



Explaining the Paradox of Hedonism

Alexander Dietz

To cite this article: Alexander Dietz (2018): Explaining the Paradox of Hedonism, Australasian Journal of Philosophy, DOI: [10.1080/00048402.2018.1483409](https://doi.org/10.1080/00048402.2018.1483409)

To link to this article: <https://doi.org/10.1080/00048402.2018.1483409>



Published online: 19 Jun 2018.



Submit your article to this journal [↗](#)




Article views: 34



View Crossmark data [↗](#)



Explaining the Paradox of Hedonism

Alexander Dietz 

University of Southern California

ABSTRACT

The *paradox of hedonism* is the idea that making pleasure the only thing that we desire for its own sake can be self-defeating. Why would this be true? In this paper, I survey two prominent explanations, then develop a third possible explanation, inspired by Joseph Butler's classic discussion of the paradox. The existing accounts claim that the paradox arises because we are systematically incompetent at predicting what will make us happy, or because the greatest pleasures for human beings are found in certain special goods which hedonists cannot enjoy. On the account that I develop, the paradox is a consequence of a theory about the nature of pleasure, together with a view about the requirements of rational belief. Which of these explanations is correct, I argue, bears on central questions about how to understand the nature and extent of the paradox.

ARTICLE HISTORY Received 28 August 2017; Revised 16 April 2018

KEYWORDS paradox of hedonism; pleasure; Joseph Butler; evidentialism

1. Introduction

The *paradox of hedonism* is the idea that making pleasure the only thing that we desire for its own sake can be self-defeating.¹ A number of philosophers have recognised this idea as describing a real and important phenomenon, and a potentially serious challenge for hedonistic versions of ethical theories such as egoism and utilitarianism. For example, Mill [1873], who famously defended hedonistic utilitarianism, reports that he experienced the paradox first-hand in the form of a mental crisis, and that he found the idea so striking that he came to accept it as the foundation of his new 'theory of life'. (Mill refers to 'happiness' rather than 'pleasure', but, like others who discuss the paradox, typically uses these terms interchangeably.² For convenience, I will follow this usage here.) He writes [ibid.: 117–18]:

I never, indeed, wavered in the conviction that happiness is the test of all rules of conduct, and the end of life. But I now thought that this end was only to be attained by not making it the direct end. Those only are happy (I thought) who have their minds fixed on some object other than their own happiness; on the happiness of others, on the improvement of mankind, even

¹ The classic statement of the paradox can be found in Butler [1726: especially 364–8]. For other influential discussions, see Sidgwick [1874: 136], Stocker [1976], and Railton [1984]. For a general introduction to the paradox, see Eggleston [2013].

² Butler [1726: 364–8] likewise uses 'happiness' and 'enjoyment' interchangeably in his discussion of the paradox. According to Irwin [2008: 501–2], English moralists from Hobbes to Sidgwick generally identify happiness with pleasure.

on some art or pursuit, followed not as a means, but as itself an ideal end. Aiming thus at something else, they find happiness by the way ... Ask yourself whether you are happy and you cease to be so.

Now, the idea that the best way to obtain pleasure is to care about things other than pleasure itself may strike us as obvious common sense. Even so, we might wonder: why might the paradox of hedonism be true? Why would pleasure's being the only thing that we desire for its own sake make us less happy? After all, many desires are not generally self-defeating. For example, we normally expect those who ultimately care about nothing other than doing well in school, sports, or work to do better than those who don't. And it is not mysterious why this would be true: if you ultimately care about only one thing, you will likely be especially motivated to pursue this thing, and will not be willing to sacrifice it for the sake of other things. So why doesn't this work for pleasure?

In this paper, I will survey three possible explanations of the paradox. The first two are currently prominent in the literature: one claims that we are systematically incompetent at predicting what will make us happy, while the other claims that the greatest pleasures for human beings can be found in certain special goods which hedonists cannot enjoy. I will next discuss an older type of account, based on insights offered by Butler [1726], according to which the paradox arises in part because of the nature of pleasure itself. While initial versions of this account face serious objections, I will try to show that Butler's insights nevertheless represent a promising and philosophically interesting alternative strategy for explaining the paradox. I will do this by offering a version of this account that avoids the earlier objections, making use of a recent theory of pleasure, and a leading view about the requirements of rationality.

I will then argue that which, if any, of these explanations is correct—or which ones, if the paradox turns out to be overdetermined—bears on two central questions about how to understand the paradox. First, does the paradox arise only because of human limitations or mere accidents, or would it arise even for seemingly ideal agents? Second, is the problem limited to those who only desire their own pleasure, or does it also extend to those who desire the pleasure of others? I will argue that the Butler-inspired account that I will develop, unlike the other two accounts, implies that the paradox does arise even for ideal agents. And this account also implies that the problem does extend to more altruistic forms of hedonism. This is because the problem, on this account, is not essentially about selfishly desiring one's own pleasure, but is rather about the nature of pleasure as an object of desire.

2. The Incompetence and Special-Goods Accounts

In the contemporary literature, philosophers who discuss the paradox tend to focus on two main accounts. The first account, as Eggleston [2013: 3794–5] writes,

appeals to the claim that when people have choices to make, they are generally poor judges of their available options' effects on happiness ... In this view, people are systematically so inept at making happiness-promoting choices that a surer route to happiness is for people to aim, when making choices, at objectives other than the promotion of happiness ...

We can call this the *incompetence* account. For example, Eggleston writes, a proponent of this account might claim 'that people generally overestimate the extent to which

additional income will increase their happiness, and generally underestimate the extent to which a longer commute between home and work will decrease it' [ibid.].

As Haybron [2008: 225–51] has argued, there is significant empirical evidence supporting the idea that people are systematically incompetent at predicting what will make them happy. In particular, Haybron argues, human beings seem to suffer from a wide variety of cognitive biases that interfere with our ability to make such predictions. For example, he cites numerous studies indicating that, when we are 'predicting how an event will make us feel, we typically imagine what it will be like at the time of the event and then project this feeling long into the future—far longer, in general, than it will actually last' [ibid.: 231].

The other prominent account claims that, for most people, happiness might depend on acquiring certain particular kinds of goods which hedonists are unable to have [Eggleston 2013]. We can call this the *special goods* account. One version of this account is offered by Stocker [1976: 457], who focuses specifically on egoistic hedonists, or those who care only about their own pleasure. Stocker writes that 'love, friendship, affection, fellow feeling, and community are among the greatest (sources of) personal pleasures', but that egoistic hedonists cannot obtain these goods. After all, egoistic hedonists, who do everything purely for the sake of their own pleasure, 'cannot act for the sake of the intended beloved, friend, and so on; thus, they cannot love, be or have a friend, and so on'.

Empirical evidence supports the idea that at least some of the goods found on Stocker's list really are among the greatest sources of pleasure for human beings. For example, in his overview of psychological research on what leads to happiness, Haidt [2006: 94] notes that '[t]he condition that is usually said to trump all others in importance is the strength and number of a person's relationships.'

Of course, for the special goods account to succeed, we would also have to think that the relevant goods really are difficult or impossible for egoistic hedonists to obtain. Would it really be impossible in principle for these hedonists to enjoy relationships of love or friendship, as Stocker seems to suggest? Or do we at least have reason to think that they will be especially unlikely to be able to form such relationships? These are difficult questions, and I will not try to resolve them here.

Overall, however, the incompetence and special-goods accounts both seem to be straightforward and plausible explanations of the paradox of hedonism, and I will not offer arguments against them in this paper. Nevertheless, I suggest that it is also worth considering an alternative account, which I will develop in the next section. This alternative is worth considering for two reasons. First, on both of the above accounts, the paradox seems to be merely an empirical phenomenon, without any special philosophical significance. After all, these accounts ascribe the paradox largely to certain seemingly mundane empirical facts about human psychology—that human beings happen to suffer from certain cognitive limitations, or that we happen to derive more pleasure from some kinds of things than from others. In contrast, I will suggest that there is a more philosophically interesting alternative. Second, as we will see, if this alternative account does turn out to be correct, this will have important implications for our understanding of the paradox.

3. The Evidentialist Account

3.1. Introduction

In this section, I will develop my alternative account. This account is inspired by Butler's classic discussion of the paradox. Rather than attributing the paradox to our

incompetence, or to the idea that we need certain special goods in order to make us happy, Butler claims that the paradox is, in part, a consequence of the nature of pleasure itself.

There are two main camps in the contemporary philosophical debate concerning the nature of pleasure. *Phenomenal* theorists identify pleasure with phenomenal qualities. For example, we might think that pleasure is just a certain kind of feeling, such as a tingle or a warm glow. In contrast, *attitude* theorists claim that what makes an experience pleasant consists in having certain attitudes, such as the attitude of desiring the experiences that one is having.³ In his discussion, Butler endorses an attitude theory of pleasure. In particular, he suggests that desires play a central role in the nature of pleasure. And he suggests that this fact can explain the paradox.

Unfortunately, as we will see, Butler's discussion of how, exactly, this is supposed to work is open to multiple interpretations, and some natural interpretations do not succeed in giving a plausible account of the paradox. However, I will try to show that his insights nevertheless represent a promising explanatory strategy. I will do this by offering a new account along Butler's lines that avoids the objections facing other interpretations. I will not try to argue that this account is what Butler himself is most likely to have had in mind, and I will not try to argue that we have conclusive reasons to accept it. But I will try to show that this account represents a plausible and philosophically interesting alternative to the accounts that we considered earlier.

My account will develop Butler's insights by showing how the paradox follows from a more recent theory of pleasure, together with a leading theory of rationality. According to the theory of pleasure, pleasure requires not only having certain desires, but also having certain beliefs. And, according to the theory of rationality, rational belief requires evidence. We can call the resulting explanation of the paradox *the evidentialist account*.

3.2. Butler's Insights

Butler's account of the paradox is brief but intriguing. It starts with a view about the nature of pleasure. For Butler, 'the very idea of interest or happiness consists in this, that an appetite or affection enjoys its object' [1726: 334]. And the reason why the desire for happiness is self-defeating, according to Butler, is that 'happiness consists in the gratification of particular passions, which supposes the having of them' [ibid.: 367]. Without these passions, self-love, or the desire for one's own happiness, is left with 'absolutely nothing at all to employ itself about' [ibid.: 334].

I will follow other commentators in supposing that Butler's 'appetites,' 'affections,' and 'passions,' are the same as or at least involve what philosophers would today call 'desires' (see, for example, Irwin [2008: 498–9]). I will also assume that the relevant desires are 'intrinsic' desires, or desires for things for their own sake. (I will use 'desire' as shorthand for 'intrinsic desire' in what follows.)

Thus, in these passages, Butler seems to be suggesting two key insights:

- (1) The nature of pleasure involves a close connection to the satisfaction of desires.⁴

³ For a brief overview of the debate between phenomenal and attitude theories, see Heathwood [2011: 90–2].

⁴ I am using 'satisfaction' in a thin sense here: for my desire for some state of affairs to be satisfied is just for that state of affairs to be the case, whether or not I know about it.

- (2) Because of the way in which the nature of pleasure is connected to the satisfaction of desires, we cannot get pleasure unless we also have desires for objects other than our own pleasure.

Butler's first insight reflects an attitude theory of pleasure, the main contemporary competitor to phenomenal theories of pleasure. While I will not try to settle the debate between attitude and phenomenal theories here, I will point out there are reasons to take attitude theories seriously. For one thing, there are serious objections to phenomenal theories. For example, phenomenal theorists, who identify pleasure with phenomenal qualities, face the challenge that pleasant experiences of very different kinds do not seem to contain any common quality. This may not be a fatal objection, but it does give us reason to consider alternative theories. In addition, it seems hard to imagine taking pleasure in something without having some positive attitude toward that thing, and it is plausible that this is because such attitudes are somehow built into the nature of pleasure.

Butler's second insight is vague, but promising. It seems promising to think that if pleasure has to do with getting what we want, this might somehow explain why, to get any pleasure, we would need first to desire something other than pleasure itself.

But what is the connection between pleasure and the satisfaction of desires, and why, exactly, does this connection make pleasure require non-hedonistic desires? As we will see, there are several ways that we might try to answer these questions, some of which are more successful than others.

3.3. First Pass: Pleasure as the Satisfaction of Non-Hedonistic Desires

For example, on one natural interpretation, Butler is claiming, roughly, that pleasure just is the satisfaction of desires of a certain kind, those that he calls 'particular passions'. Butler distinguishes these passions from what he calls self-love, or the desire for one's own happiness. Particular passions have a great variety of objects; they include 'the appetites of sense, resentment, compassion, curiosity, ambition, and the rest' [1726: 19]. These passions are 'external' and 'particular', whereas self-love is 'internal' and 'general'. Now, there are difficulties in trying to make sense of these contrasts (see Phillips [2000: 424]). But, for our purposes, it is enough to note that the list of particular passions is not supposed to include self-love. In that case, we have a very simple account of the paradox: pleasure just is the satisfaction of desires of a certain kind, a kind that does not include the desire for pleasure itself. So, someone who only desires pleasure will not be able to get any.

Unfortunately, the theory that pleasure just is the satisfaction of a certain set of non-hedonistic desires seems implausible. It seems that we *can* take pleasure in our own pleasure, and not merely in the gratification of distinct particular passions. Butler himself refers to 'the pleasure self-love would have from knowing I myself should be happy some time hence' [1726: 366]. So, this initial proposal seems unsatisfactory.

3.4. Second Pass: Pleasure as the Satisfaction of Desires

What if we remove the restriction to particular passions, and claim simply that pleasure just is, or at least requires, the satisfaction of desires of any kind?

It seems that, even when we relax our account of pleasure in this way, we might still be able to generate something like the paradox. Suppose that pleasure, by its nature, requires the satisfaction of desires, but that my only intrinsic desire is for pleasure itself. Now it seems that I have found myself in a kind of circle: I can only get pleasure in virtue of the fact that my desire is satisfied, but my desire can only be satisfied in virtue of the fact that I am getting pleasure. But explanation, we might think, is asymmetric, and so can't go in circles: A can't both explain B and be explained by B at the same time. So, if the only way for a hedonist to get pleasure would be through an explanatory circle, we might think, he must be out of luck.

Unfortunately, even if the theory that pleasure just is or requires the satisfaction of desires could give us a version of the paradox, it is not a very plausible theory of pleasure. For it seems clear that we can sometimes get pleasure even when our desires are not actually satisfied, but when we falsely believe that they are. For example, Al Gore may have briefly been happy as a result of believing that he had won the 2000 presidential election [Heathwood 2006: 556].

3.5. *Third Pass: Desire, Belief, and Rationality*

How can we fix this problem? Rather than claiming that pleasure requires that we actually get what we want, we could claim that pleasure requires that we believe that we are getting what we want. That is, we could claim this:

I am pleased at t only if and because, for some state of affairs p , I intrinsically desire at t that p and I believe at t that p .

We can call this the 'desire-belief' (DB) condition for pleasure. DB is the claim about the nature of pleasure that I will rely on, in the account of the paradox that I will develop below.

DB is based on a recent theory of pleasure offered by Heathwood [ibid.: 557]. He proposes, roughly, that pleasure just consists in intrinsically desiring some state of affairs and believing that state of affairs to be the case (for similar views, see Davis [1981: 113] and Schroeder [2004: ch. 3]). However, for our purposes, we will only treat these things as necessary conditions for pleasure.

Note that DB avoids the objections that plagued the earlier claims that we considered. DB does not preclude the desire for our own pleasure from itself giving rise to pleasure. DB also allows for pleasure to arise from false beliefs about the satisfaction of our desires.⁵ And DB seems plausible on reflection. For example, it seems plausible that I will not be happy if I have absolutely no desire for things to be the way that I believe they are, or if I do not believe that things are the way that I want them to be.

However, DB is not enough on its own to generate the paradox. If the only thing that I intrinsically desire is pleasure, DB does not preclude me from getting it. It just says that I need to believe that I now, or at least will at some point, have pleasure in my life, and this belief might thereby come true. So, if we want to use DB in our account of the paradox, we need to know what could stop me from forming this belief.

Now, there might simply be some empirical fact about human psychology that would explain why hedonists would have trouble believing in their own pleasure. But

⁵ For discussion of other objections to desire-based theories of pleasure, see Heathwood [2006, 2007].

we can get a more philosophically interesting account, I will argue, by combining DB not merely with a claim about human belief, but with a claim about rational belief. If we accept a leading view about rational belief, I will argue, we can explain why any rational hedonist would have trouble believing in his own pleasure, and so would have trouble finding it.

There are several competing views about what kinds of norm govern belief. Some philosophers claim that it can be rational to form beliefs for intuitively ‘pragmatic’ reasons, a view illustrated by Pascal’s famous argument that we should believe in God because doing so would be in our best interests. This kind of view does not look promising for our purpose of explaining why a rational hedonist would have trouble believing in his own pleasure. A rational hedonist would seem to have excellent pragmatic reasons to believe in his own happiness, since doing so is necessary for him to become happy, and this is what he fundamentally wants.

Another popular idea is that our fundamental epistemic goal should be to acquire *true* beliefs (see David [2001]). This view is one interpretation of the popular metaphors that belief ‘aims’ at the truth, and that belief, in contrast to desire, has a ‘mind-to-world direction of fit’: that is, beliefs are states that are supposed to fit the world, whereas desires are states that the world is supposed to fit (see Smith [1994: 111–16]). But, as we’ve seen, DB allows that if all that I want is to be happy, once I form the belief that I will be happy, this belief could thereby come true. So, if rational belief simply aims at truth, then it seems that a rational hedonist might easily be able to acquire pleasure.

Instead, I suggest that we focus on a version of *evidentialism*, the view that a rational agent will form beliefs only on the basis of evidence. While we have seen that there are compelling alternative views, evidentialism seems highly plausible, and is currently a leading view among epistemologists [Chignell 2016].

While evidentialism might seem similar to the view that our epistemic goal should be to acquire true beliefs, these views can come apart. For example, as Dorr [2002: 99–100] writes, suppose that ‘you believed yourself (perhaps with good reason) to be watched over by a benevolent spirit, who sees to it that whenever you form a belief on a certain subject matter (say, the winners of horse-races), it is true’. Even so, Dorr suggests, ‘it is epistemically irrational for you to form a belief one way or the other in the absence of evidence’. (See also Setiya [2008: 399–401].) One way to understand the contrast here is to see evidentialism as an alternative interpretation of the metaphors that belief aims at the truth, and that belief has a mind-to-world direction of fit. In particular, we might think, when we are trying to decide what to believe on the basis of evidence, we are trying to discover what facts are already ‘out there’, independent of our beliefs, in some sense. Beliefs formed in response to evidence are supposed to ‘aim’ at, or ‘fit’, an independently existing reality. More precisely, I propose that we rely on the following view of rational belief:

Independence: A rational agent will believe some proposition only in response to what she takes to be evidence that that proposition is true regardless of whether or not she then comes to believe it.

If I am rational, then, and I know that some proposition will be true only if I now come to believe it, I will not believe it.

I will next show how, when we combine this view about rational belief with the theory that pleasure requires corresponding desires and beliefs, we can make good on

Butler's insights, and generate a new explanation of the paradox of hedonism—namely, the evidentialist account.

As an aside, it is worth noting that another strategy that we might pursue to get a philosophically interesting account of the paradox would be to claim, not that we cannot rationally believe what we know is not independently true, but rather, as Langton [2004] has suggested, that the nature of belief simply makes it impossible to form such beliefs. I will rely on the claim about rational belief in what follows, but if you find this alternative strategy more plausible, it will be easy enough to see how my arguments could be modified accordingly.

3.6. *The Account*

I will now lay out the evidentialist account, showing how our assumptions about the nature of pleasure and about the requirements of rationality imply a version of the paradox of hedonism. On this version of the paradox, hedonism will be self-defeating in so far as the hedonist is epistemically rational, and is not deceived in certain ways.

Let's start by supposing that I am a rational and well-informed hedonist throughout my life. How might I become happy?

According to DB, I can be happy only if, and because, there is some state of affairs that I both desire for its own sake, and believe to be the case. Since I am a hedonist, the only thing that I desire for its own sake is happiness or pleasure itself. So, to be happy, I'll need to believe that I'm happy, or have been, or will be. If I know that I've been happy in the past, then, I might be able to get pleasure from this knowledge. But how can I get any pleasure to begin with?

If I haven't yet been happy, and I don't falsely believe that I have been, then DB implies that I'll need to believe that I am now or will in the future be happy. But if I'm epistemically rational, Independence implies that I will form this belief only in response to what I take to be evidence about what is true independent of my now coming to believe it. But if I know that I'll be happy only if I sooner or later come to believe it, and I know that I am and will continue to be a rational and well-informed hedonist, I will now argue, then I won't believe it, and so I will never be happy.

Again, I'll be happy only if and because I form a belief in my own happiness. Since we are assuming that I am well informed, we are assuming that this belief would have to be true. And if it were true, this would have to be because of some belief in my own happiness—either this initial belief, or some belief in my own happiness that I will form sometime later. But we cannot rely on the first alternative: since I am well informed and epistemically rational, Independence implies that I will not form a belief in my own happiness that I know would be true only if I then form that very belief. Next, suppose that I will be happy because of some future belief in my own happiness—say, one that I form tomorrow. This belief could in turn be true because of itself, because of some earlier such belief, or because of some later such belief. Once again, since I will continue to be well informed and epistemically rational, Independence implies that I will not form a belief that I know will be true only if I then form that very belief. And, since I know that I will continue to be well informed and epistemically rational, I know that I will not form such a belief in these circumstances. And if tomorrow's belief will only be true because of some other such belief, we can again ask why that belief could be true, and repeat the process.

Since the truth of each belief would have to be explained either by that belief itself, by some earlier belief, or by some later belief, and we have ruled out the first option, there would have to be either an explanatory regress into the past or future, or a kind of explanatory circle, where two or more beliefs would make each other true. If I have a finite lifespan, however, the explanatory regress will have to end somewhere: I can't be happy every day because I truly believe that I was happy the day before, or will be happy the next day, because I must have started being happy at some point (as we have been assuming), and I won't always be able to rely on being happy in the future, since I will die at some point. My last hope is an explanatory circle: perhaps I will be happy tomorrow because I will believe that I was happy today, and I am happy today because I believe that I will be happy tomorrow. But, since I will continue to be well informed tomorrow, I will believe that I was happy today only if I really am happy today. But I will be happy today only if I now believe that I will be happy tomorrow. So, I will be happy tomorrow only if I now believe that I will be happy tomorrow. And Independence implies that if I know that this is true then I cannot rationally form this belief.

The details of this reasoning are complicated, but the basic idea is simple: if I know that I will be happy only if I believe in my own happiness, and that I will continue to be rational and well informed, then there will be no way for me to find independent grounds for forming this belief, and so I will never be happy.

This basic idea behind our hedonist's predicament turns out to have the same structure as Cave's [2001] *placebo paradox*. Cave describes a sick woman, Miranda, who is given a placebo pill. Because of the way that placebos work, Miranda will get better only if she believes that she will get better, just as our hedonist will be happy only if he believes that he is or will be. But if Miranda is rational, then she will only form the belief that she will get better if she gets evidence that it's true. Now, she might acquire such evidence through being misinformed or ignorant in certain ways. For example, Cave suggests that Miranda might initially believe that the pill has pharmaceutical curative properties. But if she learns the truth, Cave writes, then her belief that she will get better will be undermined, and so she will not get better.

Like Miranda, our hedonist could escape the problem if he is misinformed or ignorant in certain ways. For example, my belief that I will be happy might be formed on the basis of inductive evidence, such as the fact that happiness runs in my family. But, in that case, I will have to be either misinformed or ignorant about a crucial question: what is it, exactly, that will make me happy? I might falsely believe that I will be able to enjoy the sorts of things that make other people happy, such as travelling. This belief will be false because, as a hedonist, I will not desire travelling for its own sake, so travelling will not itself make me happy. Alternatively, I might simply fail to form any belief as to what will make me happy. But if I know that the only thing in which I will take pleasure is my pleasure itself, and so that the idea that I will be happy is not independently true, but only something that I can make true by believing it, and I am rational, then I will not believe it.

Likewise, we can see that if I were not a hedonist, or were not rational, then I might have no trouble in finding pleasure. For example, suppose that I were a non-hedonist: suppose that, rather than desiring only pleasure, I also had an intrinsic desire to run. Then I might simply start running, see and thus believe that I am running, and therefore be happy. My pleasure would thus have a solid independent foundation. Likewise, suppose that I were an irrational hedonist: I might then be happy because I come to believe I am, despite a lack of evidence. My pleasure would thus not need a foundation.

But if I am both rational and a hedonist, then my pleasure would need, but (barring misinformation or ignorance) could not be given, an independent foundation.

Now, we might think that this argument, if successful, would imply a particularly extreme version of the paradox of hedonism: it would show not only that hedonism is self-defeating, but that it is impossible for hedonists to find any pleasure whatsoever. And this conclusion, we might worry, is implausibly strong. Surely we know from experience that hedonists are not completely incapable of finding pleasure.

However, recall that we have been focusing on people who meet certain idealising assumptions—people who are rational, relevantly well-informed, and hedonistic throughout their lives. I believe that my argument has indeed shown that DB and Independence imply that these people would not be able to find any pleasure. But it seems likely that real-world hedonists are not perfectly rational, well-informed, or hedonistic throughout their lives. As a result, it might be that, when we relax our assumptions, we can instead predict a less extreme version of the paradox.

While I do not have the space here to fully defend this suggestion, here is an illustration of what it might look like to relax our assumptions. Suppose, for example, that I was not always a hedonist, and that, before I became a hedonist, I had some pleasant experiences. In that case, DB does allow that, even after I become a hedonist, I might still be happy if I can remember that I had these experiences. After all, if I intrinsically desire my own past pleasure, then DB does allow that I can get some pleasure if I (correctly) believe that I was once happy. But if we think that getting any pleasure depends on desiring what we believe to be the case, then it would be natural to expect that how much pleasure we get depends on how strong this desire is. So, as long as I do not intensely desire to have once been happy, I will not be able to derive much additional pleasure from my belief that I was once happy.⁶

Now, even if the extreme conclusion that pleasure turns out to be impossible applies only to highly idealised hedonists, some readers might find even this to be incredible. These readers might then take my argument as a *reductio ad absurdum*, showing that we must reject either DB or Independence.

Despite this, it is important to point out that these readers might still think that an argument along similar lines could provide an explanation of the paradox of hedonism. After all, we might think that, even if DB and Independence are not correct accounts of the nature of pleasure and of rational belief, respectively, similar principles might describe important facts about human psychology. In particular, it might be that a major source of pleasure for human beings is the belief that the things that we want for their own sake have been achieved, and that people tend to form beliefs only in what they take to be independently true. Since these principles would only be generalisations, this version of the argument would allow us to avoid the conclusion that pleasure is impossible even for idealised hedonists. And it could still provide an interesting alternative explanation of the paradox, albeit one that, like the incompetence and special-goods accounts, relies centrally on empirical facts about human psychology.

However, for my own part, I do not find it absurd to think that pleasure might be impossible for highly idealised hedonists. For this reason, I believe that the more

⁶ There is a bootstrapping issue here: once I get this additional moderate pleasure, I might then recognise this fact, and this belief might in turn make me somewhat happier; and so on. However, this process might have diminishing returns.

ambitious version of the argument is still worth exploring, and so I will continue to focus on this version in what follows.

4. Comparing the Accounts

4.1. *Does the Paradox Arise from Our Weaknesses, or from Our Strengths?*

We have now seen three explanations of the paradox of hedonism—the incompetence account, the special-goods account, and the evidentialist account. We will now consider how these accounts can influence our understanding of the paradox. In this section, I will argue that which account we accept has implications for whether exclusively desiring pleasure is self-defeating because of psychological features that seem like weaknesses, or at least neutral, or because of features that seem like strengths.

On the incompetence account, hedonists end up defeating their own desires because human beings are systematically incompetent at predicting what will make them happy. It seems fair to say, then, that this account attributes the paradox to human deficiencies. Plausibly, an ideal agent would be one who did not suffer from cognitive biases, and so did not run into the same problem.

On the special-goods account, hedonists end up as less happy than other people because the greatest pleasures for human beings can only be found in certain special goods, such as friendship and community, that hedonists would have trouble in obtaining. Now, it does not seem like a deficiency to get more pleasure from something like friendship than from other goods that a hedonist might have less trouble getting, such as the experience of hiking. But, plausibly, there would not be anything wrong with creatures who happened to get more pleasure from hiking than from friendship. So, the special-goods account seems to attribute the paradox to a mere accident of human psychology, with no special intrinsic value.

If the paradox does arise either from our weaknesses or merely neutral accidents of our psychology, this would be significant. As Eggleston [2013] writes, it may be tempting to see the paradox as a problem for hedonistic ethical theories, such as hedonistic versions of egoism or utilitarianism. After all, the paradox implies that these theories are, in a potentially troubling way, self-defeating: if we all uniquely desire the goal that these theories propose, we will do worse at achieving it. But, as we saw earlier, despite recognising the paradox, Mill reports that he ‘never, indeed, wavered in the conviction that happiness is the test of all rules of conduct, and the end of life’. And if the paradox arises only because of our deficiencies or merely neutral accidents of our psychology, then it seems that ethical hedonists like Mill have no reason to worry. As Eggleston [2013: 3796] writes, many philosophers would claim that hedonistic ethical theories are self-defeating in the above sense merely ‘because of an accident of human psychology, rather than because of any defect internal to the theory itself. The paradox of hedonism, these philosophers would claim, ‘may capture a genuine fact about human nature—one with important practical implications—but it has no bearing on theoretical projects claiming that happiness must, in the last analysis, be acknowledged as the foundation of morality or instrumental rationality’ [ibid.].

However, on the evidentialist account, the paradox arises not because of features that seem like weaknesses or merely neutral accidents, but because of features that seem like good things. Whereas the incompetence account attributes the paradox to our irrationality and lack of self-knowledge, the evidentialist account claims the

opposite: hedonists will be unable to find pleasure, on this account, in so far as they form their beliefs rationally, and in so far as they are not deceived about their own psychology. And these seem like desirable qualities, qualities that we would typically imagine idealised versions of ourselves as having.

Now, whether the paradox of hedonism ultimately poses a successful challenge to hedonistic ethical theories is not something that I will try to resolve here. But we can conclude that if the evidentialist account is correct then ethical hedonists cannot dismiss such challenges on the grounds offered above.

4.2. Is the Paradox Ultimately about Selfishness?

Next, many philosophers treat the paradox as essentially a problem of selfishness. For example, Butler introduces the paradox in order to argue that there is no necessary conflict between self-love, or the desire for one's own happiness, and benevolence, since we cannot make ourselves happy unless we have desires for things other than our own happiness, such as the happiness of others [1726: 364–8]. And we saw that Mill likewise suggests that we can escape the problem by caring more about the happiness of others.⁷ However, other philosophers, including Sidgwick [1874: 405–6] and Parfit [1984: 27–8], have suggested that the paradox might also have non-egoistic versions.

Does the paradox essentially have to do with selfishness? That is, is the paradox limited to the desire for one's own pleasure, or might more altruistic desires for pleasure also be self-defeating? In this section, I will argue that how we explain the paradox can shed light on this question.

Recall that, on the special-goods account, exclusively desiring pleasure is self-defeating because the greatest sources of pleasure for human beings are found in certain particular sorts of goods that hedonists cannot obtain. According to Stocker [1976: 457], the relevant goods are found in certain kinds of relationships with other people, including 'love, friendship, affection, fellow feeling, and community'.⁸ But, clearly, this problem arises not from the desire for pleasure as such, but rather from the desire for one's *own* pleasure. It seems plausible that those who desire the pleasure of *other* people would not necessarily have any problem, or would at least have less of a problem, in finding love, friendship, and so on. So, Stocker's version of the special-goods account suggests that the paradox is mainly a problem for egoistic hedonists, not altruistic hedonists.

The evidentialist account, however, has different implications. I will now show that, under certain conditions, the evidentialist account implies that non-egoistic hedonists face analogues to the problems facing their egoistic counterparts.

Consider a group of altruistic hedonists—those who desire only the pleasure of others in the group. For simplicity, suppose that you and I are the only members of this group. So, my only intrinsic desire is for your pleasure, and your only intrinsic desire is for my pleasure. (Perhaps we are madly in love.) Thus, DB implies that you will be happy only if and because you believe that I will be happy at some point, whether in the past, present, or future, and I will be happy only if and because I believe that you

⁷ Other such philosophers include Sober [1992: 97], who, in his commentary on Butler, defines hedonism as a species of egoism, and Setiya [2014: 4], who describes Mill as discussing 'the paradox of egoism'.

⁸ However, not all special-goods accounts focus on these sorts of goods: for example, Eggleston [2013] mentions rewarding work as another possible privileged source of pleasure.


will be happy. This situation is an interpersonal analogue to the intertemporal explanatory circle that we encountered earlier, when we considered whether I could be happy today because I believe that I will be happy tomorrow, and happy tomorrow because I will believe that I was happy today. And similar problems arise here. Suppose that I am trying to determine whether you will be happy. Again, you will be happy only if and because you believe that I will be happy. And if you are well informed, you will believe that I will be happy only if it is true that I will be happy. And I will be happy only if I believe that you will be happy. So, you will be happy only if I believe that you will be happy. Once again, Independence implies that, if I know that this is true, I cannot rationally believe that you will be happy, and so I will not be happy. And, since parallel reasoning applies to you, if you are well informed then you will not rationally be able to believe that I will be happy, and so you will not be happy either.

We have now seen that if the evidentialist account is correct, then a version of the paradox of hedonism will re-emerge in groups of altruistic hedonists. This suggests that the problem ultimately arises because of the object of our desire—pleasure—rather than because we only want this object for ourselves. In particular, the problem arises because pleasure is by its nature parasitic: it requires an independent foundation in the satisfaction of non-hedonistic desires, as long as we are rational and well-informed. So, if all of us only desire pleasure, then, regardless of whether or not our desires are selfish, we will have nothing in which to take pleasure. In Butler's terms, benevolence, and not only self-love, will have 'absolutely nothing at all to employ itself about' [1726: 334].

5. Conclusion

In this paper, we have seen three proposed explanations of the paradox of hedonism—the incompetence account, the special-goods account, and the evidentialist account. While the evidentialist account is my own contribution, I do not claim to have shown that it is correct. As we have seen, this account rests on controversial assumptions: attitude-based theories of pleasure and evidentialist theories of rational belief both have formidable rivals. Rather, what I hope to have shown is that Butler's insights still represent a promising explanatory strategy. This kind of explanation deserves to be considered as a serious contender, I believe, because it is philosophically interesting in itself, because it can be made highly plausible, and because it can have striking implications for what we should think about the nature and extent of the paradox.⁹

ORCID

Alexander Dietz  <http://orcid.org/0000-0003-2469-2227>

References

Butler, Joseph 1726 (1991). *Fifteen Sermons Preached at the Rolls Chapel*, in *British Moralists, 1650–1800*, Vol. 1, ed. D.D. Raphael, Indianapolis, IN: Hackett: 325–77.

⁹ For helpful discussion and feedback, I would like to thank Ara Astourian, Stephen Finlay, Joe Horton, Nathan Robert Howard, Abelard Podgorski, Jonathan Quong, Mark Schroeder, Ralph Wedgwood, Jonathan Wright, and audiences at the Rocky Mountain Ethics Congress, the Eastern and Pacific Division Meetings of the American Philosophical Association, and the British Society for Ethical Theory.

- Cave, Peter 2001. Too Self-Fulfilling, *Analysis* 61/270: 141–6.
- Chignell, Andrew 2016. The Ethics of Belief, *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*, ed. Edward N. Zalta, URL = <http://plato.stanford.edu/entries/ethics-belief>
- David, Marian 2001. Truth as the Epistemic Goal, in *Knowledge, Truth, and Duty: Essays on Epistemic Justification*, ed. Matthias Steup, New York: Oxford University Press: 151–69.
- Davis, Wayne 1981. A Theory of Happiness, *American Philosophical Quarterly* 18/2: 111–20.
- Dorr, Cian 2002. Non-Cognitivism and Wishful Thinking, *Noûs* 36/1: 97–103.
- Eggleston, Ben 2013. Paradox of Happiness, in *The International Encyclopedia of Ethics*, ed. Hugh LaFollette, Malden, MA: Wiley-Blackwell: 3794–9.
- Haidt, Jonathan 2006. *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*, New York: Basic Books.
- Haybron, Daniel M. 2008. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, New York: Oxford University Press.
- Heathwood, Chris 2011. Desire-Based Theories of Reasons, Pleasure, and Welfare, in *Oxford Studies in Metaethics*, Vol. 6, ed. Russ Shafer-Landau, Oxford: Oxford University Press: 79–106.
- Heathwood, Chris 2006. Desire Satisfactionism and Hedonism, *Philosophical Studies* 128/3: 539–63.
- Heathwood, Chris 2007. The Reduction of Sensory Pleasure to Desire, *Philosophical Studies* 133/1: 23–44.
- Irwin, T.H. 2008. *The Development of Ethics: A Historical and Critical Study: Vol. II: From Suarez to Rousseau*, Oxford: Oxford University Press.
- Langton, Rae 2004. Intention as Faith, in *Agency and Action*, ed. John Hyman and Helen Steward, Cambridge: Cambridge University Press: 243–58.
- Mill, John Stuart 1873 (1989). *Autobiography*, New York: Penguin.
- Parfit, Derek 1984. *Reasons and Persons*, Oxford: Clarendon Press.
- Phillips, David 2000. Butler and the Nature of Self-Interest, *Philosophy and Phenomenological Research* 60/2: 421–38.
- Railton, Peter 1984. Alienation, Consequentialism, and the Demands of Morality, *Philosophy and Public Affairs* 13/2: 134–71.
- Schroeder, Timothy 2004. *Three Faces of Desire*, New York: Oxford University Press.
- Setiya, Kieran 2008. Practical Knowledge, *Ethics* 118/3: 388–409.
- Setiya, Kieran 2014. The Midlife Crisis, *Philosophers' Imprint* 14/31: 1–18.
- Sidgwick, Henry 1874 (1981). *The Methods of Ethics*, Indianapolis, IN: Hackett.
- Smith, Michael 1994. *The Moral Problem*, Oxford: Blackwell.
- Sober, Elliott 1992. Hedonism and Butler's Stone, *Ethics* 103/1: 97–103.
- Stocker, Michael 1976. The Schizophrenia of Modern Ethical Theories, *The Journal of Philosophy* 73/14: 453–66.