The Worst Time to Die

Ben Bradley

1. TWO DEATHS

At what stage of life is it worst for a person to die? Here are two fictional examples to help frame the question.

Baby. A three-week-old baby, Baby, dies in an accident. Had Baby not died then, he would have enjoyed a happy childhood and adolescence, gone to college, entered a PhD program in philosophy, become a professional philosopher, and lived an enjoyable life until dying at age 80.

Student. A 23-year-old philosophy graduate student, Student, dies in an accident after a happy childhood and adolescence. Had Student not died then, he would have become a professional philosopher and lived an enjoyable life until dying at age 80.[1]

Which victim's death is worse for him? In what follows I defend a simple view that entails that death is typically worse the earlier it occurs, and thus that Baby's death is worse than Student's. Jeff McMahan and Christopher Belshaw have argued that Student's death is worse (McMahan 2002; Belshaw 2005); I show that their views face serious problems.[2] I focus in particular on McMahan's "Time-Relative Interest Account" of the evil of death, according to which the magnitude of the harm of death is determined in part by facts about psychological connectedness.

One might wonder: what is the point of this exercise? I find the topic of the magnitude of the evil of death inherently interesting; others might not. But those others might yet be interested in other topics of obvious practical significance. A principle that entails that Baby's death is much less bad than Student's might also entail that the death of a fetus is even less bad, or not very bad at all for the fetus; this would have implications for the moral permissibility of abortion. This is part of McMahan's strategy for defending the permissibility of some abortions (McMahan 2002, ch. 4); David DeGrazia claims that McMahan's arguments "substantially advance the case for a liberal position" about abortion (DeGrazia 2003, 442). Another example concerns the allocation of medical resources. Societies face decisions about what portion of their resources they should devote to reducing infant mortality. Devoting resources to reducing infant mortality might require taking resources away from other projects that help save the lives of older people. One consideration that seems relevant to such decisions is the relative magnitudes of the harms at stake in the decision to allocate resources in one way rather than another. (Of course, other factors are surely relevant as well.) John Broome describes many other examples of situations that require some sort of weighing of the harms of deaths of people of different ages (Broome 2004, 1-18). Anyone interested in these issues should be interested in the question about Baby and Student.

2. LCA

The most popular view about what makes death bad for its victim is the deprivation view: death is bad because it deprives its victim of a good life.[3] An obvious

way to understand the idea of deprivation is in terms of counterfactuals: a person's death is bad because, if the person had not died, she would have enjoyed more of a good life. This provides an obvious account of the extent of death's badness: *how bad* it is for someone to die depends on *how much better* things would have gone for her if she had not died when she did.[4] We may generalize and say that how good or bad any event is for someone depends on how much better or worse things would have gone for her had that event not happened. This is the Life Comparative Account (LCA). If we interpret the counterfactual conditional in the now familiar way due to Lewis and Stalnaker, here is what the view says about the overall values of events:[5]

LCA: The overall value, for a person S, of event E at world W = the value for S of W (i.e. how well-off S is at W), minus the value for S of the closest world to W in which E does not occur.

Nothing important hangs on the appeal to possible worlds; those who prefer a different account of counterfactuals should feel free to make the appropriate translations.

LCA is a threadbare schema for evaluating events; it leaves open which particular sorts of things affect a person's welfare. But given any plausible theory of welfare, LCA will entail that Baby's death is worse than Student's. There are some theories of welfare according to which Student's life contains some evils, as a result of his premature death, that Baby's life does not contain. For example, it is likely that Student has many more desires, and stronger desires, than Baby has; a desire satisfactionist will say that Student's death causes more desire frustration than Baby's death. But Student's life contains many

desire satisfactions that Baby's life lacks; so, plausibly, Student's life is better than Baby's.[6] As long as Student's life is better than Baby's, LCA entails that Baby's death is worse.

2.1 The Cure

McMahan presents a number of arguments against LCA. One is based on the following example:

> *The Cure.* Imagine that you are twenty years old and are diagnosed with a disease that, if untreated, invariably causes death… within five years. There is a treatment that reliably cures the disease but also, as a side effect, causes total retrograde amnesia and radical personality change. Long-term studies of others who have had the treatment show that they almost always go on to have long and happy lives, though these lives are informed by desires and values that differ profoundly from those that the person had prior to treatment. (McMahan 2002, 77)

McMahan says that it makes sense to refuse the treatment in this case, given how little one's present and post-treatment selves would have in common psychologically were one to receive the treatment. But LCA entails that it is better to receive the cure. The psychological disunity is irrelevant according to LCA.[7]

I think LCA gets the right result here. The decision to refuse treatment is shortsighted and irrational. It seems in many ways similar to the decision of a child to ignore the consequences of his behavior on his adult self, since he does not currently care about the things his adult self will care about. Perhaps some others will share my judgment. But not everyone will. So it is worthwhile to think about some possible sources for the judgment that it is rational to refuse the cure, and to determine whether, if the judgment is based on those sources, any problem is posed for LCA.

In the example, we are to suppose that if I take the cure, *I* will continue to exist through the changes in personality, memory, desires and values. But these sorts of examples might tend to lead people to a psychological account of *personal identity,* not merely prudential concern. When McMahan says that if you received the treatment, in the future "you would be a complete stranger to yourself as you are now" (McMahan 2002, 78), he is very close to saying that psychological continuity is necessary for personal identity (though he explicitly denies that claim). Someone in the position described in the example might well reason in the following way: *I think I'd rather take the cure and get the extra life. But I'm not entirely sure I would continue to exist after getting the treatment, given the psychological changes I would undergo. This depends on the truth about personal identity over time, and there is no consensus about what such identity amounts to. Better not to risk it.* If one reasons in this way, it is doubt about whether the future individual is the *same* individual that is the basis for the decision. In discussing a similar case, Christopher Belshaw says:

> What we want is for our own lives to continue, not for there to be some life or other, no matter how good, that starts where our own life leaves off… It's bad for him to die, but it's bad, as well, to turn into someone else, or to become a person he cannot, now, much care about. (Belshaw 2005, 48-49)

McMahan and Belshaw are clearly *tempted* to say that the pre-cure individual does not survive, but is replaced by someone else. Surely others feel the same temptation. But to the extent that this temptation forms the ground for the judgment that it is rational to refuse the cure, there is no argument here against LCA. If receiving the cure causes the patient to go out of existence, LCA entails that benefits and harms accruing to the post-cure patient are not relevant to the evaluation of the life of the pre-cure patient.

Another reason one might choose not to receive the cure is that the details might make a difference. In particular, it might matter *in what respects* the patient's desires change as a result of the treatment. We often care about our future welfare even when we know that in the future we will desire things we currently do not desire; but our caring is limited. For example, I might reflect on the sorts of changes that people typically undergo as they age, or that my older relatives have undergone, and realize that I myself am likely to undergo similar changes. I might come to believe that I will become interested in playing golf, and that my politics will become more conservative. I might well take some golf lessons in the hope of making my future self happier, but refuse to contribute money to conservative political causes (or to save money for my future self to contribute to those causes) even though that would also make my future self happier. We might say that whether we should care about the desires of our future selves depends on

whether those future selves will desire not only what we are currently not interested in, but also what we think is *bad*. If my future self will have different and (from my current perspective) worse moral values, I might choose not to be able to act on those values; this might be because I care about things other than my own well-being, or (perhaps in a more extreme case) it might be because I think I wouldn't be well-off if I held such values, even if I were pleased.

There is another possible source of intuitive support for the idea that one shouldn't receive the treatment. If one were to receive it, one's life might contain more valuable episodes, but it would lack unity. This might be even more obvious in the case of someone like the musicologist Clive Wearing, whose life seems to consist of millions of very short, unconnected experiences. Consider what McMahan says about this sort of case: "While we may think that the experiences have value individually, it is less plausible to attribute independent value to them as a collection or aggregate" (McMahan 2002, 76). And recall what Belshaw says about the cure: "it's bad… to turn into someone else" (Belshaw 2005, 49). Both McMahan and Belshaw seem to think that, even if there is a single person surviving through the cure, and the parts of the person's life are good taken individually, the life taken as a whole is not so great. Perhaps such a disjointed life scores very low on the narrative unity scale. Since attributing value to lives on the basis of shape or narrative unity is compatible with LCA, LCA can give the results McMahan and Belshaw desire in these cases.[8]

To put the point another way, consider someone who judges that the treatment is not in her interests. She might decide, before the proposed treatment date, that the treatment is not in her best interests *then*. But she might also make the *time-neutral*

judgment that a long, happy life interrupted in the way the treatment required would be less good for her than a shorter life without the treatment.[9] In order to get an argument against LCA going, McMahan and Belshaw need to argue that the time-neutral judgment is not what's doing the work.

Some might find the cure unappealing because of the *way* in which one's desires and values change.[10] They change as a result of the medical intervention – not as a result of more ordinary processes like maturing and aging.[11] The new desires and values are, in a way, imposed upon the patient rather than arising naturally. Some might find such a life worse than a shorter one in which the desires and values arise as a result of ordinary processes, even though the change results from the patient's own decision to receive the treatment.[12] Again, McMahan and Belshaw need to argue that time-neutral judgments about the disvalue of having one's desires altered by a medical intervention are not driving the argument here.[13]

To recap briefly, there are several reasons not to receive the cure that are compatible with the truth of LCA. (1) Receiving the cure would cause the patient to cease to exist, and to be replaced by a different individual. (2) We care about more than our own well-being, and receiving the cure might be incompatible with other things we care about (e.g., doing what we now take to be the morally right thing). (3) One's lifetime welfare level is partly determined by certain sorts of psychological unity. (4) One's lifetime welfare level is determined not only by whether one's desires are satisfied, but also by the way in which they arose in the first place. (Note that one cannot accept all four explanations at once! But only one is necessary.) Since I think I am rationally required to take the cure, I do not endorse any of these responses; but those who think it

is rational not to take the cure should consider whether they think so because they believe one of these things, rather than because they think LCA is false.

2.2 Abortion

The most serious problem McMahan sees for LCA is that it seems to entail that in typical cases, the worst possible time for a person to die is just after she comes into existence. In the following paragraph he targets this apparent implication of LCA:

> If identity [rather than psychological connectedness] were what matters, the worst death, involving the most significant loss, would be the death of an individual immediately after the beginning of his existence. But the loss that would have occurred if that individual had simply been prevented from beginning to exist would not have been significant at all. This is hard to believe. It suggests that it is profoundly important to prevent the existence of an individual who would die within seconds of beginning to exist. (McMahan 2002, 171)

McMahan finds it hard to believe that it should matter so much whether a being fails to come into existence at all, or comes into existence only to die moments later. But LCA apparently entails that it does indeed matter a lot. McMahan thinks most people will agree that it doesn't matter, and cites as evidence the fact that most people are not extremely concerned to stop spontaneous abortions; if people really thought that earlier

deaths were worse, they would view "the vast number of spontaneous abortions that occur as a continuing tragedy of major proportions" (McMahan 2002, 165).

Since this is McMahan's primary argument against LCA, it will be helpful to spell it out in some detail. As I see it, the argument goes something like this:

P1. If LCA is true, then death is extremely bad for its victim at the earliest stage of life.

P2. Failing to come into existence is not bad for a person. ("The loss that would have occurred if that individual had simply been prevented from beginning to exist would not have been significant at all.")

P3. If P2, and if death is extremely bad for its victim at the earliest stage of life, then it is extremely important to prevent someone from coming into existence if he would otherwise die just after coming into existence.

P4. It is not extremely important to prevent someone from coming into existence if he would otherwise die just after coming into existence.

C. Therefore, LCA is not true.

P3 should be rejected. Before I explain why, let me briefly touch on the other premises. P1 is true, provided that the victim's life would have been worth living. I can imagine someone taking issue with P4, in the following way: Coming into existence is a big deal in the life of an individual. Many things are true of a person after she comes into existence that weren't true before. One of those things is that she can die. Another is that she can be harmed. Insofar as we care about preventing harms to people, then, we must

think it is important to prevent someone from coming into existence if he would otherwise die shortly after coming into existence.

P2 raises interesting questions. Accepting P2 requires us to take controversial stands on unsettled metaphysical questions. Suppose I'm considering some harms I might have suffered in the past. If I had died shortly after coming into existence, I would have been harmed – on this point all agree (disagreeing only about the extent of the harm). Suppose instead I had been prevented from coming into existence. McMahan says that wouldn't have been bad for me, since in such a case, there would have been no subject of harm, no victim.[14] Perhaps in order to be harmed (or have any properties at all) at a world, one must exist at that world. But that's not obvious. We might consider a world in which my existence is prevented, and say that a number of things are true *of me* – the actual me – at that world: e.g. that it's not the case that I exist, that it's not the case that I enjoy myself or have any knowledge. We might conclude that an event that brought about these sorts of negative facts about me would be harmful to me at that world, even though I never exist there.[15]

Investigation of these tricky matters of existence would require another paper. So let us suppose for the sake of argument that P2 is true. That leaves us with P3. P3 is true only if it is always important to prevent great harms, and it seems clear to me that it is not. This should be agreed upon by philosophers of all persuasions, as I will now show.

First, we might make a distinction between the extent to which an event *harms an individual*, and the extent to which that harm *matters morally*. It is entirely consistent to say both that the death of an infant or a fetus harms it greatly and that the harm does not matter morally as much as a harm to an adult.[16] To explain how this could be so, we

might appeal to the notion of desert.[17] Perhaps a harm matters more morally when it is suffered by someone who deserved to be benefited than when suffered by someone who didn't deserve any benefits. Or we might appeal to a notion of moral status. We might say that certain individuals have moral status, while others lack it, and that harms matter morally only when suffered by individuals with moral status; or perhaps that moral status comes in degrees, and that harms matter more when suffered by individuals with greater moral status.[18] (In fact, McMahan himself seems sympathetic to such ideas.[19]) If we hold such a view, we can reject P3.

Of course, not everyone finds notions like desert and moral status unproblematic. Consider a consequentialist who thinks that the permissibility of an action depends on the amount of value produced by the action and its alternatives, where that value is not determined in any way by facts about moral status or desert. Suppose this consequentialist accepts P1 and P2. In a situation where, were a person to come into existence, he would die immediately afterwards, here is what the consequentialist says. There are only two alternatives: (a1) prevent the person from coming into existence; (a2) let the person come into existence, only to die immediately. Even though nothing bad happens to the person if a1 is performed, and a great harm is suffered by the person if a2 is performed, and (we can suppose) nobody else is affected, the consequentialist has no reason to favor one of these alternatives over the other. The same amount of value would be produced either way. The consequentialist should say that when comparing two actions, *it does not matter* whether one of the actions causes an event that is harmful to someone while the other does not; the intrinsic values of the two outcomes are all that matter.[20]

It might seem surprising that consequentialists should say that harms do not matter. But in fact, everyone should agree about this – whether consequentialist or not – once the claim is properly understood. Remember that death is a harm that is merely *extrinsically* bad for its victim. Death is harmful because of what it prevents; it has a sort of extrinsic value. When determining whether there is any reason at all not to perform the action, we should never consider the extrinsic values of the results of the action. For example, suppose you save my life by curing me of a debilitating disease, and as a result I live an extra ten very happy years. I then die from an unrelated cause, and my later death deprives me of yet another ten happy years. Your saving my life caused my later death, and my later death was very bad for me; but my later death was merely *extrinsically* bad, and in evaluating how good it was that you saved my life, the extrinsic badness of my later death is irrelevant.[21] Before my death I would not be justified in complaining: *yes, you saved my life and gave me ten happy years, but you also caused an event that will prevent me from getting another ten happy years; so the bad you caused cancels the good you caused*. The reason this would be an unjustified complaint is that your saving my life, rather than letting me die, did not prevent me from getting those later ten years of a happy life that my later death prevents me from having. Similarly, in the case at issue in McMahan's argument, the act of allowing a person to come into existence, rather than preventing his existence, does not prevent that person from getting the goods of which his death deprives him. Even though the act of allowing the person to come into existence causes another event that is very harmful to the person, that fact is not relevant to the evaluation of the act.

Thus P3 depends for its plausibility on a principle that nobody should accept: the principle that it is important not to cause harmful events (i.e., extrinsically bad events) to occur. It is important not to cause an extrinsically bad event to occur only when, in so doing, one is causing something intrinsically bad to happen or preventing something intrinsically good from happening. For example, it is important not to cause a person's death because in so doing, one is preventing that person from receiving some intrinsic goods. This is not the case in McMahan's example, since allowing someone to come into existence only to die shortly afterwards does not cause any intrinsic evils or prevent any intrinsic goods. So even if death is worse the earlier it occurs, we might have no obligation to prevent very early deaths from occurring by, say, preventing doomed conceptions from occurring. In light of the falsity of this principle about harm, we must reject P3.

Naturally, there are many other problems one might find with LCA, including all the arguments given against deprivation accounts generally. I cannot discuss all of those arguments in this paper; I have focused on these particular arguments because they are the sorts of arguments that might lead one to hold a view that entails that Student's death is worse than Baby's. I now turn to such views.

3. ALTERNATIVES TO LCA

To find a way in which Student's death is worse than Baby's, we must find some important way in which Student and Baby differ. But not only that: it must be a difference that cannot be incorporated into LCA. That is, it must be shown that, *even*

though Student is deprived of less of a good life than Baby – even though Student's actual life is better than Baby's actual life – Student's death is worse. This is important to remember. One difference between Baby and Student is that Student has invested a great deal of time and effort into seeing to it that his future goes a certain way, while Baby has not. Student's death renders those investments futile. Another difference is that Student's life has begun to take a certain sort of shape or narrative; the narrative of Baby's life has not yet really begun. Student's death gives his life a very bad shape, or a tragic narrative, while Baby's death just results in his life having no real shape or narrative at all.[22] Both of these differences between Baby and Student can be accounted for by LCA, as long as it is intrinsically bad to have one's investments rendered futile or to have a life with a bad narrative structure. And if those things are not intrinsically bad, why think they are relevant to the evil of death?

If the best theory of personal welfare entails that it is so bad to have one's investments rendered futile, or to have a life with tragic narrative structure, that Student's life turns out to be worse than Baby's, then LCA would yield the result that Student's death is worse than Baby's. It seems very implausible to me that the best theory of personal welfare would entail that Student's life is worse than Baby's, and to my knowledge no such theory has ever been defended. But my concern here is to defend LCA, not to argue about the theory of welfare.

3.1 What We Want When We Die

Another important respect in which Baby and Student are different is that Student has fully formed desires about his life. At the time of his death, he wants the future he would have had. Baby does not have any such desires, at least not about his future beyond the next few minutes. As noted in Section Two, this does not by itself make Student's death worse, since Baby's death deprives Baby of many satisfactions of not-yet-formed desires. But we might take deprivations of satisfactions of not-yet-formed desires to be irrelevant to the value of Baby's death. Here is what Belshaw says about a case much like Baby and Student:

> Alice [the teenager] herself is already looking forward to, and planning for this future, has already made an investment in and commitment to this upcoming life. It's what she wants. Ben [the baby], of course, is hardly yet able to think beyond his next meal… In this sense the ending of his life, no matter how good it will be, doesn't represent a loss to him. (Belshaw 2005, 46)

Belshaw never states the view he is putting on the table, but here is a simple view that is suggested by his remarks: how bad it is for a person to die depends on, at the time of the person's death, how much the person *wants* the future he would have had, rather than its *value* for the person.[23] If this view were true, Baby's death would be much less bad than Student's.

But this is unacceptable. Consider a depressed teenager who commits suicide. At the time of her death, she likely does not desire the future she would have had. But her death is still very bad for her.[24] The problem is that people sometimes don't desire what

is really good for them, and in those cases, what seems relevant to the evaluation of their deaths is what would really have been good for them, not what they want (except insofar as just getting what one wants is good).

An obvious fix to this problem would be to take not the person's actual desires at the time of death, but rather some idealization of those desires. But we face an uncomfortable dilemma. Either the idealized desire set contains appropriate desires (i.e. desires of the right strength) for what is really good for the person, or it doesn't. If the latter, the view will get the wrong results, e.g. in the case of the depressed teenager. If the former, the view will get the same results as LCA, and in particular will entail that Baby's death is very bad – the very result the view was supposed to avoid. For on that assumption, Baby's idealized desire set will contain desires for all the future goods his death takes from him.

Perhaps a clever person could get around these sorts of problems. I am skeptical. In any case, Belshaw himself does not really seem to endorse the view that how bad it is for a person to die depends on her desires at the time of death. Rather, he thinks it is facts about psychological connectedness that are really important:

> Although it's bad for young children to die, then, we might with reason think it's less bad than a similar death for an older child, or an adult. Why? Because there's not the same degree of psychological connectedness, or integration, running through their lives. (Belshaw 2005, 46)

Belshaw seems to be endorsing something like McMahan's "Time-Relative Interest Account" (TRIA), which is inspired by Parfitian considerations about psychological connectedness. I now turn to that view.

3.2 The Time-Relative Interest Account

According to LCA, the value of a death is determined by facts about the values of lives. TRIA is different in one way that seems superficial: according to TRIA, the value of a death is determined by facts about "interests." Interests are not desires; they are to be understood in terms of well-being. What is the relationship between interests and well-being? One's interests, according to McMahan, "reflect what would be better or worse for one's life as a whole"; for one to have an interest in something is "to have a stake in it" or "for one's well-being to be engaged with it" (McMahan 2002, 80). These remarks do not make clear exactly what McMahan takes the relation between interests and well-being to be. I will assume that he thinks that something is in a person's interests if and only if it is intrinsically good for that person; that is, a person has an interest in something if and only if she has egoistic reason to care about it. On the other hand, "one's present time-relative interests are what one has egoistic reason to care about *now*" (McMahan 2002, 80). A person's present time-relative interests can diverge from his interests *simpliciter* provided that the *prudential unity relations* holding between his present and future selves are weak – i.e. provided that his present self has less egoistic reason to be concerned about his future self than about his present self.

Of course, if the prudential unity relation relates person S1 at time t1 to person S2 at time t2 iff (and to the extent that) S1=S2, then since identity does not come in degrees, one's interests and one's time-relative interests would never diverge. McMahan argues that the prudential unity relation is not identity, but psychological unity. (Here McMahan's view is in the spirit of Parfit, who says that "personal identity is not what matters" (Parfit 1984, 217).) The extent to which S, at time t1, has a t1-relative interest in event E at time t2 depends in part on the extent to which S at t1 is psychologically unified with S at t2. Psychological unity comes in degrees, so interests and time-relative interests can differ in strength. Here is how McMahan characterizes psychological unity: "The degree of psychological unity within a life between times t1 and t2 is a function of the proportion of the mental life that is sustained over that period, the richness or density of that mental life, and the degree of internal reference among the various earlier and later mental states" (McMahan 2002, 74-5). By "internal reference" McMahan has in mind things like memories of past experiences and acts that satisfy past desires (McMahan 2002, 74).[25]

Thus the extent to which an event is (instrumentally) bad for a person depends not only on how (intrinsically) bad its effects are for the person at the time the effects occur, but on the strength of the psychological unity between the person at the time of the instrumentally bad event and the person at the time of the event's bad effects. When the connection is not strong, the intrinsic value of the effects is multiplied by a fraction corresponding to the strength of the connection – the weaker the connection, the smaller the fraction (McMahan 2002, 80). So long as there is "physical and functional continuity of the areas of the brain in which consciousness is realized" (McMahan 2002, 79), the

fraction will be greater than zero; such continuity is sufficient for *some* degree of egoistic concern even when psychological unity is lacking.[26]

Given that background, here is how McMahan states the relevant parts of TRIA:

> We cannot, I have argued, assess how bad it is for a person to die simply by ascertaining how much better his life as a whole would have been if he had not died when and how he did… In addition to asking how much good a person's future life would have contained, we must also ask… How close would the prudential unity relations have been between the individual as he was at the time of his death and himself as he would have been at those later times when the goods of his future life would have occurred?… The badness of the loss must be discounted for the absence of [psychological connections]. (McMahan 2002, 183-4)

When we try to state TRIA more carefully, it turns out to be a fairly complicated view. It is best formulated in two parts. First, there is a part that tells us, for any actual or possible event, and any time t, the extent to which a person has a t-relative interest in that event occurring. Let us introduce a piece of technical jargon here: the extent to which a person has a t-relative interest in an event occurring will be called the *t-relative interest value* of the event for the person.[27] Here is what seems to be McMahan's account of t-relative interest value:

TRIA1. The *t-relative interest value* of an event E for subject S = the intrinsic value of event E for subject S, multiplied by a fraction corresponding to the level of psychological unity between S at t and S at the time of E. (In the case of merely possible events, the fraction corresponds to what *would have been* the level of psychological unity between S at t and S at the time of E had E occurred.)

TRIA1 is just a definition of a piece of jargon. It doesn't provide us with an account of the *overall value* of anything – i.e., the value something has for a person taking into account not only its intrinsic value for that person, but also its value in virtue of what it brings about and prevents for him. Thus TRIA1 doesn't provide a way to evaluate deaths; TRIA needs a second principle that gives us the overall value of an event for a person based on the time-relative interest values of other events for that person. Given that someone's death causes and prevents certain events, and given the t-relative interest values of those events (for some t), how do we determine the value of that death for its victim? McMahan does not answer this question. I tentatively suggest the following simple principle:

TRIA2. The overall value of an event E, occurring at time t1, for subject S = [(the sum, for every event E* caused by E, of the t1-relative interest value of E* for S) minus (the sum, for every event E* caused *not* to occur by E, of the t1-relative interest value of E* for S)], plus the intrinsic value of E for S.

The idea here is to incorporate into a single value three things: (i) the time-relative interest values of the good and bad things E brings about, (ii) the time-relative interest values of the good and bad things E prevents from coming about, and (iii) the intrinsic value of E itself (which, in the case of death, presumably turns out to be irrelevant). In stating TRIA2, we must choose a time to which to relativize the interests, and we must choose a way to determine the overall value of an event based on the values of its various consequences. The simplest way to determine overall values is to add and subtract values in obvious ways, much like in LCA; I don't think the arguments that follow rest importantly on this decision. The choice of a time to which to relativize the interests will be a point of contention shortly. But for now, let us tentatively take TRIA to be the conjunction of TRIA1 and TRIA2. Whether this is McMahan's view or not (and we will soon see reasons to think that McMahan's view is substantially more complicated than even this complicated view), it seems like a view worth considering.

Here is how TRIA works in a simple case. Suppose Jan dies at t1. Had Jan not died, she would have received a pleasant experience several years later. That pleasant experience would have had an intrinsic value of +10 for her. However, the psychological connections between Jan at t1 and Jan at the time of the pleasure would be so weak that the t1-relative interest value of the pleasure for Jan would have been only +5. As McMahan might say, it is almost as if the pleasure she would have gotten would be received by someone else. Thus, supposing that this one experience of pleasure is the only good or bad experience Jan would have received had she not died, the overall value of Jan's death for her, according to TRIA2, is –5. If TRIA were true, Jan's death would be only half as bad as it would be if LCA were true.

One possible source of confusion needs to be addressed. Sometimes McMahan writes as if he is not giving an account of the badness of a death for a person *simpliciter*, but rather an account of the badness of a death for a person *at a time*. Consider, for example, the following passage: "When the prudential unity relations that would bind an individual to himself in the future would be weak, death matters less *for that individual at the time*" (McMahan 2002, 172, emphasis his). This suggests that TRIA does not say anything about how bad death is overall for someone *simpliciter* (i.e. not relative to any time).[28] Thus Broome says, understandably, that we must think of TRIA as giving an account of "badness for the person, relative to that particular time" (Broome 2004, 249). Were this the correct way to understand TRIA, it might forestall some of the criticisms to come. In many passages, however, McMahan makes no mention of relativity to a time; for example: "Notice, however, that it seems reasonable to want one's own death to be less bad – to be a lesser rather than a greater misfortune" (McMahan 2002, 172). Here there is no mention of wanting the death to be less bad at some particular time, such as when it occurs. I think TRIA is best construed in the second way – as making a claim about non-time-relative overall values of events for people.[29] The time-relativity McMahan really seems concerned about comes in TRIA1, where the time-relative interest value of an event for a person is determined on the basis of when that event occurs, not in TRIA2, which gives the overall value of an event for a person *simpliciter* based on the time-relative interest values of some other events. If TRIA were merely making a claim about the time-relative overall values of events for people, it would not be a competitor to LCA at all, and a large portion of McMahan's book would be an interesting and provocative *non sequitur*.

Assume that a three-month-old baby is not strongly connected psychologically with its adult self. We can get TRIA to yield the result that Student's death is worse than Baby's, provided that the multiplier for psychological unity is sufficiently small in the case of Baby. Assuming that the level of well-being in each life is fairly constant throughout, the multiplier would have to be roughly .75 or lower, given that Student is deprived of about three-fourths of the amount of a good life of which Baby is deprived. TRIA also gets the results McMahan wants in *The Cure*, since the patient would not be very strongly connected psychologically pre- and post-treatment. And TRIA entails that it is not very bad for a person to die just after coming into existence, since at such an early stage, the person would have few psychological connections with his future adult self.

However, TRIA must be rejected. Consider this example. Suppose my infant son, Son, has a trust fund that he may use when he turns 25. I intend to drain the fund secretly, and prevent Son from ever finding out about its existence. Let t1 be the time when Son is three weeks old; let t2 be the time when he is 23 years old. Does it matter whether I steal the money at t1 or t2? It is very hard to see how it could possibly matter, but if TRIA is true, the harm of the theft depends on when it occurs. Were I to leave the money alone, Son would have had a much better life upon turning 25; the theft would deprive him of some goods, and the time-neutral intrinsic value of those goods for him would be the same no matter when I steal the money. But by TRIA1, the t1-relative interest value of those goods for Son would be less than the t2-relative interest value of them for him, given that the psychological connections between his three-week-old self and his 25-year-old self are much weaker than those between his 23-year-old self and his

25-year-old self. So by TRIA2, we get the result that the early theft would be less bad for Son than the later one. That is clearly wrong.[30]

Furthermore, TRIA will face problems with events that occur before a person exists yet still manage to harm the person. Consider a version of the trust fund case in which the trust fund for Son exists, and is drained by me, *before* Son exists. TRIA seems to entail that my draining the fund is not harmful to Son at all, since there are *no* psychological connections of any sort between Son in his twenties and Son before he exists. (There isn't even physical or functional continuity between parts of my son's brain at those two times, since he lacked a brain when he didn't exist.) But of course it is harmful to him.[31]

McMahan is aware of this sort of problem. He discusses a similar argument against TRIA: "If the Time-Relative Interest Account implies that it would be less objectionable to cause a person to be sterile by injuring him when he was a fetus than to cause him to be sterile by injuring him later, say at age fifteen, this is seriously damaging to the account's credibility" (McMahan 2002, 282).[32] He gives this response to the argument:

> In short, the objection assumes that, if we are to explain why the infliction of prenatal injury is wrong by reference to the victim's time-relative interests, we must focus on only those time-relative interests that the victim has prenatally, at the time the act is done. But that restriction is arbitrary; we must evaluate the act in terms of its effect on all those time-relative interests it affects, present or future. (McMahan 2002, 283)

If we apply this reasoning to the trust fund case, the idea seems to be that the early theft is just as harmful as the later one, since the early theft, like the late one, frustrates Son's time-relative interest in having the money *at the time he would have received it*, and that time-relative interest is relevant to the evaluation of the theft.

This response suggests that TRIA, as formulated, does not capture McMahan's view. I have assumed, in TRIA2, that in order to determine the value for a person of an event that takes place at time t1, we look at the *t1*-relative interest values for the person of all the goods and evils caused and prevented by the event. I made this assumption because McMahan explicitly says that when someone *dies* at time t1, it is the t1-relative interest values of the goods and evils prevented by the death that are relevant to evaluating the death. But this seems to be only because of special features of death cases. McMahan is now suggesting that in other cases not involving death, there are times t, distinct from t1, such that the t-relative interest values of those goods and evils are also relevant. What is the guiding principle here? Is there a principle that gives us the results McMahan wants – that Baby's death is less bad than Student's, that it is rational to refuse the cure, and that stealing Son's trust fund is equally bad for Son no matter when the theft occurs? That is, in determining the value of an event, to what time(s) do we relativize the values of the goods and evils it causes and prevents? Which time-relative interest values are relevant?

Recall that McMahan claims that "we must evaluate the act in terms of its effect on *all those time-relative interests it affects*, present or future" (McMahan 2002, 283; my emphasis). McMahan seems to be denying that we relativize to a single time; rather, we

relativize to *all* times. That is, the value of an event E is determined by the t-relative interest values, *for all t*, of the events caused by E. This suggests the following revision to TRIA2:[33]

TRIA2'. The overall value of an event E for subject S = [(the sum, for every event E* caused by E, *and every time t*, of the t-relative interest value of E* for S) minus (the sum, for every event E* caused *not* to occur by E, *and every time t*, of the t-relative interest value of E* for S)], plus the intrinsic value of E for S.

Let TRIA' be the conjunction of TRIA1 and TRIA2'.

TRIA' requires us to add a lot more numbers than does TRIA.[34] Perhaps it gets the right result in the trust fund case, since there will be many future times at which Son will have time-relative interests in getting the money. But it fails to get the results McMahan wants in *The Cure*. After receiving the cure, the patient would receive a lot of goods, and there would be many post-cure times *t* such that receiving those goods has a high *t*-relative interest value for the patient. So if TRIA' is true, the decision to receive the cure has a very high value for the one who gets it, and refusing the cure is bad overall.

We might be able to distinguish *The Cure* from the trust fund case by adopting a version of *actualism* about interests. The general idea would be this. When one decides to refuse the cure, one is deciding not to have one's interests changed in a certain way. The interests one would have had, had one taken the cure, *are not one's actual interests*. Hence, their counterfactual satisfaction or frustration does not affect the value of one's actual choice to refuse the cure. Whether the choice is a good one is determined only by

one's actual interests.[35] In the trust fund case, on the other hand, stealing the money frustrates some of Son's actual interests, and frustrates the same ones no matter when the theft occurs.

Distinguishing actual from possible interests is a tricky matter. Recall that interests, according to McMahan, are not desires; to say that something is in my interest is to say that it enhances my well-being. The problem is: how could it be the case that something actually fails to enhance my well-being, but would enhance my well-being if things had gone differently? Some things are only contingently good for me: doing crossword puzzles, for example, might actually be good for me, but not be good for me if I failed to enjoy it. But that is because doing crossword puzzles is only extrinsically good for me. It is good for me because it makes me happy. Death is not bad because it prevents us from getting extrinsic goods; it is bad because it prevents us from getting intrinsic goods. So the real problem is: how could it be the case that something is *intrinsically* good for me given one outcome, but not given another outcome? On many accounts of well-being, this idea is simply incoherent. Suppose hedonism were the correct account of well-being, for example; how could an experience of pleasure be intrinsically good for me *only contingently*? If hedonism is true, it is necessarily true.

There is one account of well-being that allows for the sort of actualism McMahan wants: the *object-version* of preferentism or desire satisfactionism.[36] According to this view, if someone intrinsically desires that P, and P obtains, P's obtaining is intrinsically good for that person. For example, suppose I intrinsically desire that the Yankees win the World Series. Then if the Yankees do indeed win, the fact that they win is intrinsically good for me; it is in my interest. But that fact is good for me only contingently, since,

had I not desired that they win, their winning would have failed to be intrinsically good for me. On this view, "interests" are to be understood as the objects of intrinsic desires, and the value of a state of affairs for me is *relative to a world*.

Given this account of well-being, and given the introduction of world-relative intrinsic value, we need to redefine time-relative interest value. I propose the following definition:

TRIA1*. The *t-relative interest value* of an event E for subject S *at world w* = the *w-relative* intrinsic value of event E for subject S, multiplied by a fraction corresponding to the level of psychological unity between S at t *at w* and S at the time of E. (In the case of merely possible events, the fraction corresponds to what *would have been* the level of psychological unity between S at t and S at the time of E had E occurred.)

And I propose the following account of the overall values of events for people:

TRIA2*. The overall value of an event E for subject S *at world w* = [(the sum, for every event E* caused by E, and every time t, of the t-relative interest value of E* for S *at w*) minus (the sum, for every event E* caused *not* to occur by E, and every time t, of the t-relative interest value of E* for S *at w*)], plus the *w-relative* intrinsic value of E for S.

Let TRIA* be the conjunction of TRIA1* and TRIA2*.

Here is how TRIA* works in *The Cure*. Add the following details to the story: I actually am a Yankees fan and do not care about hockey, but if I were to take the cure, I would become a hockey fan, and would want the New York Rangers to win the Stanley Cup. So if I were to receive the cure, it would be intrinsically good for me if the Rangers won, but were I to refuse, it would not be. Suppose I refuse the cure at time t_c. At time t_r ($t_r > t_c$) the Rangers win the Stanley Cup. The t_c-relative interest value of the Rangers' winning for me, *at the actual world*, would be zero. So I would, in essence, be deprived of nothing in virtue of the Rangers' winning, even though, if I had taken the cure, their winning would have been intrinsically good for me. According to TRIA*, whether refusing the cure is good for me depends on (i) how many other things would have happened that would have been in my *actual* interest, and (ii) the extent to which the values of those goods are discounted in virtue of the diminished psychological connectedness that would have obtained between myself at t_c and myself at the time I received the goods. Since many of the things that would have been in my interest were I to have taken the cure, like the Rangers' winning, are not in my actual interest, and since the discount for lack of psychological connectedness is high in this case, TRIA* seems to entail that refusing the cure is good for me. TRIA* also seems to get the right results in the trust fund case, since the same time-relative interests are frustrated no matter when the theft occurs.

But TRIA* faces new problems. It presupposes a view about well-being that very few people have ever held, whereas LCA is compatible with every theory of well-being. This is a major advantage of LCA.

More importantly, TRIA* yields some bizarre results. For one thing, it entails that whether an event would be good for someone might depend on whether the event occurs or not.[37] Consider *The Cure*. I explained how TRIA* seems to entail that refusing the cure would be good. But this was on the assumption that I actually refuse the cure. If I were to accept the cure, then my actual desires would be different. I would actually desire that the Rangers win the Stanley Cup, and so their winning would be intrinsically good for me. So it seems that TRIA* might well entail that, were I to (actually) receive the cure, my (counterfactual) refusal would have been bad for me. Thus, in *deliberating* about whether to take the cure or not, one important thing I will have to consider is *whether I will in fact take the cure or not*. I cannot evaluate which choice is better without knowing what I will in fact do. This is an unacceptable consequence of TRIA*.

This problem is a result of TRIA*'s actualist presuppositions. The problem could be solve by moving from actualism to *necessitarianism*. Instead of restricting the relevant interests to actual interests, or what is actually intrinsically good for someone, we would restrict the relevant interests to *necessary* interests; we would look at what her interests would be, what would be intrinsically good for her, *no matter what happens*.[38] But necessitarianism is little better than actualism. Suppose Jim has no interest in trying a certain sort of food, but Jane says to him: "Try some. You'll like it." Jim tries it, and likes it. As a result of his decision to try the food, he desires that he be eating that food, and since he's eating it, his eating it is intrinsically good for him. According to necessitarianism, his decision to try the food was not good for him at all. Had he chosen not to try it, he would not have had the desire to be eating it, so his eating it would not have been intrinsically good for him.

None of these modifications of TRIA results in a plausible view. There are many other ways one might modify TRIA, but it seems very difficult to find one that is principled, motivated, and gets the results McMahan wants. Either some such version of TRIA must be provided, or McMahan must abandon the examples he uses to motivate the move to TRIA, such as *The Cure*.

4. FINAL THOUGHTS

Surely there is some kernel of truth in the vicinity of TRIA that has attracted thoughtful philosophers. It does seem true that at any given time t, a person values more highly those goods that are in store for him at times when he will be closely psychologically connected to himself at t than those goods in store for him at other times. But it is unclear what follows from this fact. We might well wonder whether such a bias is rational, much as we wonder whether the bias towards the future is rational.[39] But even if this bias were rational, nothing would follow straightaway about the overall values of events for people. If the arguments of the previous section succeed, then we should deny that the overall value of an event for a person is determined by facts about psychological connectedness; so if the rationality of preferring the occurrence of some event is determined by facts about psychological connectedness, it follows that there is no close connection between the overall value of an event and the rationality of preferring (at some time) the occurrence of that event. How rational it is to prefer a given event will not be proportional to the overall value of that event. If so, then the mistake made by

McMahan and Belshaw is to suppose that how bad it is for a person to die is proportional to how rational it is for that person to care about her death at the time of its occurrence.

I have tried to show that LCA can get around the problems McMahan and Belshaw raise, and that views that are compatible with Student's death being worse than Baby's, such as TRIA, face serious problems of their own. I conclude that LCA is more plausible than TRIA, and Baby's death is worse than Student's. Of course, there might be some other view that could account for the view that Student's death is worse without encountering the sorts of problems detailed here. And I have discussed only two arguments against LCA. So I don't pretend to have provided a complete defense of these conclusions.[40] But I think I have at least given good reason to be skeptical about the possibility of developing a plausible view that entails that Student's death is worse.

REFERENCES

Belshaw, Chris. (2005) *10 Good Questions about Life and Death*. Malden, MA: Blackwell Publishing.

Benatar, David. (2006) *Better Never to Have Been Born: The Harm of Coming Into Existence*. New York: Oxford University Press.

Broome, John. (2004) *Weighing Lives*. New York: Oxford University Press.

Bykvist, Krister. (1998) *Changing Preferences: A Study in Preferentialism*. PhD dissertation, Uppsala University.

Carlson, Erik. (1999) "Consequentialism, Alternatives, and Actualism." *Philosophical Studies* 96: 253-68.

DeGrazia, David. (2003) "Identity, Killing, and the Boundaries of Our Existence." *Philosophy and Public Affairs* 31: 413-442.

Feit, Neil. (2002) "The Time of Death's Misfortune." *Nous* 36: 359-83.

Feldman, Fred. (1991) "Some Puzzles About the Evil of Death." *The Philosophical Review* 100: 205-227.

Feldman, Fred. (1992) *Confrontations with the Reaper: A Philosophical Study of the Nature and Value of Death*. New York: Oxford University Press.

Harman, Elizabeth. (2003) "The Potentiality Problem." *Philosophical Studies* 114: 173-98.

Harman, Gilbert. (2000) *Explaining Value*. New York: Oxford University Press.

Korsgaard, Christine. (1989) "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18: 101-132.

Lewis, D. (1973) *Counterfactuals*. New York: Basil Blackwell.

MacIntyre, Alasdair. (1981) *After Virtue*. London: Duckworth.

Marquis, Don. (1989) "Why Abortion is Immoral." *The Journal of Philosophy* 86: 183-202.

McMahan, Jeff. (2002) *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press.

Nagel, Thomas. (1970) "Death." *Nous* 4: 73-80.

Parfit, Derek. (1984) *Reasons and Persons*. New York: Oxford University Press.

Rabinowicz, Wlodek and Jan Osterberg. (1996) "Value Based on Preferences: On Two Interpretations of Preference Utilitarianism." *Economics and Philosophy* 12: 1-27.

Silverstein, Harry. (1980) "The Evil of Death." *The Journal of Philosophy* 77: 401-424.

Stalnaker, R. (1984) *Inquiry*. Cambridge, Mass.: MIT Press.

Velleman, David. (1991) "Well-Being and Time." *Pacific Philosophical Quarterly* 72: 48-77.

Williams, Bernard. (1973) *Problems of the Self*. Cambridge: Cambridge University Press.

[Deletion 1]

---

[1] Perhaps one will wonder: how could such facts be known? But the examples are merely hypothetical; there are no secret facts to be known about them. The details are stipulated. In any *actual* case, there would be a problem of knowledge; the best we could do would be to rely on statistics and actuarial tables to determine what sort of life it is most reasonable to suppose the victim would have had. But the topic here is not how we can *know* how bad someone's death is, but what *makes* it bad.

[2] David Benatar discusses a similar case, comparing a thirty-year-old and a fetus (Benatar 2006, 159). Based on what he says about that case, he might agree with McMahan and Belshaw. David DeGrazia endorses McMahan's approach as well (DeGrazia 2003, 433).

[3] For examples of deprivation accounts, see Nagel (1970), Silverstein (1980), Feldman (1991), Feit (2002), Broome (2004), and [Deletion 2].

[4] As in Feldman (1992, 150) and Broome (2004, 249). Throughout the paper I will use words such as 'value,' 'good,' 'bad,' and the like. In all such cases it should be

understood that it is prudential value that is at issue, even if this is not explicitly stated. The concern is how bad death is *for its victim*.

[5] See Lewis (1973) and Stalnaker (1984).

[6] [Deletion 3]

[7] Of course, the discontinuity could be relevant if it causally affected the intrinsic value of the life in some way, such as by making it less pleasant. This is supposed to be ruled out in the example by stipulation.

[8] McMahan also suggests that the weakening of prudential unity relations could be a misfortune in itself (McMahan 2002, 174). If what he means is that it is intrinsically bad for a person to have the prudential unity relations between her present and future selves weakened, then his suggestion can also be incorporated into LCA, providing yet another possible line of defense.

[9] [Deletion 4]

[10] [Deletion 5] see Korsgaard (1989), pp. 122-123.

[11] This consideration seems to be mitigated by the fact that it is the agent herself who is choosing to undergo the treatment and thereby have her desires changed.

[12] Thanks to an anonymous referee for suggesting improvements to this point.

[13] Another idea, which I will not pursue here, would be to say that while there is a single person who survives the cure, that person lives two lives. Perhaps this is the way Bernard Williams would describe the case; see Williams (1973), 92-94.

[14] Parfit also suggests that one cannot be harmed by being prevented from existing (Parfit 1984, 489).

[15] Kris McDaniel argues for this point in greater detail in an unpublished manuscript. The point is of course not uncontroversial; Williams seems to deny it, saying that we "cannot think egoistically" about such a situation (1973, 87).

Note that this strategy would not account for an *actual* case where an act of contraception prevents a person from coming into existence (as opposed to a case where we take some actually existing person and imagine a world in which that person never existed), at least not without some additional metaphysical assumptions; for in such a case, there is no actual person we can point to and say that not existing would have harmed *her*. Is the merely possible person who would have existed but for the contraception actually harmed? Our options here depend on whether there are merely possible objects or not. Either way, there is no problem for LCA. First, suppose there are some merely possible people. Then we can say, of the merely possible person who would have existed but for the contraception, that it is just as bad *for her* to be prevented from coming into existence as it is for her to die shortly after coming into existence. There is no great disparity between the two harms. On the other hand, suppose that there are no merely possible objects. Then there is no subject of harm here, but there is still no problem for LCA, because there is no person such that LCA entails that it is a good thing for that person that she was prevented from coming into existence rather than dying shortly after coming into existence. Again, there is no great disparity between harms (in this case because there are no harms). So we don't get an undesirable result for LCA whether we admit merely possible objects or not.

[16] The view that how much a harm matters depends solely on its magnitude is endorsed, for example, by Donald Marquis in his famous paper on abortion (Marquis 1989, 194).

[17] Feldman (1992) endorses a view like this.

[18] Against this view, see Harman (2003).

[19] See his discussion of the "Intrinsic Worth Account" and the "Two-Tiered Account" of the wrongness of killing, and his discussion therein of the ideas of a "threshold of equal worth" and a "threshold of respect" (McMahan 2002, 243-65).

[20] [Deletion 6]

[21] Thanks to an anonymous referee for the example.

[22] McMahan takes both the investment facts and the narrative facts to be relevant to the value of death (McMahan 2002, 183). For more on narrative unity see Velleman (1991) and MacIntyre (1981).

[23] McMahan agrees that a person's desires at the time of his death are relevant to the value of his death (McMahan 2002, 183). But he says that other factors are relevant as well.

[24] This argument is analogous to one given by Marquis in discussing the morality of abortion, though Marquis does not directly address the badness of death (1989, 198).

[25] Compare to Parfit (1984, 284-5).

[26] This is in contrast with Parfit, who says "it would not matter if my brain was replaced with an exact duplicate" (Parfit 285).

[27] [Deletion 7]

[28] [Deletion 8]

[29] In conversation McMahan confirmed that this is indeed the way he was thinking of TRIA.

[30] Broome gives a somewhat more complicated counterexample that seems to target the same feature of McMahan's view (Broome 2004, 251).

[31] See Feldman (1991, 219) for a defense of the claim that an event occurring before a person exists can harm that person.

[32] One potential complication here is that McMahan is discussing the wrongness of inflicting harm, rather than the badness of the occurrence of an event.

[33] An anonymous referee suggested a view along these lines as an interpretation of McMahan.

[34] For the purposes of this paper I suppose that time is discrete, since if time were continuous, TRIA' and its descendants would face devastating problems involving infinite value; even ordinary cases would involve adding and subtracting infinities.

[35] In personal correspondence McMahan has suggested that this is his view.

[36] See Rabinowicz and Osterberg (1996) and Bykvist (1998, 64-65).

[37] This problem mirrors one for certain versions of consequentialism pointed out by Erik Carlson (1999, 256-257). Carlson attributes the point to Wlodek Rabinowicz.

[38] In personal correspondence McMahan suggests he endorses necessitarianism.

[39] See Parfit (1984, Ch. 8) for the classic discussion of this question.

[40] [Deletion 10]