

## Butler's Stone

JOHN J. TILLEY  
INDIANA UNIVERSITY–PURDUE UNIVERSITY INDIANAPOLIS  
jtilley@iupui.edu

*Pacific Philosophical Quarterly* 99(4) (2018): 891–909.  
Please cite the published version.

**Abstract:** In the eleventh of Joseph Butler's *Fifteen Sermons* we find his best-known argument against psychological hedonism. Elliott Sober calls that argument *Butler's stone*, and famously challenges it. I consider whether Butler's stone has value. In doing so I examine, and reject, two possible responses to Sober's objection. This involves, in part, discussing Lord Kames's version of the stone argument, which has hitherto escaped scholarly attention. Finally, I explain an important value of Butler's stone, which I have not found previously discussed. Butler's stone blocks an inference, persuasive to many people, which purports to show that we intrinsically desire only pleasure.

### 1.

In one of the most oft-cited passages in the history of ethics, Joseph Butler argues thus:

That all particular appetites and passions are towards *external things themselves*, distinct from the *pleasure arising from them*, is manifested from hence; that there could not be this pleasure, were it not for that prior suitableness between the object and the passion: there could be no enjoyment or delight from one thing more than another, from eating food more than from swallowing a stone, if there were not an affection or appetite to one thing more than another. (*Sermons*, 11.6)<sup>1</sup>

Elliott Sober (1992; 2013) calls this argument *Butler's stone*, or the *stone argument*. Its target, the view it aims to refute, is generally thought to be the following, which I call *intrinsic psychological hedonism* (hereafter *intrinsic hedonism*).<sup>2</sup>

Human nature is such that we intrinsically desire – that is, desire for its own sake – nothing but our own pleasure (to include the absence of pain). We desire other things, of course, but only because we see them as a means to, or a source of, pleasure for ourselves.<sup>3</sup>

---

<sup>1</sup> I cite Butler (1913 [1729]) by sermon and paragraph number. By '*Sermons*, 11.6' I mean Sermon 11, paragraph 6. When citing Butler's preface to his sermons I use 'Pr' in place of a sermon number.

<sup>2</sup> The traditional term is 'psychological hedonism.' However, I use that term for a broader position: that either intrinsic hedonism or reductive hedonism (soon to be introduced) is true. The term 'reductive hedonism' is borrowed from Nilsson, 2013.

<sup>3</sup> Some hedonists will expand or supplement this definition with a point about the strength or intensity

A common view in philosophy is that intrinsic hedonism succumbs to Butler's stone. Sober, however, objects that if intrinsic hedonism is Butler's target, the stone argument is invalid.<sup>4</sup>

My question is whether Butler's stone has philosophical value, and if so, what that value is. Butler's stone would likely have value if we could surmount Sober's objection; thus, I examine two possible ways of doing that, each of which has proponents. The first way is to argue, contrary to Sober, that careful attention to Butler's premise shows that the stone argument succeeds. In discussing this way I examine Lord Kames's (Henry Home's) version of the stone argument, which, to my knowledge, has not been examined in print. The second way is to reinterpret the target of Butler's stone – specifically, to interpret it as *reductive psychological hedonism* (hereafter *reductive hedonism*), the view that every desire reduces to (or more accurately, is) a desire for pleasure.

The first of these two ways fails. The second does so as well, if success requires no question-begging against reductive hedonists.<sup>5</sup> But as I show in section 5, Butler's stone does something important, which I have not found previously discussed. It blocks an inference, persuasive to many people, that purports to establish intrinsic hedonism.<sup>6</sup>

## 2.

In this section I clarify Butler's stone and Sober's objection. We can initially reconstruct Butler's stone as follows:

1. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire for the external thing. Indeed, the pleasure could not occur otherwise.
2. Therefore, we sometimes desire external things themselves, 'distinct from the *pleasure arising from them*.'

---

of desires – namely, that such strength varies with the amount of pleasure we expect from the desired things or with the confidence with which we expect the pleasure. A historical example is Barr, 1749, pp. 67–69.

<sup>4</sup> This is Sober's central objection to the argument, though not his only one (see note 9). It is the objection I always mean by 'Sober's objection.' See Sober, 1992, pp. 99–101; Sober, 2013, pp. 154–55. For similar objections see Green, 1885, pp. 326–27; Penelhum, 1985, pp. 54–55; Brink, 1989, pp. 224–25; and Stewart, 1992, pp. 215, 220–21.

<sup>5</sup> A critical argument can have limited success, even if not full or robust success, even if it begs the question against proponents of the view it criticizes. For it might be capable of persuading a neutral audience. My point in the footnoted sentence is that Butler's stone has no robust success against reductive hedonism.

<sup>6</sup> To show this is not necessarily to identify the only way in which Butler's stone has value. An anonymous reviewer points out another possible way: Perhaps we can plausibly read Butler's stone not as a deductive argument but as an inference to the best explanation – specifically, as an attempt to show that intrinsic hedonism, unlike Butler's position, poorly explains the fact that we receive more pleasure from some things than from others. This is an intriguing and, to my knowledge, original idea that merits research. I will not pursue it here, however. I think that an adequate treatment of it is best reserved for a separate paper. Fortunately, I need not pursue it to establish or clarify the points of the present paper.

This argument calls for four comments. First, I use ‘external thing’ broadly, to extend to anything but our own pleasure. In doing so I follow Butler, whose usage is roughly the same.<sup>7</sup> Although this usage is odd (e.g., it makes the desire to *sleep* a desire for an external thing), it is legitimate in this sense: if Butler’s stone is sound on any narrower reading of ‘external thing,’ it is also sound given the meaning I adopt.

Second, like most interpreters of Butler, I assume that desires are among the things Butler means by ‘appetites,’ ‘passions,’ or ‘affections’ (Penelhum 1985, p. 52; Henson 1988, p. 33; Sober 1992, p. 99; Phillips 2000, p. 430 n. 28). At least, we do no damage to Butler’s argument if we replace his talk of those things with talk of desires.<sup>8</sup> I have done this in my reconstruction.

Third, my reconstruction is charitable in that, whereas some readers might prefer ‘for all external things’ in place of ‘for some external things’ (arguing that Butler’s words support this replacement), my reconstruction employs the latter phrase. This prevents Butler’s assertions from being overstatements.<sup>9</sup> Consider assertion 1, for instance. Surely we *sometimes* receive pleasure from an external thing – a cool breeze, a colorful sunset – without first desiring the thing. Hence, assertion 1 would be false if we replaced ‘some’ with ‘all.’

By contrast, if we retain the word ‘some,’ assertion 1 is plausible. For it does seem that for some external things that bring us pleasure, the pleasure occurs because, and not unless, we fulfill a desire for the external thing, meaning that we come to have (obtain, consume, use, do, experience) that thing.<sup>10</sup> In these cases, although it is natural to say (as I often will) that the *external thing* produces pleasure, it is perhaps more accurate to say that the *fulfillment of our desire* creates the pleasure. At any rate, I will not challenge statement 1.<sup>11</sup> Nor will I defend it; my main interest is not the statement itself, but its implications.

My fourth comment concerns my wording of Butler’s premise. On a more literal wording it would say merely that for some external things from which we receive pleasure, the pleasure could not occur without our fulfillment of an antecedent desire for the external things. To say this is not to make the further assertion that in those cases, the pleasure occurs *because* we fulfill the desire.<sup>12</sup>

---

<sup>7</sup> He uses ‘external thing’ to include anything but our own ‘happiness, enjoyment, satisfaction’ (*Sermons*, 11.5), and he often uses ‘enjoyment’ and ‘satisfaction’ interchangeably with ‘pleasure’ (see, e.g., *Sermons*, 11.6).

<sup>8</sup> At least, this is true as long as we use ‘some’ rather than ‘all’ as described in the next paragraph in the text. For pertinent remarks see Jackson, 1943, pp. 116–17; and Stewart, 1992, pp. 216–17.

<sup>9</sup> It thus armors them against the familiar charge that they overstate the case. See, e.g., Penelhum, 1985, p. 52; Stewart, 1992, p. 217; and Sober, 2013, p. 154.

<sup>10</sup> The parenthetical terms are included owing to the variety of ‘external things.’ For instance, our beneficent deeds are external things, but they are things that we do, not that we have.

<sup>11</sup> This is not to say that statement 1 is beyond question. See Stewart, 1992, pp. 217–18.

<sup>12</sup> As I point out in section 5, on a natural interpretation, the claim that *A* occurs because *B* occurs implies that *B*’s occurrence is what *explains* *A*’s occurrence. However, not just any event necessary for *A*’s occurrence (whether causally or logically) is an explanation of *A*’s occurrence.

But elsewhere Butler reveals that he accepts that assertion. For instance, in the preface to his *Sermons* he claims that we ‘find delight in such and such objects ... *because* we have particular affections towards them’ (*Sermons*, Pr.37, italics added). This claim appears not just anywhere in Butler’s preface, but in his preview or synopsis of the stone argument.<sup>13</sup> So I take him to intend a similar claim in the passage quoted earlier. Indeed, the idea Butler is famous for has two parts, though their difference is seldom highlighted: first, fulfillment of a desire for an external thing is often necessary for receiving pleasure from the thing; and second, in such cases the pleasure arises (if it does arise) because we fulfill that desire.<sup>14</sup> My wording of Butler’s premise reflects this idea.

Let us now turn to Sober’s objection, which we can present in two steps. First, Butler’s stone refutes intrinsic hedonism only if it refutes the hedonist’s view that whenever we want something besides pleasure itself, we do so only because we believe that were we to have it (obtain it, etc.), we would receive pleasure (which we intrinsically desire). We can express this as the view that an expectation of pleasure (as we might call the belief just mentioned) explanatorily precedes any desire for an external thing. In short, the desire occurs owing to the expectation, not vice versa.

So Butler’s stone refutes intrinsic hedonism only if its conclusion, statement 2, means this:

- 2.1. Therefore, we sometimes desire external things, where no expectation of pleasure explanatorily precedes our desire for them.

Second, if 2.1 is the meaning of 2, Butler’s stone is invalid. Statement 1 says that the pleasure we derive from an external thing explanatorily follows, rather than precedes, our desire for that thing. This does not rule out the claim that an expectation of pleasure explanatorily precedes our desire. *Pleasure* is one thing, an *expectation* of pleasure is another; so even if the first of those things follows the desire, the second may precede it. So 2.1 does not follow from 1.

Sober’s contention, then, is that Butler’s stone is invalid if it aims at intrinsic hedonism. But let us not leave the matter there. Let us see if we can

---

<sup>13</sup> See also Butler’s claim that ‘happiness or enjoyment . . . arises from such gratification alone,’ where ‘such gratification’ means ‘the gratification of certain affections, appetites, passions, with objects which are by nature adapted to them’ (*Sermons*, 11.16; see also *Sermons*, 13.5). Here Butler seems to mean that the happiness or enjoyment in question occurs because of the gratification, and further, that in no other way could the happiness or enjoyment occur. Butler makes this claim in his summary of the first fifteen paragraphs of Sermon 11. In fact, he makes it early in that summary, where he appears to be rehearsing part of the stone argument.

<sup>14</sup> Two representative remarks: ‘As Bishop Butler held, there must be a desire for sex or food or fame in the first place, if having sex or food or fame is to give us any pleasure. To suppose that we desire food for the pleasure it gives is to put the cart before the horse. Food gives pleasure because we desire it’ (Harrison 1993, p. 3). ‘Butler is surely right in claiming that there are many pleasures whose existence does depend on the satisfaction of a prior desire for something other than pleasure. . . . We are pleased [when we help the needy] because we want to help, rather than helping because we want to be pleased’ (McNaughton 2013, p. 395). See also Darwall, 1983, p. 5, final five lines.

rescue Butler's stone or at least find value in it. This is the task of the next three sections.

### 3.

One way to attempt a rescue of Butler's stone is to contend that, despite what Sober may *seem* to show, Butler's stone actually refutes intrinsic hedonism. Students sometimes sympathize with this way in their first reactions to Sober's objection. Their idea is that if we appreciate what premise 1 is saying, if we carefully reflect on its implications, we can see that it somehow rules out intrinsic hedonism.

I will take this idea seriously. I will do so by examining a version of the stone argument in which the idea is prominent, namely, that of the Scottish jurist and philosopher Henry Home, Lord Kames (1696–1782). This version of the argument, hereafter *Kames's stone*, is much the same as Butler's, with one possible improvement. It includes an intermediate conclusion which, Kames believes, reveals a valid connection between his premise and his final conclusion, the negation of intrinsic hedonism. Kames's argument is interesting for this reason, and for historical reasons as well. Despite Kames's prominence in the Scottish Enlightenment, his stone argument has escaped scholarly attention.<sup>15</sup>

Kames's stone is in his 'Foundation and Principles of Morality,' which appears in his *Essays on the Principles of Morality and Natural Religion*, first published in 1751. It occurs in a passage concerning what Kames calls the 'selfish system' of morality (2005 [1779], pp. 73–82), which he describes (not very clearly) as 'resolv[ing] all morality into self-love' (ibid., p. 73). His stone argument runs thus:

Those who hold self-love to be the only motive to action, maintain that the prospect of gratification is the only motive one can have for voluntary deeds of benevolence. I ask these gentlemen a plain question, When I have it in view to do a benevolent deed, whence arises the prospect of gratification? They must admit that it arises from my desire of performing the benevolent deed; for if I have no desire to perform, the performance will not gratify me, nor consequently will it afford me an antecedent prospect of gratification. It clearly follows, that as the desire to do a benevolent deed must always precede the prospect of gratification, the latter never can be the sole motive. The prospect of gratification may be an additional motive to act, but never can stand single. (Kames 2005 [1779], p. 80)

This argument is strikingly similar to Butler's. I believe that Kames sees himself as borrowing Butler's argument, though improving it in a certain way.<sup>16</sup>

---

<sup>15</sup> On Kames's prominence, Ross (1972) is useful.

<sup>16</sup> Kames's entire treatment of the 'selfish system' (2005 [1779], pp. 73–82) appears indebted to Butler's *Sermons*. And Kames certainly respected Butler and his work. See Kames, 2005 [1779], pp. 33–34. See also Ross, 1972, pp. 35–36.

The putative improvement has already been mentioned; its details will become clear when I reconstruct Kames's argument.

But first a word about Kames's target. Despite his claim that his opponents see the prospect of gratification as our 'only motive' for benevolent deeds, I see Kames's target as intrinsic hedonism, not reductive hedonism. I read him as challenging the view, not that we desire nothing but pleasure, but that we *intrinsically* desire nothing but pleasure. This reading is plausible partly for textual reasons – for instance, Kames's use of 'motives' in some other passages to mean fundamental motives rather than motives generally, and his reference to Claude-Adrien Helvétius (1715–71), an intrinsic hedonist, as a key proponent of the 'selfish system' (2005 [1779], pp. 74–75).<sup>17</sup> But more important, intrinsic hedonism is more plausible and prevalent than reductive hedonism. Kames's argument has more interest as a challenge to the first of those views than as a challenge to the second.

Let us now reconstruct Kames's stone, reading its target as intrinsic hedonism. To highlight its similarity to Butler's stone, let us replace Kames's talk of *doing benevolent deeds* with Butler's talk of *external things*. This revision is harmless, for given Butler's terminology, to want to do a benevolent deed is to want a species of external thing.

Our reconstruction is this:

1. For some external things from which we receive gratification, the gratification occurs because we fulfill an antecedent desire for the external thing. Indeed, the gratification could not occur otherwise.
2. Therefore, for some external things we are capable of desiring, namely, those to which 1 refers, any prospect we have of receiving gratification from them must await our desire for them. Until we desire them, the prospect of gratification cannot arise, given that the desire is crucial to receiving the gratification.
3. Therefore, contrary to intrinsic hedonism, we sometimes can desire external things, where no prospect of gratification from them explanatorily precedes our desire for them.

Premise 1 charitably captures Kames's claim that 'if I have no desire to perform [the benevolent deed], the performance will not gratify me.'<sup>18</sup> Statement 2 does

---

<sup>17</sup> For Kames's use of 'motives' to mean 'fundamental motives' see, e.g., Kames, 2005 [1779], p. 15, lines 8–13. Given the implications of the first two sentences in that passage, the third sentence is puzzling if we read it literally, as saying that as far as self-love is concerned, our *only* motives are pleasure and pain. That sentence is not puzzling at all, it agrees with the first two, if we understand 'motive' to mean 'fundamental motive.' For indications that Helvétius is an intrinsic hedonist, see Helvétius, 1759, pp. 161, 162, 163, 166, 168, 174, 224, 309.

<sup>18</sup> Kames is similar to Butler in thinking not just that the fulfillment of a desire is necessary for gratification, but also that the fulfillment of a desire is what produces or explains the gratification. For instance, in the paragraph directly preceding his stone argument (2005 [1779], pp. 79–80), Kames

the same for what Kames sees as an implication of premise 1: that ‘the prospect of gratification [from the benevolent deed] ... arises from my desire of performing the benevolent deed.’ As he also puts it, ‘the desire to do a benevolent deed must always precede the prospect of gratification.’ Statement 3 charitably captures Kames’s conclusion that ‘the prospect of gratification ... never can be the sole motive [for the benevolent deed],’ though it ‘may be an additional motive.’ I take the words ‘may be’ to indicate that sometimes, the prospect of gratification is not even an *additional* motive. This is to imply that sometimes, no prospect of gratification explanatorily precedes a benevolent deed, and hence no such prospect explanatorily precedes the desire that produces the deed.

It is mainly by including statement 2, an intermediate conclusion, that Kames’s argument differs from Butler’s. Statement 2 asserts a connection between two claims. The first, stated in 1, is that *receiving gratification* from (some) external things depends on having a prior desire for them. The second, stated in 2, is that *having a prospect* of receiving the gratification depends on having that desire. Through its use of ‘therefore,’ 2 asserts that the first of those claims entails the second. The rough idea here is that if receiving gratification depends on having a particular desire, then until we have that desire we have no prospect of receiving the gratification. But if the second claim is true – that is, if 2 stands up – then it seems that 3 is true. For it seems that if, as 2 says, the prospect of gratification must await a desire for the external things, then, as 3 says, the prospect does not explanatorily precede the desire. When a prospect of gratification explanatorily precedes a desire, it also temporally precedes and helps produce the desire.

Before we evaluate Kames’s argument some rewording is desirable. This stems from Kames’s terms ‘gratification’ and ‘prospect,’ which have multiple uses.

By ‘gratification’ Kames apparently means a species of pleasure that attends the fulfillment of any desire (2005 [1779], p. 79). However, this meaning is too narrow for his purpose. To refute intrinsic hedonism, he must use ‘gratification’ to mean what ‘pleasure’ means in the intrinsic hedonist’s thesis. He must use it to mean, not a particular species of pleasure, but simply pleasure of one or more kinds. So let us substitute ‘pleasure’ for Kames’s term ‘gratification.’

Next, ‘prospect of gratification’ could mean any of several things – for instance, the *likelihood* of gratification or, instead, the *expectation* of gratification. Given that intrinsic hedonism claims that desires for external things spring from *expectations* of pleasure, let us remove ambiguity by substituting ‘expectation’ for ‘prospect.’

Kames’s argument is now this:

---

‘prepare[s] the reader’ for his stone argument by ‘premis[ing] the following *data*, First, that the accomplishment of desire produces a pleasant feeling, termed gratification of the passion. Next, that where there is no desire, there is no gratification.’

1. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire for the external thing. Indeed, the pleasure could not occur otherwise.
2. Therefore, for some external things we are capable of desiring, namely, those to which 1 refers, any expectation we have of receiving pleasure from them must await our desire for them. Until we desire them, the expectation of pleasure cannot arise, given that the desire is crucial to receiving the pleasure.
3. Therefore, contrary to intrinsic hedonism, we sometimes can desire external things, where no expectation of pleasure from them explanatorily precedes our desire for them.

This argument is unsound. Statement 1 does not entail 2. From the claim that our pleasure from  $x$  occurs because, and not unless, we fulfill a desire for  $x$ , it does not follow that until we have that desire we can expect no pleasure from  $x$ . We might expect pleasure from  $x$  – that is, we might think that by having (obtaining, etc.)  $x$  we would receive pleasure – owing to many things that enter our minds prior to, and apart from, any desire we develop for  $x$ . (This is consistent with the claim that our expectation quickly sparks a desire for  $x$ .)

Nor will it do to delete the word ‘therefore’ from 2 and let 2 stand on its own. For as just indicated, for any external things we might want, we might expect pleasure from them owing to many things besides a desire – for instance, to the testimony of someone we trust or to a resemblance to known sources of pleasure.

Although unsuccessful as it stands, Kames’s stone is suggestive. So let us try to improve it, preserving the idea that to examine premise 1, to uncover exactly what it implies, is somehow to find a disproof of intrinsic hedonism.

To start, consider the following arguable view. If, as premise 1 says, we receive no pleasure from certain things unless we first desire them, then until we desire them we have no *strong evidence*, meaning evidence of exceptional force, that they will bring us pleasure. We might expect pleasure from them, and even have evidence for doing so. However, until we desire them – that is, until we have the psychological state that enables them to produce pleasure – we have no *strong evidence* for our expectation.

On the basis of this view, let us revise premise 2 as shown below. It now concerns the conditions, not for *having* an expectation of pleasure, but for having a *strongly evidenced* expectation of pleasure. Also shown are further revisions to Kames’s argument, which link the new version of 2 with Kames’s conclusion.

1. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire for the external thing. Indeed, the pleasure could not occur otherwise.



2. Therefore, for some external things we are capable of desiring, namely, those to which 1 refers, any strongly evidenced expectation we have of receiving pleasure from them must await our desire for them.
3. For some of the external things to which 1 refers, any expectation we have of receiving pleasure from them is strongly evidenced at the time it arises. (It would be an odd coincidence if this were not true of *any* of the external things to which 1 refers.)
4. Therefore (from 2 and 3), for some external things we are capable of desiring, any expectation we have of receiving pleasure from them does not temporally precede our desire for them. The desire is there when the expectation arises, as a condition for the strong evidence the expectation enjoys.
5. If the expectation to which 4 refers does not temporally precede the desire to which 4 refers, then it does not explanatorily precede that desire. An expectation explanatorily precedes a desire by causing (or partly causing) the desire. But that involves practical reasoning (conscious or unconscious), which consumes at least an instant of time.
6. Therefore (from 4 and 5), contrary to intrinsic hedonism, we sometimes can desire external things, where no expectation of receiving pleasure from them explanatorily precedes our desire for them.

Unfortunately, this argument has at least one fatal flaw. The flaw resides in the ‘arguable view’ that spawned our revision of premise 2. Contrary to that view, the ‘strong evidence’ in question need not await a desire. Hence the revised version of 2 does not follow from 1.

To see this, assume that 1 is true, and let  $x$  be any of the external things to which 1 and 2 refer. Of course, even though 1 is true of  $x$ , we may not know that. Suppose that what we know about  $x$  is the following: first,  $x$  shares many properties with  $y$ ; second, each of the many times we have obtained something with even a few of those properties, we have received pleasure; and third,  $x$ ’s other properties, the ones that  $y$  does not share, have never frustrated our pleasure when we have met with them. On this basis we expect to receive pleasure from  $x$ . More fully, given what we know about  $x$ , we believe that were we to obtain (experience, consume, etc.)  $x$ , we would receive pleasure from it. Suppose further that our belief is unopposed by evidence that  $x$  will not bring us pleasure. For example, the fact, if it is a fact, that as yet we have no desire for  $x$  is no evidence that we will receive no pleasure from  $x$ . For it does not preclude our soon having such a desire. In fact, that we will do so is very likely, given our belief that  $x$  offers us pleasure.

In this situation, which is possible, our expectation of pleasure from  $x$  is strongly evidenced the moment it arises. For there is strong evidence in support

of it, and no evidence against it. But as just indicated, we may not desire  $x$ . Nothing in our situation ensures that we do. Very likely, our expectation of pleasure will quickly produce a desire for  $x$ , but we may lack that desire at the instant of our expectation.

Hence the example shows that 2 could be false even if 1 is true. It shows, first, that for any of the external things to which 1 refers, it is possible that, while lacking a desire for those things, we have a strongly evidenced expectation of pleasure from them. Second, this is possible even if, as 1 says, the pleasure we ultimately receive from those things occurs because, and not unless, we fulfill a prior desire for them. So the step from 1 to 2 is invalid.

Let us consider two possible replies. The first is that even though 2 does not follow from 1, premise 2 is true. The idea here is that the possible situation I described, which shows that 2 *could* be false, never occurs in the actual world. Nor does any other situation in which 2 is false. In this world, prior to desiring an external thing of the kind to which 1 refers, any grounds we have for expecting pleasure from it are never of the type described in my example – and more generally, never of a type that counts as strong evidence for our expectation.

But why believe this? Given that 2 does not follow from 1, what evidence do we have for 2? Premise 2 is not *obviously* true; nor does the argument for 6 contain support for it.

The second reply is that although the things mentioned in my example, the various facts about  $x$  and  $y$ , indeed provide strong evidence for the relevant expectation, that evidence is not as strong as the evidence furnished by a desire for  $x$  – or furnished by adding such a desire to the other facts about  $x$  and  $y$ .

I find this reply doubtful, but suppose it is true. How does it challenge my argument about 1 and 2? It does so by prompting a revision of 2 – a revision in which the phrase ‘strongly evidenced expectation we have of receiving pleasure from them’ is replaced with ‘expectation we have of receiving pleasure from them that enjoys evidence of strength  $S$ ,’ where  $S$  is the remarkable strength of evidence which, by hypothesis, stems not from the facts I listed about  $x$  and  $y$ , but from a desire (or the addition of a desire) for the external things. But now, to preserve the validity of the inference to 4, we must make 3 say that for some of the external things to which 1 refers, any expectation we have of receiving pleasure from them rests not merely on strong evidence, but on evidence of strength  $S$ , at the time it arises. With this revision, 3 is not very plausible. For instance, if it were false, would that really be a rare coincidence? I see no reason to think so.

It remains to point out that we cannot repair premise 2 by tinkering with the description of the external things it concerns – for instance, by replacing ‘namely, those [external things] to which 1 refers’ with either ‘namely, at least a few of those to which 1 refers’ or ‘namely, at least a few of them [the things we are capable of desiring], which may or may not be among the things to which 1

refers.’ The resulting versions of 2 do not follow from 1. Nor do they rest on other compelling evidence.

At this point, we have carried our discussion far enough to conclude, reasonably, that Kames’s stone does not refute intrinsic hedonism. It does not overcome Sober’s objection. Nor, more generally, does the strategy it embodies, that of unpacking the implications of premise 1. So let us proceed to a different attempt to find value in the stone argument.

#### 4.

Some philosophers believe that Butler’s stone aims not at intrinsic hedonism but at reductive hedonism, the view that no matter what its apparent object, every desire is actually a desire for pleasure.<sup>19</sup> If this belief is true, Butler’s stone sidesteps Sober’s objection.

Reductive hedonism differs sharply from intrinsic hedonism. Intrinsic hedonists do not claim that every desire is for pleasure. They think that our desire for, say, a new watch is for exactly that – a new watch. The fact that it subserves a fundamental desire for pleasure does not mean that its immediate object, what it is directly for, is pleasure or something that includes pleasure. Reductive hedonists, by contrast, claim that our desire for a new watch – or what we call or see as our desire for a new watch – aims directly at pleasure. In that sense, our desire ‘reduces’ to a desire for pleasure.

H. M. Zellner (1999, pp. 194–96) gives two reasons to think that reductive hedonism is Butler’s target. The first, endorsed by Peter Nilsson (2013, p. 548) (who adds further textual support for it), is that Butler’s own words sometimes suggest that he aims at reductive hedonism. In a portion of Butler’s preface to his *Sermons*, where Butler gives a preview of the sermon containing his stone argument, Butler speaks of ‘a strange affectation in many people of explaining away all particular affections, and representing the whole of life as nothing but one continued exercise of self-love’ (*Sermons*, Pr.35). Zellner thinks that arguably, this ‘strange affectation’ involves a tendency to see all ‘particular passions’ – Butler’s term for desires for things besides pleasure – as really desires for pleasure.

Why would Butler want to refute reductive hedonism, a thesis which, frankly, has little intrinsic plausibility? This brings us to Zellner’s second reason for thinking that Butler targets reductive hedonism. Although the authors

---

<sup>19</sup> Pristine examples of such philosophers include Zellner (1999, pp. 193–97) and Nilsson (2013, pp. 548–49). See also Albert Lefevre (1900, p. 167), who thinks that in the stone argument ‘Butler attempts to refute the hedonistic contention, that pleasure is the only object of desire.’ Stephen Darwall (1983, p. 5) may have Butler’s stone in mind when he says that ‘one of the most enduring of Butler’s contributions’ is ‘his critique of psychological theories that seek to reduce all motivation to some dominant self-regarding desire, such as the general desire for one’s own pleasure or happiness in life.’ Nilsson (2013, pp. 546, 548) maintains that C. D. Broad sees Butler as attacking reductive hedonism. He does so because Broad (1930, p. 63) holds that in the stone argument, Butler’s ‘task is to show that [“particular impulses”] cannot be reduced to self-love.’

Butler opposes – Hobbes and La Rochefoucauld, for instance – may consciously hold views other than reductive hedonism, their choice of words sometimes suggests reductive hedonism.<sup>20</sup> Hence, given the influence of those authors, it would not be odd for Butler to take a critical interest in reductive hedonism.

Zellner could have added that Butler’s eighteenth-century allies against egoism – Francis Hutcheson and David Hume, for instance – sometimes use words which suggest that Hobbesians hold reductive hedonism (Hutcheson 2002 [1728], p. 134; Hume 1997 [1772], pp. 164–65). Given the instances of that suggestion in Butler’s time, reductive hedonism may have had some currency then, particularly among advocates of Hobbesian psychology. So again, it would not be odd for Butler to take a critical interest in reductive hedonism. Also, his criticism of it would have value if it refuted its target.

Let us assume, then, at least for the sake of argument, that Butler’s stone aims at reductive hedonism. Some philosophers think that it refutes that thesis.<sup>21</sup> I contend that it does not, at least if a condition for truly refuting that thesis is that we beg no questions against reductive hedonists.<sup>22</sup>

Let me explain, first by noting, as Zellner does, that it is difficult, perhaps impossible, to find clear, published endorsements of reductive hedonism. Even so, we can imagine people who hold that thesis. Do they mean by it that we have just one desire, the desire for pleasure *simpliciter*?<sup>23</sup> Or do they mean that we have multiple desires, each for pleasure of some kind, where pleasure *simpliciter* is the broadest of those kinds? For my purposes, it does not matter which of these views they mean; so let us take them to mean the second one. They believe that we indeed desire pleasure *simpliciter*, but we also desire the pleasure *we connect with a hot bath*, the pleasure *we connect with a cold beer*, and so on. (The meaning of ‘connect with,’ they will say, varies with context. It often means ‘expect from,’ but it can sometimes mean ‘recall from,’ ‘associate with,’ etc.) These desires – our desire for pleasure *simpliciter* and our desires for pleasures of specific kinds – are the only desires we have. This is the reductive hedonists’ thesis.

We might be inclined to dismiss this thesis. But if we do not, if we instead take it seriously enough to challenge it, we must treat it fairly. Let us do that, and thus assume that reductive hedonists have thought of the following question and have addressed it consistently with their theory: Why, if we desire only

---

<sup>20</sup> The passages Zellner quotes appear in Hobbes (1962, p. 221) and La Rochefoucauld (2007 [1665], p. 147).

<sup>21</sup> Zellner (1999, pp. 193, 200) and Nilsson (2013, pp. 546, 549) think so, and I believe that Lefevre (1900, pp. 167–68), Hume (1997 [1772], pp. 165–69), and possibly Broad (1930, p. 63) think so as well. And of course we should add Butler to this list, if Zellner’s reading of him is accurate.

<sup>22</sup> See note 5.

<sup>23</sup> This, apparently, is how Zellner (1999, p. 200) and, possibly, Nilsson (2013, pp. 548–49) interpret them.

pleasure, do we commonly say such things as that we want a new watch, or that we have a desire for a new watch, leaving pleasure unmentioned?

The reductive hedonists' answer, no doubt, is that ordinary speech is often inaccurate. To speak accurately, we should say that we want *the pleasure we connect with owning a new watch*. In fact, because desires are propositional attitudes, to be fully accurate we should say that we want *that we receive the pleasure we connect with owning a new watch*.<sup>24</sup> Reductive hedonists think that unless our statement is equivalent to this latter statement, it is false.

Of course, this view of ordinary speech about desires is not necessarily plausible. My point about it is simply that reductive hedonists hold it, if we interpret them fairly. They see it as a corollary, or a crucial auxiliary, of their thesis. For it would be obtuse to hold reductive hedonism on the one hand, but not hold, on the other, a view of the above kind about ordinary speech.

Let me now argue, in four steps, that unless Butler's stone begs questions against reductive hedonism, it fails to refute that thesis. First, Butler's stone refutes reductive hedonism only if we read Butler's stone as shown below. Here we have replaced 2.1 (mentioned in section 2) with 2.2, which negates reductive hedonism.

1. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire for the external thing. Indeed, the pleasure could not occur otherwise.
- 2.2. Therefore, we sometimes desire external things, where the immediate object of our desire is simply the external thing; it includes no pleasure. In other words, the propositional content of our desire is simply that *we have (obtain, consume, etc.) the external thing*.

Let me pause for a clarifying point: I assume that when we say 'We want (or desire, or have as the content of our desire) . . .' and add a 'that' clause – for instance, 'that we have *x*' – we are speaking in no elliptical or colloquial way. At least, this is how I intend such constructions. So 2.2 is about the full and literal propositional content of the desire in question. The same goes for 1.1 and 1.2, soon to be introduced.

To come to the second step of my argument, statement 1 contains the expression 'desire for the external thing.' To make this expression non-colloquial, most of us will read it to say 'desire that we have (obtain, etc.) the external thing.' Reductive hedonists, however, will read it to say 'desire that we receive the pleasure we connect with having the external thing.' We thus have two non-colloquial readings of statement 1:

---

<sup>24</sup> Suppose we take the minority view that some desires are not propositional attitudes. This would require only a minor rewording of some of my later statements.

- 1.1. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire that *we have the external thing*. Indeed, . . .
- 1.2. For some external things from which we receive pleasure, the pleasure occurs because we fulfill an antecedent desire that *we receive the pleasure we connect with having the external thing*. Indeed, . . .

Third, 1.1 is glaringly unacceptable to reductive hedonists. They will reject it on sight, without even asking what conclusions it may aim to support. It takes the literal meaning of the phrase ‘desire for the external thing,’ which appears in 1, to be ‘desire that we have the external thing,’ which makes no mention of pleasure. It thus conspicuously implies that some desires do not have pleasure as part of their objects. Thus, we must read 1 as 1.2 to avoid begging the question against reductive hedonism.

Fourth, if we read 1 as 1.2, it fails to entail 2.2. The only desire mentioned in 1.2 has pleasure as part of its immediate object. Its propositional content refers to a form of pleasure. But 2.2 says that some desires have immediate objects that exclude pleasure. So 1.2 does not entail 2.2.

In sum, insofar as Butler’s stone begs no questions against reductive hedonism, it consists of 1.2 and 2.2. But 1.2 fails to establish 2.2.

Allow me to tie up some loose ends before closing this section. First, earlier I mentioned two possible meanings of the claim that every desire is for pleasure. I said that for my purposes, it does not matter which of those meanings reductive hedonists intend. My point was that given either meaning, insofar as Butler’s stone begs no questions against reductive hedonism, it fails against that thesis. Let me briefly explain this concerning the first meaning: that we have just one desire, the desire for pleasure *simpliciter*.

This thesis, like the other one, comes with an account of ordinary speech about desires. It is not hard to see what that account is, what forms of it are viable. Those who hold it, and with it the hedonism we are now considering, will reject Butler’s premise unless it means roughly this:

- 1.3. For some external things from which we receive pleasure, the pleasure occurs because, by achieving the external thing, we fulfill an antecedent desire that *we receive pleasure, meaning pleasure simpliciter*. Indeed, . . .

Unless we read Butler’s premise as roughly 1.3, we beg the question against the hedonists we are envisioning. But 1.3 fails to entail 2.2, the version of 2 that would rule out our hedonists’ thesis.

To tie up a second loose end, let me say that the results of this section do not depend on my particular reconstruction of Butler’s stone. They result from Butler’s own words – for instance, from his phrase ‘affection or appetite to one thing more than another.’ This phrase, reductive hedonists hold, means the

same as the less colloquial phrase, ‘affection or appetite to the pleasure we connect with one thing more than to the pleasure we connect with another.’ But if we read the phrase that way, thereby preventing Butler’s stone from begging questions, Butler’s stone does not refute reductive hedonism.

As a final loose end, let me speculate that most people find 1.1 more plausible than 1.2 or 1.3. To the extent that they accept premise 1, they understand it as 1.1. So if, on some occasion, they were tempted by reductive hedonism, Butler’s stone might prevent them from accepting it. For they would read 1 as 1.1, which, by entailing 2.2, rules out reductive hedonism.

Hence, Butler’s stone has some success against reductive hedonism. It has persuasive force for some audiences.<sup>25</sup> But as this section shows, it does not refute reductive hedonism in a robust way – a way that avoids any question-begging. Not surprisingly, the same goes for Kames’s stone, which contains the phrase ‘desire for the external thing’ (literally: ‘desire of performing the benevolent deed’). If we read that phrase so that it begs no questions against reductive hedonism, Kames’s argument does not refute that thesis.

## 5.

The stone argument has no success against intrinsic hedonism and only limited success against reductive hedonism. Even so, it just *seems* to do something important, something that goes beyond the limited success just mentioned. And indeed it does. Its premise, if true, blocks a nondeductive inference that many find tempting, an inference that purports to establish intrinsic hedonism. Perhaps Butler has no aim of doing this; even so, his premise, if true, refutes the inference.

The inference in question is not the one ethics texts commonly reject, sometimes basing their rejection on Butler’s premise.<sup>26</sup> Nor is my treatment of it the same as theirs.

The inference in those texts, formulated to cover different versions of it, is this:

Whenever we complete an action or satisfy a desire, we invariably (or usually) receive pleasure. Therefore, the object (or ultimate object) of our act or desire (or of any of our acts or desires) was that particular pleasure (or pleasure, period).

---

<sup>25</sup> Perhaps this modest point is what Zellner, Nilsson, and some of the others in note 21 intend. I say ‘perhaps’ because they do not explicitly say so.

<sup>26</sup> See, e.g., Frankena, 1973, pp. 21–22; Harrison, 1993, p. 3; Luper, 2002, p. 71; Hinman, 2013, p. 98; and Pojman and Fieser, 2017, p. 80. Incidentally, it is not merely ethics texts that discuss a form of the inference or use Butler’s premise to reject it. See, e.g., Henson, 1988, pp. 38, 50.

Owing partly to the parenthetical remarks, this inference has many variations. Many of them are grossly implausible; it is the minimally plausible ones that receive criticism in ethics texts.

All of them differ from the inference I will discuss. First, they are deductive. This makes them easy to refute, given their content. Often, the texts that discuss them simply point out that they lack validity, that a claim about the *results* of acting or fulfilling a desire entails nothing about the direct or ultimate *object* of the act or desire. Also, most of them conclude with something slightly different from intrinsic hedonism – for instance, with the claim that the ultimate object of any act is the specific pleasure the act produces.

The inference I will discuss is this:

Normally, we receive pleasure when we fulfill a desire for an external thing. So most likely, we desire only those external things from which we expect pleasure, pleasure being the only thing we desire intrinsically.

This inference is more plausible than the invalid inference mentioned above. Also, it is more significant in what it concludes than most forms of that inference. Its premise (which I grant, at least *arguendo*) I call the *correlation claim*, for it asserts a correlation between fulfilling desires for external things and receiving pleasure. Its conclusion is intrinsic hedonism.

It is in conversations about intrinsic hedonism that I often encounter the above inference. Those who make it apparently think that what best explains the correlation claim – that is, what best explains the fact it asserts – is either intrinsic hedonism or something that entails that view. They say things like this: ‘Clearly, we normally receive pleasure from fulfilling our desires for things. Isn’t that because pleasure is what we basically want, making us generally desire only pleasurable things?’ They seldom flesh out this idea, so let us do it for them.

At the bottom of their idea is the following assumption, which, if true, would both explain the correlation claim and support (by including) intrinsic hedonism.

- (A) We desire only those external things from which we expect pleasure, pleasure being the only thing we desire intrinsically. Also, we have a knack for knowing, prior to desiring external things, which ones hold out pleasure and which ones do not. So our expectation of pleasure from an external thing is usually true.

Let me digress to note that those who accept intrinsic hedonism on the basis of (A) are clearly *intrinsic* hedonists, not reductive ones. So throughout this section, ‘desire for the external thing,’ variations of which appear in (A), has its usual meaning: ‘desire that we have (obtain, etc.) the external thing.’ Likewise, ‘Butler’s premise’ denotes Butler’s premise on an ordinary reading.



Charitably interpreted, those who infer intrinsic hedonism from the correlation claim reason as follows: The correlation claim is true, and the best way to explain it is to accept (A). But (A) implies intrinsic hedonism. Thus, intrinsic hedonism is well supported.

The key premise here is that thesis (A) is what best explains the correlation claim. This premise is false if Butler's premise is true. This is the case whether (A) itself is true or not.

Butler's premise essentially says (in part) that some of the pleasure the correlation claim concerns – that is, some of the pleasure that comes from fulfilling desires for external things – occurs *because* we fulfill the relevant desires. On a natural and charitable reading, this implies that what *explains* the pleasure is the fulfillment of those desires, not the ultimate motives from which the desires arose. True, those ultimate motives are among the causes, broadly speaking, of the pleasure (so too is the Big Bang). But they are not the *explanatory* causes, the things to be cited if asked 'Why did the pleasure occur?' So if Butler's premise is true, some of the pleasure the correlation claim concerns has an explanation which, although it refers to certain desires, leaves unmentioned the motives at their source. It says nothing about whether those motives are self-interested. And that means that the best explanation of the correlation claim will do the same. Unlike (A), it will say nothing about the motives just mentioned.

In sum, if Butler's premise is true, thesis (A) is not the best explanation of the correlation claim, contrary to the key premise in the inference we are examining. So Butler's premise, if true, has considerable value. It undermines an inference that lures some people to accept intrinsic hedonism.

A question remains: If Butler's premise is true, meaning that (A) is not the best explanation of the correlation claim, then what *is* the best explanation of that claim? It is something like this:

Sometimes, we receive pleasure from fulfilling our desires for external things because we antecedently want pleasure and believe, accurately, that by obtaining the external things we could receive pleasure. In the remaining cases (or in most of them) we receive the pleasure because we antecedently desire the external things. It is *because* we desire them that upon obtaining them we receive pleasure.

Predictably, this explanation does not entail intrinsic hedonism. For all we know, the desires to which its second sentence refers originate in no desire for pleasure.

Let me close this section with an earlier point: that this paper neither challenges nor defends Butler's premise, which says that sometimes, we receive pleasure from a thing because, and not unless, we fulfill a desire for the thing. This premise is not beyond question, but I lack space to examine it adequately. Certainly it is plausible, and I find that many people accept it, including a great

many philosophers. They may have not thought of it before encountering Butler's argument, but they accept it upon reading his argument. Hence my results are important. To accept Butler's premise is tacitly to reject the nondeductive inference discussed in this section.

## 6.

To conclude: It has long seemed to philosophers that Butler's stone scores a point, an important one, against psychological hedonism – indeed, against the most plausible form of that view, intrinsic hedonism. And in fact it does, even if that point is often misidentified. Butler's stone scores a point because it employs a plausible premise that many people readily accept, and which, if true, refutes an inference that many find tempting. I mean the inference from the correlation claim to intrinsic hedonism. As long as that inference is alluring to some minds, Butler's stone will have value in debates over intrinsic hedonism.<sup>27</sup>

## REFERENCES

- Barr, J. (1749). *A Summary of Natural Religion*, 2nd ed. (London: R. Dodsley).
- Brink, D. O. (1989). *Moral Realism and The Foundations of Ethics* (Cambridge: Cambridge University Press).
- Broad, C. D. (1930). *Five Types of Ethical Theory* (New York: Harcourt, Brace and Company).
- Butler, J. (1913 [1729]). *Fifteen Sermons Preached at the Rolls Chapel*, 2nd ed., ed. J. H. Bernard (London: Macmillan).
- Darwall, S. L. (1983). 'Introduction,' in Joseph Butler, *Five Sermons*, ed. Stephen L. Darwall (Indianapolis: Hackett), pp. 1–8.
- Frankena, W. K. (1973). *Ethics*, 2nd ed. (Englewood Cliffs, NJ: Prentice-Hall).
- Green, T. H. (1885). 'Introduction to Hume's *Treatise of Human Nature*,' in *Works of Thomas Hill Green*, 3 vols., ed. R. L. Nettleship (London: Longmans, Green, and Co.), vol. 1, pp. 1–371.
- Harrison, J. (1993). 'Introduction,' in *Challenges to Morality*, ed. Jonathan Harrison (New York: Macmillan), pp. 1–40.
- Helvétius, C. (1759). *De L'Esprit: or, Essays on the Mind, and its Several Faculties* (London: Dodsley, et al).
- Henson, R. G. (1988). 'Butler on Selfishness and Self-Love,' *Philosophy and Phenomenological Research* 49(1), pp. 31–57.
- Hinman, L. M. (2013). *Ethics: A Pluralistic Approach*, 5th ed. (Boston: Wadsworth).
- Hobbes, T. (1962). *Body, Man, and Citizen: Selections from Thomas Hobbes*, ed. Richard S. Peters (New York: Collier).
- Hume, D. (1997 [1772]). *An Enquiry concerning the Principles of Morals*, ed. Tom L. Beauchamp (Oxford: Oxford University Press).

---

<sup>27</sup> I'm grateful to Chad Carmichael for letting me pick his brain on an issue or two. I'm especially grateful to the anonymous referees, whose comments led to significant improvements.

- Hutcheson, F. (2002 [1728]). *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. Aaron Garrett (Indianapolis: Liberty Fund).
- Jackson, R. (1943). 'Bishop Butler's Refutation of Psychological Hedonism,' *Philosophy* 18(70), pp. 114–39.
- Kames, H. H., Lord (2005 [1779]). *Essays on the Principles of Morality and Natural Religion*, 3rd ed., ed. Mary Catherine Moran (Indianapolis: Liberty Fund).
- La Rochefoucauld, F., duc de (2007 [1665]). *Collected Maxims and Other Reflections*, trans. E. H. Blackmore, A. M. Blackmore, and Francine Giguère (New York: Oxford University Press).
- Lefevre, A. (1900). 'Self-Love and Benevolence in Butler's Ethical System,' *The Philosophical Review* 9(2), pp. 167–87.
- Luper, S. (2002). *A Guide to Ethics* (Boston: McGraw-Hill).
- McNaughton, D. (2013). 'Butler's Ethics,' in *The Oxford Handbook of the History of Ethics*, ed. Roger Crisp (Oxford: Oxford University Press), pp. 377–98.
- Nilsson, P. (2013). 'Butler's Stone and Ultimate Psychological Hedonism,' *Philosophia* 41(2), pp. 545–53.
- Penelhum, T. (1985). *Butler* (London: Routledge and Kegan Paul).
- Phillips, D. (2000). 'Butler and the Nature of Self-Interest,' *Philosophy and Phenomenological Research* 60(2), pp. 421–38.
- Pojman, L. P., and Fieser, J. (2017). *Ethics: Discovering Right and Wrong*, 8th ed. (Boston: Cengage).
- Ross, I. S. (1972). *Lord Kames and the Scotland of His Day* (Oxford: Clarendon).
- Sober, E. (1992). 'Hedonism and Butler's Stone,' *Ethics* 103(1), pp. 97–103.
- Sober, E. (2013). 'Psychological Egoism,' in *The Blackwell Guide to Ethical Theory*, 2nd ed., ed. Hugh LaFollette and Ingmar Persson (Oxford: Blackwell), pp. 148–68.
- Stewart, R. M. (1992). 'Butler's Argument against Psychological Hedonism,' *Canadian Journal of Philosophy* 22(2), pp. 211–21.
- Zellner, H. M. (1999). 'Passing Butler's Stone,' *History of Philosophy Quarterly* 16(2), pp. 193–202.