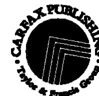
PHILOSOPHICAL PSYCHOLOGY, VOL. 13, NO. 1, 2000



The choice between current and retrospective evaluations of pain

STEPHANIE BEARDMAN

ABSTRACT *Daniel Kahneman and his colleagues have made an interesting discovery about people's preferences. In several experiments, subjects underwent two separate ordeals of pain, identical except that one ended with an added amount of diminishing pain. When asked to evaluate these episodes after experiencing both, subjects generally preferred the longer episode—even though it had a greater objective quantity of pain. These data raise an ethical question about whether to respect such preferences when acting on another's behalf. John Broome thinks that it is wrong to add extra pain in order to satisfy a person's preference for a better ending. His explanation for this intuition is that pain is intrinsically bad. I argue against this explanation, and raise several doubts about the moral intuition Broome endorses. In doing so, I offer alternate interpretations of Kahneman's data, and show that these each yield different values which are relevant to the ethical question.*

There are two main ways of evaluating an episode as painful: the first is as the experience is occurring, the second is after it has occurred [1]. One might think these two evaluations—we can call them the direct and the retrospective—are intimately related, in that the retrospective evaluation simply reflects, or should reflect, the sum of overall pain experienced directly. There is, however, evidence that the two evaluations come apart, from a series of studies conducted by Daniel Kahneman and his colleagues. These data have interesting ethical implications. After explaining some of the data and identifying the ethical question to which it gives rise, I will investigate possible avenues of addressing the ethical question. I will argue, against the predominating view expressed by Kahneman and others, for the recognition of an intrinsic value connected with retrospective evaluations.

I will focus the discussion by responding to John Broome's "More pain or less?" (1996). In that paper, Broome expresses a strong ethical intuition regarding the moral relevance of a subject's retrospective evaluations, and I want to question its basis. Broome's view implicitly relies upon an interpretation of the data that is expressed by Kahneman himself in his recent "Objective happiness" (1999). My main concern will be to establish an alternative interpretation of the data that neither

Editor's note: This paper was the winner of the 1999 William James Prize of the Society for Philosophy and Psychology, awarded for the best paper by a graduate student presented at the Society's annual meeting.

ISSN 0951-5089 (print) ISSN 1465-394X (online)/00/010097-14 © 2000 Taylor & Francis Ltd

Kahneman nor Broome seem to notice. The larger point is that various possible interpretations of the data carry substantive assumptions about value, assumptions that are debatable even when the focus is restricted to hedonic value.

One experiment in particular demonstrates that direct and retrospective evaluations of a painful experience come apart in such a way that subjects are moved to choose more pain over less. This is the "cold pressor" experiment (Kahneman *et al.*, 1993). Subjects were told they would undergo three episodes of immersion of one hand in painfully cold water. During the episodes, the subjects recorded their current levels of discomfort. After each trial, they were asked to evaluate the episode as a whole. They experienced both a "short" and a "long" session of pain. The short episode consisted of 60 seconds of immersion of one hand in 14° C water. The long episode consisted of 60 seconds of immersion in 14° C water, plus 30 seconds during which the water was slowly warmed up 1° C. (The order in which the episodes were administered, and which of each subject's hands was immersed, made no difference to the results.) After experiencing the two episodes, the subjects were permitted to choose which they wanted to repeat for their upcoming third session.

The two episodes were identical except for an extra period of immersion in painfully cold (though marginally warmer) water added to the long episode. It would seem that the rational choice would be the shorter episode. After all, the only relevant criterion for choice in such a situation seems to be narrowly hedonic: maximize pleasure and minimize pain. As Kahneman points out, there was no indication that subjects were guided by any other motive. So it may seem surprising that 81% of those who experienced diminishing pain in the longer episode later chose to repeat it rather than the shorter one.

Call the quantity identified by retrospective evaluations of an episode its "global utility," and that identified by direct evaluations the "local utility" of the moments which make up the episode. In the cold pressor experiment (and in the rest of the experiments in the series), people's assessments of global utility can be predicted by a formula which Kahneman *et al.* call the "Peak and End rule," whereby judgments of global utility are equal to the average of the utility directly assigned to just two moments of the experience: its worst moment (peak) and the final moments (end). Though there is some evidence that other qualities of the experience are accurately remembered, they carry little weight in assessing global utility. Most notably, the duration of the experience has almost no effect on such assessments. So an experience with a greater number of painful moments can be evaluated as less painful overall, depending on the relation between the peak and end moments.

Not only do people's direct and retrospective evaluations conflict, but the impact of retrospective evaluation on future (*ex post facto*) preferences eclipses that of direct evaluation. The sum of local utility assessments is overridden by assessments of global utility. Global utility—a function of just two salient moments of the experience—is what is maximized.

When the two evaluations come apart—or when they cannot come together—and when the salient criterion is whether the experience is painful, which evaluation *should* be accorded greater authority? In practice, for the experiencing subject, there may be no live choice here. For, much of the time, preferences are informed by a

person's retrospective, and not direct, evaluations. The effect of retrospective evaluations on choice is considerable since, as Kahneman tells us, "[t]he only utility that people (and other organisms) can learn from personal experience to maximize is the utility that they store in memory" (1999, p. 20). In other words, our past experiences are always mediated by our memory of these experiences, so—left to our own assessments of the quality of our experiences—we can only learn to maximize "the expected utility of anticipated memories" (1994, p. 22).

But several important questions remain. One is the question of the rationality of the subjects' choices in the experiments: is it sometimes rational to prefer more pain to less? Another question, the answer to which is determined in large part by one's intuitions about the first, is the *ethical* question, concerning whether a person's preference for global utility should be honored by others. Kahneman illustrates the ethical problem vividly with the following scenario:

> Imagine a physician conducting a colonoscopy; the patient is in intense pain. The examination is complete and the physician could terminate the procedure now, providing instant relief—and a permanently negative evaluation of the whole episode. Should the physician seek the patient's consent to extend the pain for a while in order to form and retain an improved opinion of the procedure? It is unlikely that consent would be granted. However, the cold pressor experiment tells us that a patient who has had two otherwise identical procedures that differ in the abruptness of relief will prefer to return to the physician who inflicts more total pain but provides a better end. Surely, no consent can be better informed than a choice between two experiences one has already had. When the experiencing self and the remembering self disagree, whom are we to believe? (1994, p. 21)

Kahneman assumes that people will not give consent for the experience to be prolonged. And he may be right, even for people who know of the peak and end phenomenon. But setting this aspect aside, the physician in this case faces the following ethical dilemma: shorten an experience of pain and let the patient's experience be, as a whole, one of which she has a worse subsequent impression, or lengthen the actual pain experienced and thereby lessen the awfulness of the pain that the patient will later think she experienced. Does the fact that such an extension of a painful episode will be appreciated later provide a good reason to extend it now? Should a later, longer-term appreciation for having experienced X override a current, short-term aversion to experiencing X?

The choice is between what seem to be two very different properties of pain: that available to direct evaluation, and that available only to retrospective evaluation. To a first approximation, the common response to such a choice is that what matters more is the moment by moment experience—i.e., that which is available to direct evaluation. Generally, when one speaks of minimizing pain—say, in the world, or over the course of a life, or during a particular day—one usually has in mind minimizing the additive severity of the phenomenological experience as it is occurring. Adding more pain to an experience makes it objectively worse. Kahneman and

Broome share this intuition. They seem to conceive of the issue in terms of the choice between the objective disvalue of pain versus the subjective retrospective evaluation of pain. In "Objective happiness" (1999), Kahneman is explicit about endorsing this formulation of the problem. Such a way of conceiving the problem is, in fact, the bedrock of his entire methodology, which is to grant introspective authority to direct evaluations, and to take local utility as the basic unit of analysis, from which an "objective" or "true" measurement (as Kahneman puts it) of global utility can be calculated (1999, p. 4).

I have not formulated the issue in terms of the objective disvalue of pain versus the subjective retrospective evaluation of pain because I want to push the question further back, to consider in effect which of the utilities identified should be given objective weight. As should become clear, I believe that "objectivity" is not the real issue, in any case: the real question is *what it is rational to desire for its own sake*. Talk of "objective" and "actual" pain masks a value judgment, which itself relies upon a specific interpretation of the data, an interpretation that I will argue below is quite possibly not justified by the studies completed so far.

Broome goes a step further than Kahneman in offering an explanation for the intuition that giving more pain, to increase global utility, is not morally right. He offers the following reason: pain is intrinsically bad. It is "a bad thing *in itself*" (1996, p. 117). Broome takes this to support the intuition that people's retrospective judgments should not be followed when we act on their behalf. Both his explanation and ethical intuition can be questioned. Accordingly, I will address:

1. Broome's view (largely implicit) of what makes something intrinsically bad; and
2. whether accepting the claim that pain is intrinsically bad is sufficient to establish an answer to the ethical question.

These two issues are related. If one makes the assumptions Broome makes with respect to intrinsic value, then one is likely not to recognize legitimate values that can override the intended force of pointing out that pain is intrinsically bad.

Intrinsic value and temporal perspectives

To explain what he means when he says that pain is intrinsically bad, Broome points to an intuitive asymmetry between pain and pleasure:

> It does not matter who experiences it, or where it comes in a life, or where in the course of a painful episode. Pain is bad; it should not happen. There should be as little pain as possible in the world, however it is distributed across people and across time. Nice things are different. The value of enjoyment, for instance, is that it is nice for the person who is having it. Consequently, we value it *through* the valuation the person herself makes of it. So if the person would rather do without a particular bit of enjoyment, perhaps because it slightly spoils the ending of an episode, we will not think she should have it nonetheless. (1996, p. 117)

Pain is different. Because it is intrinsically bad, Broome says, a person should not have more pain just because she prefers it to less. There are two ambiguous elements in this line of reasoning that I would like to address. One is in the temporal perspective from which a person (dis)values pain, and the other concerns the relationship of intrinsic values to people's attitudes. Let's start with the latter of these.

There is lots of evidence of widespread asymmetries between our attitudes towards pain and pleasure (as there are in our intuitions about gains and losses, acts and omissions, and a host of other phenomena) [2]. But, while the asymmetrical intuitions are well confirmed, there's no good reason to think they depend on whether the values in question are intrinsic. It seems quite implausible to think that the asymmetrical intuitions can be adequately explained by invoking a theory of intrinsic value. In the first place, as Broome himself points out, there is good reason to think that the intuitions themselves cannot be made fully coherent [3]. But there is a larger assumption lurking behind Broome's appeal to the supposed asymmetry between pleasure and pain, concerning the relation between people's attitudes and intrinsic value. Broome seems to be saying that while pain is intrinsically bad, pleasure is only extrinsically good because it is made so by the subject's attitudes. He implies that what makes painful experiences bad has nothing to do with one's attitude towards one's experience, while what makes pleasurable experiences nice is contingent on such attitudes.

But this seems confused. The question of intrinsic value is orthogonal to that of whether something's having value depends on the subject's attitudes. Surely pain can be intrinsically bad even if its being bad requires that a subject be averse to experiencing it. Broome's comments suggest that if someone wants to feel a pain, and has no conflicted feelings about this desire, that experience is nonetheless intrinsically bad. This seems intuitively wrong. If one wants a certain pain (and has no conflicting desires), then it is not a bad that needs to be minimized without regard to other features of one's circumstance. However, one is usually averse to experiencing pain, and rejects it for its own sake; this makes pain intrinsically bad.

One might wonder at this point if I have the right sense of intrinsic value in mind. There are several senses in which something might be said to have intrinsic value. One of these should be set aside immediately as irrelevant to the present discussion. By "intrinsic value," one might mean to refer to value which obtains in the world independently of the evaluative attitudes of anyone [4]. (This is might be called *objective* value—though I do not want too much to hang on this terminology, since "objective" itself has a multiplicity of meanings). Whether there is intrinsic value in this sense is a matter of debate. Non-subjectivists argue that there is: they hold that the *source* of value is not in any way dependent upon the attitudes of subjects, but rather is an objective property of states of affairs. In contrast, subjectivists hold that all value is in some sense dependent on the attitudes of evaluating subjects. We needn't differentiate the variety and complexities of the possible versions of these views here. The important point to note here is that these views about the existence of objective value are logically independent of the evaluative claims about pain that Kahneman, Broome, and I want to make. We are not

concerned with whether pain's being bad is an objective truth, in a sense that would require denying subjectivism about values. In short, my present disagreement with Broome doesn't turn on the *metaethical* status of evaluative claims.

There is a second sense of "intrinsic value," whereby an object is valued independently of its effects. This is the notion of *non-instrumental* value. Something has intrinsic value if it is valued for its own sake, as opposed to being valued for the sake of something else. It has value as an end in itself, independently of any instrumental value it may have. On this understanding, pain has (dis)value independently of any further consequences that may follow from it. And this could be true even if the badness of pain were constituted in part by having feelings of aversion. But this could not be the sense that Broome has in mind, for he denies that pain's badness has anything to do with a subject's attitudes towards pain.

To make the disagreement at this point a little clearer, consider the issue of the painful experience of a person, Jane, who is suffering miserably. Looking at her pain from a third person perspective, we can judge that this pain is a bad thing. Broome and I agree so far. But now, we can ask the following question: is Jane's pain bad independently of her attitude of aversion, or is the badness of Jane's pain at least in part a function of this? I find it natural to think that the pain is bad precisely *because* Jane does not like it. One can pick the latter answer—that Jane's pain is bad in virtue of her aversive feelings toward it—and still be able to say that her pain is intrinsically bad, in the sense of being non-instrumentally bad.

In contrast, Broome seems committed to the view that the badness of pain is independent of Jane's desires and aversions. It might seem that Broome is using "intrinsic" in yet a third sense. This might be called a *Moorean* sense of "intrinsic value," according to which something has value in virtue of its non-relational properties. According to this sense, the intrinsic value of something is determined by the "intrinsic nature" of that thing: the non-relational, internal, properties of pain make pain bad. On this reading, if something is valuable in virtue of some attitude someone has towards it, then its being valuable is a contingent and relational fact about it, and thus it cannot be intrinsically valuable. That is, if the badness of pain is independent of any subject's attitudes, and the goodness of pleasure is *contingent* upon such attitudes, then pain is intrinsically bad, and pleasure only extrinsically good.

But even if Broome were right in claiming that the badness of pain is independent of the sufferer's mental attitudes, it is unclear how this could shed light on the ethical question with which we began. The question was whether an episode of pain should be extended in order to give the experience a better ending. The only way that Broome's comments seem relevant to this question is if we assume that the badness of pain were determined by its direct—rather than its global—utility. It turns out that the relation of people's attitudes to intrinsic value is not the real issue raised by Kahneman's studies. The really central question concerns the proper weight to give retrospective attitudes (or, global utility) when comparing them to direct ones.

It is striking that this central issue arises even if Broome were to *grant* that the disvalue of pain is a function of the sufferer's present desires. For what Broome (and

Kahneman) might be seen as particularly committed to denying is that the badness of pain can be affected by one's *retrospective* desires. In fact, I think the main claim Broome *must* be concerned to resist—in order to explain his intuition that it is wrong to increase a painful episode in order to give it a better ending—is not that the value of pain can be constituted by a person's attitudes in general, but the more specific claim that there is a genuine value of pain that is constituted in part by retrospective desires.

This brings us to the first ambiguity mentioned above, concerning the temporal perspective of valuation in the quote from Broome. Recall the distinction between direct and retrospective evaluations of pain. A person's desires can be a product of either sort of evaluation. Thus, one might desire to avoid pain that one is feeling at a particular moment. One can also desire to avoid repeating experiences that on the whole seemed painful in the past. Whether Broome meant his point to apply to the desires formed by direct evaluations is unclear, but as we've just seen, his claim—that pain is bad independently of a person's attitude toward pain—would be implausible if he did. A person's concurrent desires—or aversions—do seem to play a role in the (dis)value of pain. I think Kahneman, if not Broome, might agree with this. Broome's point is more plausible, then, if it is taken to be that a subject's *retrospective* attitudes are irrelevant to the badness of her experience of pain. The idea here is that the valuation of painful episodes *ex post facto* are *objectionably* dependent on that person's attitudes, to such an extent that *the value expressed in retrospective evaluations* can't possibly count as "intrinsic" in the Moorean sense. I take this to be the relevant idea behind Broome's intuition: that the intrinsic (dis)value of pain must be independent of a person's *retrospective* desires [5].

The claim would then be that a preference formed *after* a painful experience has no impact on the experience's real value. The question is whether this is true. Is there also an intrinsic value connected with pain that is in part constituted by a retrospective desire? That is, can a retrospective preference issue in an intrinsic value or disvalue? The rest of the paper will explore an answer to this question.

Global utility as an intrinsic value

Let's grant that pain, as concurrently disvalued, is intrinsically bad in the Moorean sense. I want to argue that several intrinsic values are relevant to the ethical worry, and that it is at least an open question how they are to be balanced. I will offer some reason to take seriously the value of global utility: it may even be better to cause more pain than less if it gives the experience a better ending.

There are at least two intrinsic values that may compete with that of avoiding pain:

1. the value that memories of pain may have *qua* memories; and
2. the value of the overall shape of an extended pain experience.

Each is related to a different interpretation of Kahneman's data. The most common interpretation—which Kahneman explicitly encourages, and which I suspect is behind Broome's thoughts on the matter [6]—is that retrospective evaluations are

"simply wrong" (Kahneman *et al.*, 1993, p. 403); the Peak and End rule produces faulty evaluations of the past. Retrospective evaluations require the implementation of two processes that are not needed for direct evaluations: retrieval from memory of the direct evaluations, and a summary of that information which yields a single evaluation of the experience as a whole. Both of these processes are fallible. This is why Kahneman grants introspective authority to direct evaluations, and not to retrospective ones. Furthermore, as he points out, since many decisions can be based only upon retrospective evaluations, and since these are subject to error, decisions based on such cognitive errors may be easily misguided.

> These results illustrate a general fact of life: except for acts that escape current pain (removing a hand from a flame), the sovereign masters that determine what people will do are not pleasure and pain, but fallible memories of pleasure and pain ... Where retrospective evaluations distort actual experience, subsequent preferences are governed by the distorted evaluation, not by the experience. (Kahneman, 1999, p. 20)

This is, in fact, what Kahneman thinks is going on with his subjects: their choices for the longer episode of pain over the shorter one are the result of erroneous memories. The processes involved in producing retrospective evaluations distort the "true" global utility. On this interpretation, subjects take their global assessments to accurately reflect the totality of local experiences. Unbeknownst to them memory leads them astray. The episodes' endings somehow skew retrospective evaluation so that the total quantity of pain experienced is incorrectly remembered. If this is the right explanation, one can initially see support for Broome's conclusion. For we generally do not think we should do what people want if their preferences stem from mistaken beliefs.

Even if we grant this interpretation, however, it seems to me that Broome's ethical intuition is not supported by the claim that pain is intrinsically bad. It is possible that there is *some* weight to be attached to one's retrospective evaluations of events. It is possible for those memories to have intrinsic value *qua* memories. I am not referring here to the disvalue of a memory that itself produces a painful, or at least unpleasurable, experience. Certainly, there are cases in which a person lives with nightmares or disturbing "flashbacks" of a traumatic experience. In such cases, it seems obvious that it might be better to extend the amount of pain experienced directly if that would diminish the disturbing effects of remembering it. But these cases present a distraction from the ethical question at issue. The reason is that they represent cases in which memories have disutility in virtue of their *effects*: their value is not a function of the intrinsic properties of the memories themselves. (So, we should assume that the retrospective evaluation of pain is not itself a painful or unpleasant experience.) We are looking for *intrinsic* values to rival that of the badness of pain. Can memories have intrinsic value independently of the pleasure or pain they give rise to?

Consider an analogy with optimism. There are two ways in which optimism might be valued. It might be valued instrumentally, because of the chance that one might be a happier person if one is optimistic about the future. Or, it might be

valued intrinsically: one might hold that it is good to be optimistic regardless of its effects on one's general state of well-being—simply because "that's the way to be," for example. Something similar might be true of pleasant memories.

There are further reasons to doubt that direct preferences should always outweigh retrospective preferences. We live with our retrospective evaluations longer than with our direct ones. Perhaps the length of time that we are aware of something shouldn't make a normative difference. Or perhaps it should, given the central role that memories play in our lives. Though in general a person's preferences should perhaps not be honored if based on mistaken memories, it may be that there are cases in which the preference skewed by memory should be taken seriously. In any case, it is *possible* that memories *qua* memories can have intrinsic value.

Still, there is something unattractive about this value. The intrinsic value of having nice memories is intuitively not very strong compared with that of avoiding pain, since it comes at the cost of having false belief. This may account for why the existence of such a value has been ignored or overlooked. To the extent that the memories are distorted, falsely representing an experience as being less painful than it was in fact—the value of memories *qua* memories is not a compelling one. If a person actually endorsed such a value, and unambiguously wanted to have the greater period of pain in order to have a nicer, distorted, memory of the experience, we can understand how paternalistic considerations might preclude satisfying such a desire when acting on their behalf. I just mention this here to acknowledge again how strong our intuitions are against endorsing the value of memories *qua* memories. My aim is not to argue that it is a compelling value, but merely to point out that accepting that pain is intrinsically bad does not on its own settle the ethical question.

There is, however, an alternative interpretation of the data. This yields a quite different retrospective intrinsic value, one that I take to be an attractive intrinsic value to rival that of avoiding pain. On this interpretation, an event's ending plays its own role in retrospective evaluation, independently of whether information is lost. There is some indication that subjects in the experiments retain relatively accurate information about how long the pain lasted. Those theorists who favor the distorted memory account presented above could accommodate this fact by saying that subjects may notionally remember the structure and duration of the experience over time, but that they nevertheless wrongly underestimate the badness of the severity of pain over time. But this explanation of evidence that subjects' memories are not malfunctioning seems to beg the question: why can't it be that subjects remember the relative duration of pain, and don't care about it? What evidence is there to say that caring more about the *shape* of an experience than its duration represents a cognitive error? More studies are needed to test these rival explanations.

The explanation I am offering is that what is preferred is a pain that lessens rather than a pain that does not. I want to suggest that this is a legitimate preference, and one that, moreover, a benevolent third person could morally justify respecting. The intrinsic badness of pain does not show otherwise.

Suppose the right interpretation of the data is that people prefer the shape of an experience, rather than that they prefer nice (flawed) memories. (As far as I can tell, whether this is the case is an open empirical question.) This would have quite

interesting implications about the structure of our values. We may be psychologically constituted in such a way that we simply don't care about how long an episode lasts (within certain temporal parameters) as much as we care about whether it tails off nicely. If so, it may not be irrational in certain circumstances to prefer the shape of an episode to the total quantity of pain. One way of "minimizing pain" may be to give it a better ending. This yields a distinct value represented by global utility, one that is supported by a desire which is formed on the basis of a retrospective evaluation.

The interesting feature of the point I am making is not the general point that avoiding pain can be overridden by other considerations—but that *retrospective* considerations might be sufficient to override the badness of total amount of pain experienced. Broome cannot have meant to deny the first claim. Recognizing that pain is "intrinsically bad" does not commit one to the view that avoiding pain overrides all other considerations. For example, minimizing pain might be achieved by letting someone die rather than administering tests and treatment; but one can rationally decide to accept pain if that is the cost of a necessary colonoscopy. So there are cases in which what is intrinsically bad can be rationally preferred to something else. Of course. But the point at issue concerns the *type* of consideration that might be a candidate for overriding the value of minimizing pain—and, moreover, for doing so systematically.

We should note that the value of a global experience of pain is itself an *intrinsic* bad. This fact is missed by Broome's explanation. It is intrinsic in the Moorean sense: the global value supervenes on the intrinsic properties of the painful experience. The question I am raising is this: if external properties (such as pain's being an unavoidable consequence of a procedure that may save one's life) can override the intrinsic badness of pain, then why can't other internal properties of pain do so?

The total quantity of pain is one property that we intrinsically value or disvalue. But I am suggesting it is not the only such property. This raises the possibility of conflicts of value over *internal* properties of pain. One wants the least amount of pain, but one also wants the pain to tail off nicely.

Broome might make the following objection to my suggestion. The overriding external properties mentioned above can be construed as hedonic properties, in the sense that they concern how *much* pain or pleasure is experienced overall. I could reply that this is too crudely hedonistic about values. And this would be correct. But then Broome would ask us to remember a significant fact about the studies in question: these test subjects *had no other objective than a narrowly hedonic one*. Well, in one sense, that is right: subjects had no incentive for enduring greater pain rather than less. None of the usual considerations that might override the badness of pain were relevant to their situation. And yet the majority chose a greater aggregate of pain.

My response is just this: we may have to redefine (at least, speaking theoretically—a redefinition may not be necessary in practice) what we mean by "pain," as well as by "hedonic property." The shape of an event may very well affect its overall hedonic value. To avoid word games, one might stipulate that a "hedonic" property is one that bears directly on the total aggregate of pain or pleasure. But then one

would be begging the question if one said that this was the only goal relevant to the decision the subjects were asked to make in the experiments.

Let me now consider a different sort of objection to my suggestion. One might acknowledge that shape is indeed a relevant value for many experiences, and still doubt whether it could plausibly be attributed to the study subjects with respect to the sorts of experiences generated in these studies (loud unpleasant noises, the medical procedures of colonoscopies and lithotripsies, hand immersion in cold water, short plotless films of pleasant and unpleasant scenes). Many philosophers (among them: Velleman, 1991; Griffin, 1986; Stocker, 1990) have argued that the global value of an experience is not always a mere aggregate of the value of the sum of its parts. And this seems intuitively right, for a large class of experiences: just consider the experience of a symphony, a play, a date, a phone conversation, a relationship that ends particularly well or badly, or an entire life. As Mane Hajdin has noted, "nobody expects my evaluation of a symphony to be simply a sum of my evaluations of its moments, and most definitely not a sum of my evaluations of its individual notes. Similarly nobody expects my evaluation of a movie to be the result of straightforwardly adding up my evaluations of its individual scenes" [7]. But the experiences generated in the studies considered here are not like that. There are many differences between these types of experiences, but for now let's just focus on one.

Remember that it is plausible to assume—as Kahneman does—that subjects who are offered the choice of removing their hand from the cold water or keeping it in to experience a better ending would choose to remove their hand immediately. Hajdin has pointed out that an analogous assumption for more complex cases, such as watching a boring movie or sitting before an unpleasant rendition of a symphony, is not plausible. If they know that the ending will be somewhat less bad than the rest, people—unsurprisingly—might prefer to prolong an unpleasant experience of a movie rather than walk out in the middle of it. "They may reasonably expect that staying till the end will make their experience rounded off in a way that is desirable even when the individual elements of the experience are undesirable" [8]. In these cases, the experience is recognized *while* it is going on to be of the sort that shape matters. In contrast, in the episodes featured in the experiments, there is no reason to think that the shape of the experience is thought by the subjects to be relevant while it is going on.

I agree with this depiction of the difference between people's current, in-the-midst-of-the-experience judgments regarding the two sorts of experiences. But I am not sure that it is relevant. For my argument to work, shape does not have to be recognized as a relevant factor from the perspective of the direct evaluations. It is enough if it is judged so *ex post facto*. And, in fact, neither does it have to be *consciously* so recognized, in order for shape to influence the retrospective judgment. It is a mistake to think that one must be able to extrapolate shape judgments from subjects' *direct* judgments in order to acknowledge that shape judgments are nevertheless relevant. Two considerations support this. First, the desire to avoid pain may affect the prudential judgment to maximize global utility, without threatening my suggestion that shape may be what subjects are actually choosing for *ex post*. Pain

can swamp prudential or instrumental considerations. Second, setting aside the special effects of pain, there may be some situations—such as those created in the experiments—in which we are simply more or less hard-wired to value shape. I want to suggest that it may nevertheless be true to say of such cases that the retrospective judgment represents the value of the shape of the experience.

Still, I do admit that there is a difference between cases in which shape is recognized as a relevant value while the experience is unfolding, and others in which it is not. In fact, though I have not the space to elaborate the point here, I think the difference between these cases is significant and substantial. Hajdin suggests that the difference lies in the degree of complexity. He calls the experiences featured in Kahneman's experiments "utterly simple." Without exploring what exactly could be meant by this, I acknowledge that the so-called "simple" cases are significantly unreflective, and the goal—avoiding pain—is also very simple (narrowly hedonic). As a result, I think one should be *very* careful about making generalizations about either of "complex" or "simple" experiences based solely upon evidence concerning the other sort. Having said that, there are also intriguing comparisons to pursue.

For example, when speaking of the evaluation of well-being over the course of entire lives, Velleman (1991) argues that, at least for lives with some minimal "narrative structure," the overall value of a life is not a function of the value of individual moments within that life. His treatment of the issue suggests that the value of an experience may be not only time–context sensitive, but also largely a matter of how one interprets one's "life story" (as well as the substories that make it up). But it's hard to see how interpretation could play a role in the judgments made by Kahneman's subjects. On the contrary: it's plausible to think that one does not have a choice about how to "see" the global utility of the sorts of experiences induced in the experiments [9]. So at a minimum, there are differences in terms of narrative structures and the possibility of reinterpreting them. In the experiments featured in Kahneman's studies, the story is a story that a dog could appreciate: it was bad, and it got better. Narrative structure is present. But the experiments give us reason to think that the very notion of narrative structure itself has structure. I am suggesting that there are primitive narratives that are unambiguous. Interpretation is not an option with these sorts of "shapes."

The intrinsic value revealed by this alternative explanation of the data is more appealing than the value of memories *qua* memories. For one thing, the value of the overall shape of an extended episode is not predicated on making any sort of cognitive error. For another, global utility may be intricately bound up with our conception of who we are, in a way that is not simply a matter of how long we live with our retrospective evaluations. Though this alternate value depends on memory, it is not about memory. Different values are formed from different perspectives. It may be that we identify closely with the perspective from which the shape of an event is intrinsically valuable. These are speculations for another paper. I merely suggest here that the intrinsic value of an experience as a whole might outweigh the intrinsic badness of its separate moments combined [10]. Allowing for this possibility does not of course itself establish that such a value *does* outweigh the badness of

pain, or that it should for a third person. The ethical question is one of how to weigh different intrinsic bads.

Acknowledgements

Many thanks to Ruth Chang, Frankie Egan, Mane Hajdin, Brian Loar, Barry Loewer and Stuart Rachels for helpful comments on earlier drafts of this paper.

Notes

[1] Or similarly, before an experience which may occur in the future. Evaluations which I call "temporally removed" may be retrospective or prospective. Here we are concerned only with retrospective evaluations. I have more to say about temporally removed evaluations in general in my "All's well that ends well: cognitive psychology and the bias toward endings" (unpublished manuscript).

[2] For some discussion and a few citations of the asymmetry in attitudes toward pleasure and pain, see Kahneman (1999, pp. 17–19).

[3] To further support his account, Broome appeals to the intuition that "there is no duty to have a child who would have a good life, but there is a duty not to have a child who would have a painful life" (1996, p. 117). While I cannot fully explore this intuition here, and while I grant Broome's own doubt that such asymmetrical intuitions can be made fully coherent, I think an explanation of it can begin with the deep-rooted psychological fact that we often think pain is much worse than pleasure is good. There is, in Kahneman's words, "differential sensitivity to aversive and positive events," which is captured in the phenomenon called "loss aversion." This might account for the intuition. Again, I see no reason to think that the asymmetry between pleasure and pain has anything to do with whether the values in question are intrinsic, or with the relation of intrinsic value to people's attitudes.

[4] Throughout this discussion of intrinsic value, I am indebted to John O'Neill's treatment of the notion in his paper, "The varieties of intrinsic value" (1992).

[5] Harry S. Silverstein misses this distinction between the two distinct preferences (the direct and the retrospective) when he argues that "to refuse to accede to [the subjects'] preferences seems straightforwardly, and offensively, paternalistic" (1998, p. 149). The charge of paternalism misses the mark, since the question is *which* of the subjects' preferences—that expressed by direct evaluation or that expressed retrospectively—should be honored. Silverstein might respond that it is objectionably paternalistic to insist on one of the preferences over the other. This might be right. But in that case, the charge of paternalism makes sense only if we recognize what I will argue for below: the intrinsic value of the individual preferences. (I came across Silverstein's paper after I had finished writing this paper.)

[6] Though Broome doesn't say so in his short (1996) piece, there is a good deal of evidence for this in the theory he develops in his *Weighing goods* (1991), especially Chapter 11.

[7] From the comments presented by Mane Hajdin (1999) in response to an earlier version of this paper.

[8] Hajdin (1999).

[9] I suspect there may in fact be many more than two perspectives from which a person can evaluate her experience, and in some cases, it may not be a matter of choice which of these perspectives one adopts. Furthermore, it seems to me that each of these perspectives would spawn its own set of values.

[10] An interesting alternative possibility is that the values in question may be incomparable. Though this might make our ethical situation more dramatic, the main points of the argument would still stand.

References

BROOME, J. (1991). *Weighing goods*. Oxford: Blackwell.

BROOME, J. (1996). More pain or less? *Analysis*, *56*, 116–118.

FELDMAN, F. (1997). On the intrinsic value of pleasures. *Ethics*, *107*, 448–466.

GRIFFIN, J. (1986). *Well-being: its meaning, measurement, and moral importance*. Oxford: Clarendon Press.

HAJDIN, M. (1999). Comments on Stephanie Beardman's paper. Paper presented at the *Pacific Division Meeting of the APA*, 1 April.

KAHNEMAN, D. (1994). The cognitive psychology of consequences and moral intuition. Tanner Lecture on Human Values given at the University of Michigan, unpublished manuscript.

KAHNEMAN, D. (1999). Objective happiness. In D. KAHNEMAN, E. DIENER & N. SCHWARZ (Eds) *Well-being: the foundations of hedonic psychology*. New York: Russell Sage Foundation.

KAHNEMAN, D., FREDERICKSON, B.L., SCHREIBER, C.A. & REDELMEIER, D.A. (1993). When more pain is preferred to less: adding a better end. *Psychological Science*, *4*, 401–405.

O'NEILL, J. (1992). The varieties of intrinsic value. *The Monist*, *75*, 119–137.

SILVERSTEIN, H.S. (1998). More pain or less? Comments on Broome. *Analysis*, *58*, 146–151.

STOCKER, M. (1990). *Plural and conflicting values*. Oxford: Clarendon Press.

VELLEMAN, J.D. (1991). Well-being and time. *Pacific Philosophical Quarterly*, *72*, 48–77.

Copyright of Philosophical Psychology is the property of Carfax Publishing Company and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.