and in any event it can conceivably be simulated. If, therefore, the dish I am eating is not a cheesecake, no matter: I am not after a cheesecake *per se*, but after a particular type of taste it produces. The state of affairs which I enjoy is: [I am eating  $a$  dish with *this* taste]. Actually not only the nature of the dish becomes irrelevant, but also the very process of eating, which can be simulated as well. A more accurate description of the state of affairs I enjoy would be: ‘I am experiencing *this* taste.’ Then no deception would be possible. I cannot be deceived about enjoying or not enjoying my taste-experiences, though I can be deceived about the proximate causes of these experiences. Thus the pleasure I derive from taste-experiences should not be adjusted.

---

6 The rule, in a slightly different form, is suggested in Feldman (2004, p. 112) where it is used in the defence of another version of hedonism to be addressed below in section 4.

In the second case, by contrast, I take pleasure in eating the cheesecake itself. The state of affairs I enjoy is: [I am eating a cheesecake]. It is essential to my enjoyment that the cheesecake and, say, the process of eating are part of the state of affairs. The nature of the dish now makes all the difference for my enjoyment. So, had I been deceived about the nature of the dish, the pleasure I take in the episode would have been adjusted. And in general, there are many cases where the object of pleasure is distinct from my own activities. Imagine a tennis fan who derives great pleasure from shaking the hand of Roger Federer. Now the fan has no interest in shaking the hand of a Federer lookalike, nor does he relish the very feeling of shaking hands. The fan is excited about the opportunity of meeting Federer and shaking *his* hand. The identity of Roger Federer is crucial for the level of the agent's pleasure. That event, let us say, took place in 2008. In 2009 it turns out that in fact the fan shook the hand of Federer's lookalike. According to the Qualifying Rule, the fan's pleasure should be adjusted. Given the circumstances, it is likely to be close to zero or even negative.

According to this argument, the properties of the past state of pleasure (the pleasure the agent had in 2008) cannot, and should not, be revised. The idea of 'qualified S-hedonism' is to distinguish pleasures, still understood as facts of psychological life, from the contributions these pleasures can make to welfare. We interpret the Qualifying Rule as suggesting only that the episode's contribution to welfare should be revised. The pleasure itself should remain constant throughout.

This view is open to a familiar and rather straightforward objection. It was first articulated in *Principia Ethica*, section 48:<sup>7</sup>

Mill's judgement of preference [i.e., the doctrine of quality of pleasure], so far from establishing the principle that pleasure alone is good, is obviously inconsistent with it. If one pleasure can differ from another in quality, that means, that *a* pleasure is something complex, something composed, in fact, of pleasure *in addition to* that which produces pleasure. Mill, therefore, in admitting that a sensual indulgence can be directly judged to be lower than another pleasure, in which the degree of pleasure involved may be the same, is admitting that other things may be good, or bad, quite independently of the pleasure that accompanies them. (Moore, 1903, pp. 79–80, with omissions)

We can develop Moore's idea as follows. The contribution  $W_P$  of the given pleasure  $P$  is determined, at least in part, by the intrinsic properties of  $P$ . But the only intrinsic property we can find is the psychological intensity of  $P$ . On the other hand, we certainly allow  $P$  to stand in many external relations. Their instances would include '\_\_\_\_\_ is remembered at  $t$ ', '\_\_\_\_\_ is discussed by  $X$ ', and also '\_\_\_\_\_ is directed at a state of affairs  $Y$ '. There are generally many entities in the world standing in an external relation to  $P$ . We could allow the value of these entities also to influence  $W_P$ . However, if we were to say that the relation of  $P$  to  $Y$  influences the

---

7 See also Sidgwick (1907, pp. 94–95).

value of  $W_p$ , then we would have accepted  $Y$  as another *independent* factor in the determination of  $W_p$ . And if we were to accept that, then, so far as  $Y$  is not a psychological state, the supervenience of welfare over psychological attitudes would have been violated.

Call this objection ‘the heterogeneity objection’.<sup>8</sup> An altogether different objection begins with the assumption, possibly also endorsed by Moore, that only intrinsic properties of  $X$  constitute its value (call it ‘the externality objection’).<sup>9</sup> Since the only intrinsic property of pleasure is its intensity, the value of pleasure must be determined by that intensity. But, on the other hand, the value of life is determined by the values of pleasures the subject had. So it should not be determined by the values of all sorts of items externally related to these pleasures.

Thus stated, the externality objection misinterprets the formula of S-hedonism and the corresponding adjustment made by the Qualifying Rule. The value of life (the person’s welfare), according to S-hedonism, is determined by the quantity of pleasure the subject has. So far as pleasure is a psychological state, its quantity is wholly fixed by its intensity. What value in general should be ascribed to that quantity is a question not settled by S-hedonism. In effect it means that the quantity  $W_p$  contributed by the given pleasure to the overall welfare might not be identical to the value of pleasure. So we should be able to apply the Qualifying Rule to  $W_p$ , but not to the value of pleasure. The qualified S-hedonist can simultaneously endorse the following formulae (where  $V(x)$  stands for ‘value of  $x$ ’):

$$V(\text{Life}) = \sum_i W_{P_i} = I(P_1)n_1 + I(P_2)n_2 + \dots + I(P_m)n_m$$

$$V(P_j) = I(P_j).$$

The externality objection is thereby dismissed. The value of pleasure is determined by its intrinsic property of intensity. But the value of life is *also* determined by its intrinsic properties. A life spent in fighting wars is intrinsically distinct from a life spent in an experience machine simulating wars. And a life spent in writing books is intrinsically distinct from a life spent in an experience machine simulating book writing. An activity and its simulation are two intrinsically distinct activities. Thus by the lights of the Qualifying Rule value is still determined by intrinsic properties.

The heterogeneity objection, by contrast, does not hinge on the assumption that only intrinsic properties of  $X$  constitute its value. The objection evidently allows, if only for the sake of argument, the possibility of value contributions of extrinsic properties. But had such properties – e.g., various external relations – influenced the value of  $X$ , that value would have therefore depended upon the value of the

8 Not to be confused with Sidgwick’s ‘heterogeneity problem’. See n. 3.

9 Thanks to the anonymous referee for insisting on greater clarity on this issue.

relata. Thus, to borrow Shelly Kagan's example, the pen used in signing the Emancipation Proclamation would partially derive its value from the value of the Proclamation.<sup>10</sup> The same holds for qualified S-hedonism. Whether or not the property '\_\_\_\_\_ is spent in writing books' is intrinsic to life is a moot question. But to allow this property to influence the prudential value of life is to recognize a factor of that value other than pleasure. The Qualifying Rule and the formula for  $V(\text{Life})$  above reflect this commitment.

The culprit is the theory of pleasure. Qualified S-hedonism remains orthodox in treating pleasure as a certain *psychological* state. Ontically pleasure remains the same as in classical unqualified hedonism. Granted, we do not insist on construing pleasure as a special kind of sensation. As mentioned before in section 1, further research can tell us what psychological phenomenon pleasure is (whether, e.g., it is an emotion, a propositional attitude, or a sensation). But if its psychological nature is thus reaffirmed, how could its *contribution* to welfare – that is, its value – depend on a manifestly non-psychological factor? Welfare is now functionally determined by the subject's psychological attitudes *and* by certain states of the world that may, or may not, be causally connected to these attitudes.

#### 4. Attitudinal Hedonism

Hedonism, in its most general form, is the view that welfare supervenes on pleasure. As our formulation of S-hedonism made explicit, it combines two theories: the theory of welfare and the theory of pleasure. The theory of pleasure tells us that pleasure is nothing but a psychological state. The theory of welfare tells us that welfare supervenes on pleasure. It is expressed in the first premiss of our *reductio*. Qualified S-hedonism, in challenging the premiss 2, was in effect led to deny the premiss 1. That was fatal for its cause, since evidently no hedonist worth his salt, qualified or non-qualified, can abandon that theory of welfare.

But a hedonist can nevertheless try to modify the theory of pleasure. This is the avenue explored by Fred Feldman. Pleasure is an attitude we take towards a state of affairs. S-hedonists conceive the attitude/state of affairs relation in causal terms. Attitudes are caused by states of affairs, and it is possible for two distinct states of affairs to cause the same attitude. That is why we also believe that attitudes and states of affairs stand in an external relation. But what if we were to insist on the intrinsic relation between them? Feldman writes:

[F]ollowing Mill, we can say that when a person takes intrinsic attitudinal pleasure in some moral, intellectual, or aesthetic matter, his episode of pleasure as a whole is "higher", and thus more valuable in itself. If we allow ourselves a certain amount of conceptual leeway, we can construct the

---

10 See Kagan (1998).

theory in such a way that the objects of attitudinal pleasure are intrinsic elements in the episodes of pleasure. (Feldman, 2004, p. 73)

The states of affairs become themselves part of your pleasure. The identity of your particular pleasure depends not only on the intentional mental states it is associated with, but also on what causes these mental states. The objects of attitudes should then stand in an external (causal) relation to mental states, but ‘pleasures’, as currently interpreted, should have both of them as intrinsic parts.

Here is a sketch of Feldman’s proposal. To begin with, pleasure is not any singular kind of sensation. This is a familiar observation often used against Bentham’s utilitarianism (see n. 2). We are able to enjoy things so diverse that it is not reasonable to believe that there is a common mental denominator in our enjoyments. It is better to replace the talk about ‘feeling pleasure’ or ‘feeling pleased’ with the terminology of ‘taking pleasure in  $\xi$ ’. We take pleasure in eating cheesecakes and in listening to Mozart. And we take pleasure in the fact that our bank account is in such excellent shape. So we take pleasure in states of affairs, or equivalently, in propositions. Thus the proposal is to think of pleasure as a propositional attitude.<sup>11</sup>

As a second step, we make the object of the pleasure-attitude the integral part of this attitude. That this is not an unreasonable move we can see by comparing pleasure to another attitude, such as desire. I desire to look at the Parthenon. You desire to look at a carefully crafted copy of the Parthenon. The satisfaction conditions of our desires are different, even though the copy may be perceptually indistinguishable from the original. If I am looking at the copy and you are looking at the original, neither of us has satisfied his desire. The same goes for pleasure.

The proposal, then, is analogous to the well researched idea of wide and narrow contents, already mentioned in section 2. ‘Jon experienced Federer’s handshake’: this statement cannot be true, unless Jon shook the hand of Roger Federer. If Jon never met Federer, but only bumped into his look-alike, he could not possibly have experienced shaking Federer’s hand. In one sense, experience necessarily depends on states of affairs outside the subject’s mental sphere. In another sense, experiences denote mental states characterized purely psychologically, from the inside. Jon has never met Federer. Yet, we wish to say, if he were to meet him, his feelings, his emotions, his sensations, would have been exactly the same – as when he shook the hand of Federer’s look-alike. That is the narrow content of his experience. The term ‘pleasure’ can also refer either to a certain kind of a mental state, or to a propositional attitude. Only the latter should feature in the formulation of hedonism.

---

11 There are legitimate worries whether *all* instances of enjoyment reports should signal a predicative mode of expression and whether, therefore, the object of enjoyment is *always* to be assimilated to a proposition. See Anscombe (1981). Dialectically, the most interesting consequence of these worries is, I think, the fact that the concept of attitude, so crucial to A-hedonism (and to other views, such as Sumner’s authentic happiness), is underdeveloped. But we should be willing, for the sake of argument, to grant the A-hedonist his propositional concept of attitude.

The outcome is a different form of hedonism. We maintain the supervenience of welfare over pleasure, but we modify the theory of pleasure so as to reject the premiss 2:

*Attitudinal hedonism (A-hedonism).* (1) Pleasure is a complex object combining a psychological state (an attitude) with a certain state of affairs (the object of the attitude). (2) Necessarily, individuals indiscernible in terms of their pleasures are necessarily indiscernible in terms of their welfare: For every world  $u$  and every world  $v$ , if an individual  $a$  in the world  $u$  has exactly the same pleasures as the individual  $b$  in the world  $v$ , then  $a$  has the same amount of welfare as  $b$ .

There are two objections I wish to raise against this view. They will be developed in the remainder of this section.

### *Incredulity*

An immediate difficulty facing A-hedonism is the ‘incredulous stare’ objection.<sup>12</sup> A-hedonism offers a drastic revision of the concept of pleasure. Our pre-theoretic concept of pleasure has phenomenal character. We may agree, indeed, that pleasure is not one unique sensation. And we may agree, readily, that some experiences we designate as pleasant should not be classified as sensations at all. But at a minimum we also agree that pleasure is a fact of our psychological life. It of course stands in many causal relations with non-psychological entities. But very different entities could cause essentially the same pleasure, whether we construe it as sensation, or as psychological ‘attitude’, or as psychological ‘episode’. Making non-psychological entities an intrinsic element of pleasure represents, for some of us, a departure from the pre-theoretic concept too big to be credible.

The incredulous stare objection is not conclusive and is not actually intended to be so. It is an attempt to portray the view criticized as especially esoteric – esoteric, indeed, to such an extent as to make it not really interesting. But we can try to make it a little more constructive and specific. Suppose we were to accept A-hedonism. Then we are still able to discern attitudinal objects (that is, states of affairs) in pleasure. The idea is not to treat pleasure as a primitive entity. In addition, we must be able to discern in pleasure a certain psychological state. Thus suppose we were now to think of pleasure as a complex  $\langle S, O \rangle$  constituted by the psychological state  $S$  and the attitudinal object  $O$  (other elements may appear too, but we can ignore them now). And we insist that  $O$  stands in an intrinsic relation to the complex  $P$ , though perhaps not to the state  $S$ . Well, what should prevent us from identifying pleasure with  $S$ , rather than with  $P$ ? In other words, it is possible that we managed

---

<sup>12</sup> See Lewis (1986) for this terminology.

to define and describe a certain complex entity. There is, however, no evidence that pleasure is to be identified with it. It is as though I were to take my desk and my phone sitting on it and stipulate an entity ⟨desk, phone⟩. Nothing prevents me from further stipulating the term ‘desk’ to refer to the complex of which phone is an intrinsic element. Its first element I now propose to call ‘table’. But I have not shown thereby that there is anything wrong with stipulating otherwise and treating the first element of the pair as the *real* desk.

Another question one should be asking is this: what is the relation between the subject and his pleasure? It can be either intrinsic or non-intrinsic. A-hedonism allows it to be non-intrinsic. I may take pleasure in the state of affairs [Federer wins Wimbledon 2009], to which I stand in a non-intrinsic relation. Since this state of affairs is an intrinsic part of my pleasure, the relation between me and this pleasure of mine, we conclude, is also non-intrinsic. And of course it should seem strange to say that the subject can be in a non-intrinsic relation with his pleasure. The A-hedonist would reply that this is no stranger than saying that the subject stands in a non-intrinsic relation with his thoughts. But there are many theorists who believe precisely this claim. The quick answer here is that the two cases are not that similar to have parallel justifications. Thought is supposed to represent the way things are. It can then be argued that representations constitutively depend on what they actually represent. Pleasures, however, are not supposed to represent anything.<sup>13</sup> Saying that they do marks a departure from our shared conception of pleasure; such a departure is absent in the case of thought.

The incredulous stare objection requires the A-hedonist to give us more evidence as to why his stipulations manage to capture the way things are – why, that is, his ‘pleasure’ has anything to do with what *we* mean by ‘pleasure’. Perhaps, however, it is not terribly clear what we mean by ‘pleasure’. And furthermore, while acknowledging the departure from the common meaning, the A-hedonist is likely to emphasize the pragmatic virtues of his theory. The primary such virtue is in the ostensible success in tackling the argument from illusion. A back and forth would follow, questioning the theoretical status of pragmatic virtues. The outcome of this exchange is uncertain. So we should think of a more constructive objection.

#### *Heterogeneity objection sustained*

Suppose we were to accept the A-hedonist theory of pleasure and think of pleasure as a complex  $P = \langle S, O \rangle$ . We wish to maintain, in contrast to qualified S-hedonism, that the level of welfare is determined entirely by the level of pleasure. But what is the level of pleasure, the quantity  $W_P$ , which it contributes to overall welfare? One

---

13 But see Schroeder (2004, pp. 89–97) where pleasure is interpreted as representing the individual’s desires. For a rejection of this view see Persson (2006, pp. 58–60). See also Katz (2009, section 3.2) and Brax (2009, pp. 52–56).

answer here is that it should be identified with the intrinsic value of pleasure (i.e., with the quantity associated with that value). And the intrinsic value of  $P$  will be determined by  $S$  and  $O$ . It is not a helpful answer. For we would only have this formula for intrinsic value if we accepted its supervenience on intrinsic properties. However, as noted earlier in section 3, the supervenience assumption might not be granted. The quantity  $W_p$  and the intrinsic value of  $P$  may then come apart.

Instead we ought to seek a different starting point and bypass any qualms about the theory of intrinsic value. The intent of A-hedonism is to provide a response to the argument from illusion. Thus we should assume that the quantity  $W_p$  is influenced by the intensity of the mental state  $S$  and by another quantity associated with  $O$ . And how are we supposed to set the value of  $O$ ? Our criteria in doing so should be external to hedonism. The Qualifying Rule thus stipulates that actual states of affairs will be more valuable than non-actual ones.<sup>14</sup> Other rules may be that noble states of affairs or more deserving ones are more valuable than less noble or less deserving ones. But if actuality or nobility were to carry more value, we must be able to say why they do so. We can be sure, however, that they carry more value not because they produce more pleasure. On the contrary: the amount of pleasure they produce is augmented by their own intrinsic properties (i.e., their own nature). There is no escape from affirming the existence of different sources of value, each of them not reducible to the other.

Notice also that the same conclusion follows if A-hedonism resorts to the idea of organic wholes.<sup>15</sup> On the holistic interpretation of this idea, the value of  $P$  is affected by the value of a whole, in addition to the values of  $S$  and  $O$ . On the so-called 'conditional' interpretation, the value of a whole is fully determined by the values of  $S$  and  $O$ , but those values vary within the whole and in isolation. Whatever interpretation we adopt, the main idea of the present objection remains intact. The way the entity  $O$  contributes to  $W_p$  is determined, at least in part, by that entity's own intrinsic properties. As long as these properties are counted as factors of welfare, we are committed to recognizing an additional source of value.

The heterogeneity objection can, therefore, successfully be cloned if we were to make attitudinal objects intrinsic elements of pleasure. A-hedonism fares no better than S-hedonism in this regard. The significance of Moore's objection is in aligning the putative 'hedonism' with anti-hedonism: the two become indistinguishable. While S-hedonists locate intrinsic good in certain psychological attitudes (alternatively: in a property of such attitudes, their enjoyableness), A-hedonists locate intrinsic good in those attitudes and in their causes (i.e., different states of the world). Anti-hedonists may also be prepared to assign intrinsic value to psycho-

---

14 'More valuable' here is evidently not to mean 'more intrinsically valuable.' It is rather a shorthand for 'contributes more quantity to the level of welfare.'

15 See Hurka (1998) and Bradley (2002).

logical states and to many other factors, all of which cause, or could cause, enjoyment.<sup>16</sup> In the next section I will illustrate this corollary of the heterogeneity objection by citing one instance of such anti-hedonist view.

### 5. The Road to Anti-hedonism

*Nicomachean Ethics* is commonly known for its forceful rejection of hedonism.<sup>17</sup> In X.2 Aristotle castigates hedonists for a faulty argument that pleasure is the only good. What we can, and should, say, according to Aristotle, is that pleasure is a good, i.e., only one of several goods. Even that, strictly speaking, would be a dubious claim, since in X.4 a theory is developed in which pleasure is said to be a special ‘completing’ mode of activities, rather than a goal on a par with virtue or health.

Nevertheless Aristotle should be viewed as an unlikely father (or better, a godfather) of A-hedonism. In two remarkable paragraphs at the end of X.5 (1175b24–1176a29) he resolves to show that ‘classical’ hedonism – that is, our S-hedonism – is fundamentally a bankrupt view. In the first place, we ought to recognize some pleasures as good, others as bad. This is supposed to follow from the view he defended earlier in the book, in which pleasures accompany, or ‘complete’, activities. If an activity itself is good, whatever completes it is also good. Correspondingly, if an activity is bad, then the pleasure completing it is bad. Hedonism is accused of not distinguishing between these ethically different kinds of pleasure.

Such an objection cannot of course be made from within S-hedonism (as we witnessed in section 3). Once we insist on the idea that phenomenal pleasure alone is good, then there are no grounds left for maintaining that some pleasant activities could still be bad. Aristotle should be understood as challenging the premisses of S-hedonism. We should not think of phenomenal pleasures as the only good, *since* we have an independent reason to judge an activity as good or bad, and therefore, to judge the pleasure associated with it as good or bad.

So Aristotle anticipated something very much along the lines of the Qualifying Rule. If we believe that pleasures can make any contribution to welfare, we must first adjust them. As before, we begin by gauging the intensity  $i$  of a particular pleasure. Then we check if the activity completed by pleasure is good or bad. Let  $j$  be the indicator of goodness and badness. The value contributed by the given pleasure to the overall welfare is  $i \times j$ . If the activity is good, we set  $j > 1$ . If the

---

16 Anti-hedonism is thus distinct from what Bentham called ‘asceticism’ and Feldman now calls ‘dolorism’. The latter is the view that pain, rather than pleasure, is intrinsically good. See Feldman (2004, p. 183).

17 For a dissenting view see Prichard (1935) and Berkovski (2011).

activity is bad, we set  $j < 0$ . And perhaps if the activity is trifling, but not harmful, we can set  $0 \leq j \leq 1$ . In general, let the pleasure  $P$  complete the activity  $A$ , and let the intensity of  $P$  be  $I$  and the goodness of  $A$  be  $J$ . Then the contribution  $W_p$  made by  $P$  to welfare is given by the formula:

$$W_p = I \times J, \text{ where } \begin{cases} J > 1 & \text{if } A \text{ is good} \\ 0 \leq J \leq 1 & \text{if } A \text{ is neutral} \\ J < 0 & \text{if } A \text{ is bad.} \end{cases}$$

According to Aristotle's rule, we recognize – intuitively or pre-theoretically – that some activities, such as torture or sexual perversion, are bad regardless of the amount of pleasure they produce. In fact they seem to be made worse when accompanied by intense pleasure on the part of the torturer or pervert. It is bad enough if a torturer causes pain to his victim. It is much worse if the torturer, while causing the same amount of pain, also experiences intense pleasure. We refuse to say that his pleasure is good and that it can make his life better. And we believe that the reason it is bad is precisely in the nature of the action that generates pleasure (or as Aristotle might put it, 'is completed' by pleasure).

At 1176a10 Aristotle is prepared to go even further and challenge the view of pleasure as a psychological state. We normally believe that it is possible to take pleasure in just any state of affairs (however the value of  $W_p$  is to be calculated). Facts about your hedonic attitudes may tell us a lot about your character, the attitudes may be considered noble or repugnant; yet it is impossible to deny that you *actually* enjoyed that state of affairs. Or to put it in Aristotle's parlance, the activity can be as bad as any, yet it is eminently possible for the pleasure to complete it (i.e., for the agent to take pleasure in it). But now Aristotle argues that some activities are such that pleasure cannot correctly be said to complete them. One *could* feel pleasure while engaged in  $\varphi$ -ing, or generally take pleasure in  $\varphi$ -ing. But the activity *itself* is not pleasant, unless the person is in a right state.<sup>18</sup>

The analogy Aristotle intends to draw is with qualities such as coldness and sweetness. An object could be felt as cold or sweet by a person. Yet, when this person is feverish or injured, we are not entitled to conclude that the object is itself sweet or cold. We are justified in attributing the properties of coldness and sweetness to the object itself only when a healthy or otherwise 'normal' person judges it to be so.<sup>19</sup> With these qualities, then, there is a standard of correctness such that, if a person is sick, the room may appear to him colder than it *actually* is. His judgement about the room's temperature would reveal to us not the information

18 See Gosling and Taylor (1982). For an alternative interpretation of this passage see Broadie (1991).

19 Aristotle's account of secondary qualities is, from a modern perspective, largely indefensible. But it is still heuristically useful for understanding the ethical argument.

about the state of the room, but rather the information about his own state. In such cases we should say that the room is extrinsically cold (cold-for- $X$ ), but intrinsically warm. Analogously, a deprived activity would be extrinsically pleasant, but intrinsically painful (1176a23). And again, the channel enabling us to learn the intrinsic hedonic value of an activity is the judgement of a good man (1176a16).

The revolutionary proposal consists in a new account of pleasure. Whether a given activity is pleasant is fixed by an extra-psychological criterion. Your actual enjoyment of a certain activity  $\phi$  is a reliable indicator of the actual hedonic value of  $\phi$  only if your character satisfies an ethical condition. The distinction obtained is between pleasures as proper completions of activities and pleasures as bare psychological states. Only the former count as *genuine* pleasures and as such may contribute to individual welfare. Aristotle is now able to maintain both (a) that pleasure contributes to individual welfare, and (b) that hedonism misrepresented the nature of pleasure, and therefore, provided an essentially inaccurate theory of welfare. The conclusion Aristotle wishes to draw amounts to what may be termed ‘qualified anti-hedonism’. Pleasure is indeed a factor in welfare, but only once it has been properly understood to account for the objects of pleasure (here: ‘completed’ activities). And this is precisely the conclusion an A-hedonist also draws. The Qualifying Rule adjusts the value of pleasure depending on the actuality of its object. But a similar strategy can be deployed if one wanted to adjust pleasure in relation to justice (so that manifestly unjust pleasures are not valuable) or morality (so that manifestly immoral pleasures are not valuable). Feldman pursues exactly such a strategy and uses it to show the welcome ‘plasticity’ of hedonism.<sup>20</sup>

Now if the qualifying rules can serve hedonists and anti-hedonists alike, it becomes exceedingly difficult to see just what the issue between them *is*. For a remarkable consensus has been achieved. Both parties wish to say that only a right kind of pleasure is valuable, or ‘good’, to any significant extent. And both agree, in subscribing to the adjustment rules, on how to separate the right and wrong kinds of pleasures, i.e., the pleasures that should and should not be adjusted. What evidently should remain in dispute is the exclusivity of pleasure. Consider the following proposal:<sup>21</sup>

*Criterion of hedonism.* A theory  $T$  is a form of hedonism iff according to  $T$ , welfare is entirely determined by the levels of pleasure and pain.

A-hedonist seems to satisfy the criterion. The A-hedonist, once all the adjustments have been made, insists that pleasure is the sole good. Anti-hedonism should

20 See the discussion of DAIAH and AAIAH in Feldman (2004).

21 See Feldman (2004, pp. 172–177) and the criterion H4 there. I am also inclined to believe that Feldman’s critique of H3 is inconsistent with his own attitudinal hedonism, but the details cannot be elaborated here.

fail the criterion. And indeed, a qualified anti-hedonist like Aristotle, even after all the adjustments have been made, insists that pleasure does not exhaust the good. Virtue is good, for instance, and so is knowledge.

But the difference between A-hedonism and qualified anti-hedonism is illusory: their disagreement is merely verbal. For simplicity, let us restrict our attention to the role of virtue. In a version of A-hedonism, such as Feldman's DAIAH (with desert interpreted broadly to encompass morality), the subject's pleasure  $P$  will increase depending on whether he enjoys a virtuous or a vicious act. Thus virtue will influence the contribution  $P$  makes to the overall welfare. And thus virtue will be a factor of welfare, according to A-hedonism no less than according to Aristotle. Conversely, the pleasure of a vicious act will be either minimal or negative – again, exactly as Aristotle would demand. Consider now a hard case, a life of a virtuous person who experiences various tragedies and physical pain in his life. Surely the A-hedonist, as any other hedonist, and Aristotle would diverge in their verdicts on his welfare? Not necessarily, and their arguments mirror each other. Such a person, according to both, should have a positive attitude towards virtuous acts. So the virtuous acts such a person performs would contribute positive value to his welfare. It may though be outweighed by the negative value of his many misfortunes. And so the overall welfare will be low.

The agreement in all these cases is not merely extensional. Explanatory patterns also match. A possible area of disagreement between the two accounts is the prudential value of psychological attitudes. Classical (standard) hedonism contained the following insight: welfare depends entirely on the psychological state of the agent. How things are in the world should not matter for personal welfare. What matters is their impact on the psychological condition of the agent. Qualified anti-hedonism, of the sort sketched by Aristotle, is purported to reject this idea. Virtue is a central determining factor of welfare. Pleasure necessarily accompanies a virtuous activity. Since a virtuous agent cannot fail to enjoy a virtuous activity, pleasure becomes part of that very activity. So it should seem that a psychological attitude, in isolation, has no value to contribute to welfare, nor does it have intrinsic value.<sup>22</sup>

The stance of A-hedonism is a little difficult to divine, since it hinges on the precise interpretation of the complex  $\langle S, O \rangle$  (see section 4). If the complex is either a mere aggregate or an organic whole interpreted holistically, then a psychological state  $S$  has value in isolation. Welfare would functionally depend both on the intensity of  $S$  and on a host of other factors. Such factors will be named in the Qualifying Rule. They may include the adequacy of the agent's factual beliefs, or the adequacy of his valuational system, or even the plain morality of

---

22 Textual evidence for the reasoning along these lines is in *NE*, 1177a12, 1104b4–13, 1175b27.

his actions. If, however,  $\langle S, O \rangle$  is an organic whole interpreted conditionally, then  $S$  would not have value in isolation, exactly as demanded by Aristotelian anti-hedonism.

The practical implications of this ambiguity can be illustrated as follows. A person may have an intense feeling – of what *he* calls ‘enjoyment’ – while torturing cats. The fact of that feeling may even be corroborated by a scientific procedure, such as a brain scan. From a first-person and third-person perspective that feeling may be judged similar to the feeling a generous person has while doing charity. Now, if  $\langle S, O \rangle$  were to be a conditional organic whole, no amount of enjoyable torture would be sufficient to augment the torturer’s welfare, at least to an extent comparable with the welfare of a generous person. But if  $\langle S, O \rangle$  is an aggregate or a holistic organic whole, then, given sufficiently frequent torture, the torturer would eventually obtain a higher level of welfare than a virtuous individual.

It is unclear to me how an A-hedonist will decide between these competing interpretations. Nevertheless the least we can conclude is that A-hedonism is a version of qualified anti-hedonism.

## 6. Conditional Hedonism

If the objections raised above are well-taken, a standard hedonist wishing to accommodate the argument from illusion could attempt a different strategy. Instead of tinkering with the theory of pleasure, he may attempt to qualify S-hedonism by imposing certain *de dicto* constraints. The idea is to say that S-hedonism holds only if these constraints are satisfied. One can say, in alluding to the *reductio*, that the supervenience relation featuring in S-hedonism should be premised on some further properties of the subject. The logical form of this revision is very simple. Let  $C$  be a conjunction of (non-trivial) claims about the subject. Then we revise S-hedonism as follows:

*Conditional hedonism (C-hedonism)*. (1) Pleasure is a class of psychological states. (2) For every world  $u$  and every world  $v$ , if an individual  $a$  in the world  $u$  has exactly the same pleasures as the individual  $b$  in the world  $v$ , then  $a$  has the same amount of welfare as  $b$ , but only if  $C$  is satisfied.

With usual conventions, formally we could express conditional hedonism thus:

$$\forall u \forall v (Iua \wedge Ivb \supset [(Fa \leftrightarrow Fb) \supset (Ga \leftrightarrow Gb)] \supset Cab).$$

The welfare of a subject, therefore, is determined entirely by the distribution of the attitudes of pleasure and pain, but only if a constraint  $C$  is satisfied. C-hedonism is not a single theory, but a family of theories. We can select a particular theory by selecting a particular predicate  $C$ . It is generally clear by what criteria  $C$  will be

selected. The condition will be drawn from the qualifying rules, such as Feldman's rules or Aristotle's rule. So we could say that welfare supervenes on pleasure only if the subject takes pleasure in morally impeccable activities. Or we could say that supervenience holds only if the subject is sufficiently informed about the way his more important activities are conducted. The former route will be taken by the theorists who, like Aristotle, are concerned with denying high amounts of welfare to evil agents. The latter option will serve as a response to the argument from illusion.

As an account of welfare, C-hedonism is incomplete. It does not tell us how much welfare is attributed to the subject who fails the condition *C*. For if the subject fails the condition, then supervenience fails, too. If the subject is, like Nagel's businessman, deceived about his marital life, or if he is, like Truman Burbank, deceived by virtually everyone around him about many important aspects of his life, or if he is, like Don Quixote, involved in some intricate process of self-deception – then, we say, we cannot compute his welfare by attending exclusively to his pleasures and pains. Well, then how *are* we supposed to compute it? We have a choice. In estimating welfare in problematic cases we can either ignore the psychological attitudes of the subject (pleasures and pains), or we can somehow adjust them. Neither alternative is appealing. The first one is simply not viable. There can be no sensible view which in some cases pins welfare entirely on attitudes, while in other cases utterly refuses to attach any weight to them. What is left is the second alternative. But it has already been exhausted in the kinds of qualified hedonism explored earlier.

But perhaps the intent is not to provide us with a computation procedure analogous to Feldman's Qualifying Rule. It may be that in unproblematic cases we are supposed to use the usual S-hedonistic calculation, while in problematic cases we cannot say anything definitive about the level of welfare. We should admit welfare-value gaps. Further enquiry may be needed to justify these gaps. This, however, is not the real problem here. What we have achieved in conditionalizing welfare is the creation of *another* supervenience relation. Supervenience in general expresses determination. By fixing the distribution of base properties we are able to fix the distribution of supervenient properties. Once the original supervenience relation is supplied with a condition of its application, welfare would be determined not only by the distribution of hedonic states, but also by the properties contained in the condition *C*. Two subjects indiscernible in their psychological states and in their properties named by *C* are also indiscernible in the level of their welfare. While the conclusion is thus identical in all three cases, the argumentative strategy of C-hedonism is different from the other two approaches. It does not seek to reject premiss 2 of the *reductio* by tweaking the concept of pleasure. The psychological nature of pleasure is reaffirmed, so that premiss 2 is left intact. It is premiss 1 which is directly rejected in its unconditional form.

## 7. Theoretical Unity

In creating an alternative supervenience relation, conditional hedonism, just like qualified S-hedonism and A-hedonism, falls victim to the heterogeneity objection. It represents a significant departure from any recognizable form of hedonism: pleasure is relegated to merely one among other factors of welfare. Whether or not C-hedonism deserves to be classified as ‘hedonism’, a question should still be asked about how viable it is. Perhaps, the thought goes, it constitutes the best response to be offered by hedonism to the argument from illusion. And the achievement is in defending the prominent, if not unique, role of phenomenal pleasure in the theory of welfare.

If this is how we interpret its intent, then the chief argument in its favour is pragmatic. While C-hedonism may be a weak form of hedonism, it is the least problematic interpretation of it available. But I am now going to argue that there is also a good pragmatic argument against C-hedonism.

Now, since C-hedonism is a family of theories, its particular versions will differ with regard to the condition *C* they name. Though it may be possible to evaluate C-hedonism in general, for any arbitrary non-trivial *C*, this is not my task here. Far more urgent is to examine an instance of C-hedonism that has recently gained much prominence in the literature. I am referring to the account defended by Wayne Sumner.<sup>23</sup> The proposition *C*, according to Sumner, is a conjunction of two constraints. They are intended to reflect two distinct worries. One is our problem of factual deception dramatically emphasized in the argument from illusion: the subject must be sufficiently well informed about the conditions of his life. Clearly not every kind of misinformation should count. There are trivialities that can be omitted. Equally no one can be perfectly well informed about his life. Misinformation that counts must be such that it affects the subject’s judgement of welfare. In other words, a subject is sufficiently well informed when he is not deceived about any of his ‘significant’ enjoyments. If he knew that his enjoyments are false, he would have altered his judgement about his own welfare. Sumner does not offer a precise way of putting these ideas together, but my gloss would be as follows:<sup>24</sup>

---

23 See Sumner (1996, 1995). An earlier appearance of a very similar idea is in Sidgwick (1907, pp. 111–112). For a recent criticism see Enoch (2005).

24 Interpretation of Sumner’s view is greatly complicated by his notion of happiness. Welfare is happiness constrained by a certain condition *C*. Enjoyment and happiness “nearly” converge (Sumner, 1996, p. 148), but at the end happiness is linked to one’s assessment of one’s life as a whole (pp. 149–156). So to be happy to a degree *Q* is to reflect on your life as a whole and to be satisfied with it to the degree *Q*. However, the grounds for such assessment are presumably still the enjoyments and sufferings experienced by the judging subject. Sumner, then, adopts a subjectivist theory of welfare linked directly to the

*Information Constraint (IC).* Let  $S$  be the statement ‘My life is going well.’ Let  $P_1, P_2, \dots, P_n$  be statements such that:

1.  $x$  believes  $[P_1] \wedge x$  believes  $[P_2] \wedge \dots \wedge x$  believes  $[P_n]$ .
2. (Had  $x$  believed  $[\sim P_1]$ , then  $x$  would have believed  $[\sim S]$ )  $\wedge$  (had  $x$  believed  $[\sim P_2]$ , then  $x$  would have believed  $[\sim S]$ )  $\wedge \dots \wedge$  (had  $x$  believed  $[\sim P_n]$ , then  $x$  would have believed  $[\sim S]$ ).

Then  $x$  is sufficiently well informed about his life iff  $\forall i P_i$ .

Factual illusion is not the only kind of illusion. The subject may also be under what may be called ‘valuational illusion’.<sup>25</sup> An enslaved worker, an oppressed wife, or an indoctrinated foot soldier may often believe they have a welfare-rich life. These agents may well enjoy the circumstances of their lives. For a hedonist, the level of their welfare may be very high. But since their social conditions are hopelessly degraded, we ought to conclude that their lives are not going well at all. Welfare no longer supervenes on pleasure.

Sumner’s solution is based on the notion of autonomy. We begin by asking how the agent came in possession of the values he actually has. We must then identify a historical process that led to the agent forming the values that he has. We will be looking for external factors in the agent’s value formation. Sometimes the process of indoctrination will be evident enough. The agent will be brainwashed, so that he himself does not consciously select the values in question. In many other cases, the external forces play only a partial role in the agent’s choice. In fact, as Sumner admits, we are all to some noticeable extent part of our environment. Our choices are always, to some extent, conditioned externally. The division between autonomous and non-autonomous agents depends not on the mere presence of external factors, but on the agent’s capacity to reflect on his values (and on the way he acquired them). By being unable, or unwilling, to engage in a critical examination of their values the agents jeopardize their autonomy. The second constraint takes the form:

*Autonomy Constraint (AC).*  $x$  is autonomous iff  $x$ ’s values are selected in the process of critical self-examination.

---

opinion the subject has about his life, where this opinion is shaped by hedonistic attitudes. Yet the view adopted in the present article (see section 1) is that ‘subjectivism’, for what it is worth, should be expressed by ontic supervenience. So the net result is still conditional hedonism. At all events, Sumner’s candidate for  $C$  can be discussed independently of his other commitments.

25 See Sen (1985).

According to Sumner, therefore, *C* is the proposition that the subject is autonomous and well-informed about his life.

The flaw common to (IC) and (AC) is their explanatory deficiency. (IC) is a response to deception scenarios. But there is no explanation why we should take these scenarios at their face value. (IC) gives us no ammunition against those who simply dismiss their apparent lessons. Once we have decided that there must be some limit on how far the agent can be deceived, then (IC) may give us a clue as to how to draw the line between relevant and irrelevant deceptions. But *before* we have some general reason to believe in such a limit, the employment of (IC) would be *ad hoc*. (IC), that is, would be a modification of hedonism not dictated by theoretical reasons. It would be a fix, theoretically arbitrary, to obtain a verdict matching our intuitions.

A parallel concern exists with regard to (AC) which is a response to Sen's scenarios. The requirement of self-examination points towards a substantive theory of welfare which would ultimately condition it on the possession of wisdom and virtue.<sup>26</sup> In this case the hedonistic ingredient, if it were to remain independent, should look increasingly artificial. In fact it is clear what direction the argument will take. It will drop the hedonistic factor altogether, and then face (bravely) the charge of going against our central intuitive convictions. Or it will, equally implausibly, go in the direction of modifying the theory of pleasure by making pleasure a necessary companion of wisdom. Unless one of these moves is accomplished, (AC), like (IC), will be a merely *ad hoc* response.

The worry can be put by saying that C-hedonism fails to meet the following demand:

*Integration demand.* If the amount of welfare depends on several diverse factors, there must be an explanation why all of them should contribute to welfare. Such an explanation will provide a level of generality allowing us to discern non-arbitrary connections between those factors.

The significance of this demand is not in searching for an answer why exactly *these* factors influence welfare. Such a quest is futile, if it is meant to be a quest for an impossible ultimate explanation and if we do not hope to base our theory on some analytic equivalence. In making this demand we rather claim that a theory must reduce the number of independent factors. Consequently, between two competing theories, one which names fewer distinct factors should be preferred.<sup>27</sup>

Phenomenalist hedonism meets the Integration demand by identifying the same ingredient of enjoyment and pain in the very different attitudes the subject might have. And if we begin with the premiss of crucial significance of happiness for

26 One such development of Sumner's idea is in Lebar (2004).

27 The Integration demand, as formulated here, has affinities with the 'pattern problem' discussed in Jackson *et al.* (2004, pp. 216–221) in relation to ethical particularism.

welfare, as seems to be the intent of C-hedonism, then there must be a systematic reason why their identity breaks down in deception scenarios. There must be a systematic reason linking happiness to the condition *C*. Unless we have such a reason, one is tempted to think that happiness makes the genuine contribution to welfare, while the contribution of *C* is illusory. Or alternatively, it would be *C* which makes the genuine contribution, while the significance of happiness would be illusory.<sup>28</sup>

Why accept the Integration demand? Though I have no crisp argument to offer, I can cite two sets of reasons, historical and systematic, to make it attractive. (1) In Book I of *Nicomachean Ethics* Aristotle equates welfare (*eudaimonia*) with a list of goods, such as virtue, health, friendship, family, wealth, and so forth. It may be asked if there is any one principle that holds all of them together. It may be thought that each good was added piecemeal, in order to deal with specific counter-examples. For instance, it would be implausible that a lonely person would turn out to be a happy one (at least according to Aristotle). Thus we must count friendship and family among the goods. But a closer look at the list reveals, I think, a tight connection. Aristotle divides those goods into goods of the soul, goods of the body, and 'external goods' (*ektos agatha*). Virtue and health comprise the first two classes, and the rest belong to the third one. But the role of this separate category of external goods is clearly instrumental. External goods are supposed to help us get the goods of the soul and body. Thus Aristotle says:

[Eudaimonia] needs the external goods as well; for it is impossible, or not easy, to do noble acts without the proper equipment. In many actions we use friends and riches and political power as instruments. (Aristotle, 1980, 1099a32–1099b1)

At the end we are left with the goods of soul and body. It is possible that health is also instrumental, facilitating the activity of the soul. This presumably would be so according to the doctrines of I.7-8 and X.7-8 alike:

But, being a man one would also need external prosperity; for our nature is not self-sufficient for the purpose of contemplation, but our body also must be healthy and must have food and other attention. (Aristotle, 1980, 1178b33–1178b35)

Then *eudaimonia* will be identified with excellence of the soul, evidently satisfying the Integration demand. But even if health is an autonomous good, as suggested by some readings of III.3, we should appeal to the theory of human nature of which soul and body would be essential components. Given the elements of the psycho-physical nature that we have, goods included in welfare reflect the excellence

---

28 Notice also that at this stage we cannot say that happiness and whatever factor mentioned in the condition *C* – e.g., knowledge or rationality – are parts of an organic whole. To say so is to assimilate C-hedonism to A-hedonism and to expose it to the same heterogeneity objection.

of those elements. Thus again, we should be able to meet the Integration demand.<sup>29</sup>

Other than historical, there are (2) closely related systematic considerations. A theory of welfare assigns goods suitable for every man (even if we restrict the quantifier domain here, there would still be a sufficiently large number of people our theory is supposed to serve). Then a theory of welfare must be able to explain why *X* and *Y* are the goods for this whole multitude of individuals. But what model of explanation would be appropriate here? Obviously neither a causal, nor a deductive-nomological one. The only suitable model, I think, would be the unification model. Unification here acts on two levels at once. There must (a) be a general pattern these individuals exhibit – or a pattern exhibited by the individuals and the environment they inhabit – that makes exactly *X* and *Y* suitable for all of them. So unless we want to admit a global improbable coincidence into our account, *X* and *Y* must have something in common. Thus in Aristotle's view (or one of his views, as mentioned above), their shared feature is the relation they bear to the psychophysical nature of men. The required generality in that view is reached by reference to the shared natural characteristics of the human species. But (b) there should be a standard sense of generality in which our theory recognizes a lesser number of 'brute facts', as compared to alternative theories.<sup>30</sup> Suppose we had a theory naming *X* as a sole factor of welfare, and we had another theory naming *Y* as its sole factor. Suppose further that both theories were plausible, as far as they go. It will not do to unify them by simply conjoining *X* and *Y*. There must be an account why *X* and *Y* together act to increase welfare. Such an account will reveal the underlying regularity responsible for the joint influence of *X* and *Y*.

Aristotle's discussion is once again instructive. In I.8 he lists common beliefs held about *eudaimonia* by 'many men', 'men of old', and 'a few eminent persons'. They alternately identified *eudaimonia* with virtue, practical wisdom, or philosophic wisdom, whether accompanied by pleasure and external prosperity (*eueteria*) or not. Aristotle's own view is supposed to unify these opinions which, he says, cannot be 'entirely mistaken' (1098b25–9). The instrumentalist view developed in the rest of I.8 may be interpreted in this light. External prosperity is a condition for virtuous acts, and virtuous acts are good for the soul. Pleasure is necessarily a state of the virtuous soul, since a virtuous man must take pleasure in his actions (1098b30–1099b8). Where the 'men of old' could attend only to some of the phenomena and form only a partial view of *eudaimonia*, Aristotle is now able to

---

29 The need of integration is similarly felt by W. D. Ross. After a lengthy enquiry supposed to name three diverse factors in individual welfare – virtue, knowledge and pleasure – he eventually says: "[B]y reflection on what we really think to be good, [this list] perhaps derives some support from the fact that it harmonizes with a widely accepted classification of the elements in the life of the soul" (Ross, 1930, p. 140).

30 See Friedman (1974).

bring their insights together by reference to the activity of the soul. This is a fine example of unification.

## 8. Conclusion

Let me recapitulate the main ideas of this article. (1) Standard hedonism affirms the psychological nature of pleasure. As such it is open to objections from illusion, as well as other structurally similar *reductio* arguments, such as the objection from bad pleasures. (2) In response to these objections, qualified S-hedonism modifies the way pleasure contributes to overall welfare. However, the concessions it offers are unsustainable, so far as pleasure is no longer the only source of value. (3) A-hedonism modifies the theory of pleasure by making objects of psychological attitudes intrinsic elements of pleasure. But this modification again leads to its convergence with a typical anti-hedonism. (4) It is an open question whether what was earlier termed ‘conditional hedonism’ is really meant to be a version of hedonism by its defenders. At all events it similarly carries crucial affinities to anti-hedonism. It is also plagued by problems of explanatory adhocness and lack of theoretical unity.

The moral of our story is twofold. Attempts to soften hedonism by allowing factors other than mental states to play a role in welfare lead us away from hedonism and to distinctively anti-hedonist views. And a theory which allows certain other factors along with pleasure to influence welfare will face accusations of adhocness.

On the other hand, if these modifications do not go through, committed hedonists should be impelled to re-examine the *reductio* arguments. They may dig deep by claiming that the purported rhetorical sting of these arguments tends to confuse the good-for-me with other qualities a life, or an action, might have. They may, therefore, put the onus on their opponents and challenge them to come up with a substantive theory showing why *foolproof* factual illusions, valuational illusions, bad pleasures, and the like decrease the value of life. They may further look at the cognitive mechanism of a thought experiment similar to the one employed in the argument from illusion. This thought experiment employs a deception never to be discovered by the subject. However, we, by looking at the condition the subject is in, are already aware of the ongoing deception. And it is we, outside the cloud of deception, so to speak, who are supposed to pass a judgement on the subject’s life. The cognitive setup of the thought experiment may well, therefore, be incoherent. This, I think, could be an exciting new avenue of research encouraged by the repeated failures of qualified hedonism.<sup>31</sup>

---

31 For a response in this direction see Brax (2009, p. 227).

## Acknowledgements

I am grateful to Yasemin Topac and two anonymous referees for helpful comments on an earlier draft of this article.

## References

- ANSCOMBE, G. E. M. (1981) "On the Grammar of 'Enjoy'." In *The Collected Philosophical Papers*. Vol. 2, pp. 94–102. Oxford: Blackwell.
- ARISTOTLE (1980) *The Nicomachean Ethics*. New York: Oxford University Press.
- BERKOVSKI, S. (2011) "Prichard's Heresy." *Philosophy* 86: 503–524.
- BRADLEY, B. (2002) "Is Intrinsic Value Conditional?" *Philosophical Studies* 107: 23–44.
- BRAX, D. (2009) *Hedonism as the Explanation of Value*. Lund: Media-Tryck Sociologen.
- BROADIE, S. (1991) *Ethics with Aristotle*. New York: Oxford University Press.
- BRONSTEEN, J., BUCCAFUSCO, C. and MASUR, J. S. (2010) "Welfare as Happiness." *Georgetown Law Journal* 98: 1583–1641.
- ENOCH, D. (2005) "Why Idealize?" *Ethics* 115: 759–787.
- FELDMAN, F. (2004) *Pleasure and the Good Life*. New York: Oxford University Press.
- FELDMAN, F. (2008) "Whole Life Satisfaction Concepts of Happiness." *Theoria* 74: 219–238.
- FELDMAN, F. (2010) *What Is This Thing Called Happiness?* New York: Oxford University Press.
- FRIEDMAN, M. (1974) "Explanation and Scientific Understanding." *Journal of Philosophy* 71: 5–19.
- GOSLING, J. C. B. and TAYLOR, C. C. W. (1982) *The Greeks on Pleasure*. New York: Clarendon Press.
- HAYBRON, D. M. (2008) *The Pursuit of Unhappiness*. New York: Oxford University Press.
- HURKA, T. (1998) "Two Kinds of Organic Unity." *Journal of Ethics* 2: 299–320.
- JACKSON, F., PETTIT, P. and SMITH, M. (2004) *Mind, Morality, and Explanation*. New York: Clarendon Press.
- KAGAN, S. (1998) "Rethinking Intrinsic Value." *Journal of Ethics* 2: 277–297.
- KATZ, L. D. (2009) "Pleasure." In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2009 edition.
- LEBAR, M. (2004) "Good for You." *Pacific Philosophical Quarterly* 85: 195–217.
- LEWIS, D. K. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
- MILL, J. S. (2003) "Utilitarianism." In J. Troyer (ed.) *The Classical Utilitarians: Bentham and Mill*. Indianapolis: Hackett.
- MOORE, G. E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- PALEY, W. (2002) *The Principles of Moral and Political Philosophy*. Indianapolis: Liberty Fund.
- PERSSON, I. (2006) *The Retreat of Reason*. New York: Clarendon Press.
- PRICHARD, H. A. (1935) "The Meaning of ἀγαθόν in the Ethics of Aristotle." *Philosophy* 37: 27–39.
- RÖNNOW-RASMUSSEN, T. (2011) *Personal Value*. New York: Oxford University Press.
- ROSS, W. D. (1930) *The Right and the Good*. Oxford: Oxford University Press.
- SCHROEDER, T. (2004) *Three Faces of Desire*. New York: Oxford University Press.

- SEN, A. (1985) "Well-Being, Agency and Freedom." *Journal of Philosophy* 82: 169–221.
- SIDGWICK, H. (1907) *Methods of Ethics*. New York: Macmillan, 7th edn.
- SIDGWICK, H. (2000) "Pleasure and Desire." In *Essays on Ethics and Method*, pp. 79–88.  
New York: Oxford University Press.
- SUMNER, L. W. (1995) "The Subjectivity of Welfare." *Ethics* 105: 764–790.
- SUMNER, L. W. (1996) *Welfare, Happiness, and Ethics*. New York: Clarendon Press.