Utility Monsters for the Fission Age

Introduction

At least since Parfit 1984, there have been connections drawn between puzzle cases about personal identity, on the one hand, and normative questions, on the other. The relevant cases are hypothetical cases involving apparent fission of one person into multiple people (through separation of brain hemispheres, teletransportation accidents, and so on), fusion of multiple people into one, slow changes in one organism that yields a different person at the start than at the end, and suchlike. Questions that arise include the following. What is the practically rational thing to do, from the standpoint of a person facing the prospect of splitting into two people? Should the products of such fission be held morally responsible for the actions of their progenitors? How does fission bear on special duties created by family ties, friendships, and personal debts?

However, we think reflection on puzzle cases from the personal identity literature impacts on moral theorising in another important way. To date, much of the focus has been on potential consequences for moral and practical questions focused on obligations or duties of people in particular personal relationships: how should I think about *my* future welfare how might my obligations differ between *my* friends and family and their qualitative clones, how should we distribute moral rewards and punishments, praise and blame, given that intuitively these should go to the *same person* who performed the relevant earlier actions. (See, for instance, Williams 1981, Unger 1992, Jeske 1993, Eklund 2004, and Shoemaker 2007.)

We claim that the metaphysics of identity also has implications for impersonal moral requirements: that is, obligations to people regardless of whether they happen to share a particular connection with

the agent. We aim to show that on one popular account of what is going on in cases of fission, fusion etc., put together with two popular normative accounts of what (or who) matters for moral action and how what matters should be aggregated, yields what seems to be an unacceptable moral *reductio*. While we discuss ways of responding to this clash in the final section of this paper, the cases we consider raise the prospect that, given one of the common stances in the metaphysics of personal identity, some common (plausible, frequently defended) fundamental normative commitments are not jointly tenable, and stand in need of revision.

1. Multiple Occupancy

We consider the moral implications of:

> **Multiple Occupancy** Two or more distinct, temporally extended persons can completely overlap for a span of time.

Versions of Multiple Occupancy are defended by David Lewis 1976, Denis Robinson (Robinson 1985), from whom we get the label, Eugene Mills (Mills 1993), David Braddon-Mitchell and Caroline West (Braddon-Mitchell and West 2001), John Perry (Perry 2002), Harold Noonan (Noonan 2003), and Simon Langford (Langford 2007).

One common reason for endorsing multiple occupancy is to address puzzles of fission and fusion: cases where what is apparently one person at one time becomes two at another, or two at one time apparently merge to become one. Many paradoxes of fission, for example, can be addressed if there were two people there all along before fission, sharing all their parts: then the two later people can each be identical to an earlier person, without any failure of transitivity of identity. Multiple

Occupancy can be of use in addressing fission and fusion puzzles associated with gaining and losing parts as well: one response to the ancient puzzle of Dion and Theon, where Dion loses his foot and arguably becomes identical to the pre-existing object which was Dion-less-that-foot (Theon), is to say that after the foot is lost both Dion and Theon continue to exist, but now Theon, instead of being a large part of a person, is a person in his own right (albeit one that shares all his parts with another person, Dion).

One way to endorse Multiple Occupancy is to claim that people persist by having different temporal parts at different times, which we will call *person-stages*, and that one person-stage can be a part of more than one person. According to this framework, what it is for two persons to overlap for a span of time is for them to share all of their temporal parts within that span, so that every person-stage located in that span which is part of one, is also part of the other.

While Multiple Occupancy fits naturally with an ontology of temporal parts, however, it can be endorsed by those who reject the "perdurantist" view that people have different temporal parts at different times. It is open to *endurantist* theories to endorse Multiple Occupancy (see Robinson 1985): two enduring objects may be able to share all their parts at a particular time while being distinct; or one could adopt the view that two distinct people can both be *constituted* by the same underlying physical material, even if they do not share some of their parts. The metaphysical differences between perdurantist and endurantist Multiple Occupancy are not important for present purposes: while we will continue to talk of person stages for convenience, endurantist partisans of Multiple Occupancy should reinterpret this in more endurantist-friendly language.

Lewis 1976 invokes Multiple Occupancy in fission cases, where (apparently) one person is divided into two, and also in fusion cases, where two persons (apparently) fuse into one. He begins with

Derek Parfit's concept of an *R-relation*: a placeholder for whatever relation ultimately matters to survival. Candidates for the R-relation include continuity of personality, connectedness by links of quasi-memory and quasi-anticipation[1], and physical continuity, but we do not need to settle what the R-relation consists of here. The important point is that it need not be one-one.

A person, says Lewis, is a maximal aggregate of R-interrelated person stages: an aggregate of stages such that each is R-related to each of the others, and which is not a proper part of any larger such aggregate. In the case of fission, there are two such aggregates that share an initial part; in the case of fusion, there are two such aggregates that share a final part.

Braddon-Mitchell and West suggest another reason to accept Multiple Occupancy. They argue that what counts as an R-relation is partly *conventional*: the survival and identity conditions of persons depend partly on their own attitudes, such as accepting responsibility for the actions of particular person-stages or anticipating the experiences of others, and partly on the practices of the societies they inhabit, such as the transmission of property rights from one person-stage to another or the punishment of one person-stage for another's misdeeds.[2] In other words, what counts as a continuant person depends on what gets *treated* as a continuant person: for every reasonable set of conventions, there is a different R-relation. ("R-related" thus picks out a role property with many potential realisers.) A person is a maximal R-interrelated aggregate of person stages, where the realiser of R is fixed by the practices of the person and those around her.

But what happens if someone changes her practices? Braddon-Mitchell and West (2001 pp 62-64) imagine the case of two societies: the Somataphiles, who view teletransportation as a kind of death,

---

1 Quasi-memory and quasi-anticipation are the states that constitute memory and anticipation in cases where the person-stage doing the remembering or anticipation is part of the same person as the person-stage that is remembered or anticipated, but may relate person-states that are not part of the same person.

2 For more arguments in favour of the view that the persistence conditions of persons are conventional, see Braddon-Mitchell and Miller 2004.

and the Teletransporters, who view it as a fast and convenient means of travel. Suppose a Somataphile converts to Teletransportation before stepping into a teletransporting machine for a vacation at the Sunshine Coast. Does this person survive teletransportation, or not?

Braddon-Mitchell and West claim that there are two overlapping people, one of whom (being a Teletransporter) survives teletransportation, and another of whom (being a Somataphile) does not. More generally, a person is a maximal X-interrelated set of person-stages, where X is a potential realiser of the R relation, and at least one of person-stages adopts practices that make X the realiser R relation. So in the case at hand, one person is the maximal aggregate of person-stages connected by what we might call the T-relation—the candidate R-relation picked out by Teletransporter practices. Since some of the later person-stages adopt Teletransporter practices, there is a person composed of all the T-interrelated stages related to them. This person survives teletransportation to the Sunshine Coast. Another person is the maximal aggregate of person-stages connected by what we might call the S-relation—the candidate R-relation picked out by Somataphile practices. Since some of the earlier stages adopt Somataphile practices, there is a person composed of all the S-interrelated stages related to them. This person perishes upon being teletransported.

2. Normative Assumptions

Let us adopt, as a theory of impersonal moral requirements:

> **Maximising Consequentialism** One morally ought to do whatever will result in the greatest wellbeing for everyone.

Maximising Consequentialism admits of a further key distinction: should one maximise *total*

wellbeing (Total Maximising Consequentialism), or *average* wellbeing (Average Maximising Consequentialism)? It turns that on both versions of the view, Multiple Occupancy causes trouble. The exact type of trouble will depend on the exact type of Maximising Consequentialism; we will flag these issues as they arise. (While we will focus on maximising views for simplicity, the cases we will discuss cause trouble for other distributional principles as well, in ways that are straightforward to work out.)

In tricky cases of fission and fusion, we must draw another distinction. Should we aim to bring the greatest aggregate wellbeing to persons, or person-stages? We will restrict our attention to versions of Maximising Consequentialism that give the former answer. Thus, we favour:

> **Count the Happy People** The aggregate of wellbeing in the world is given by taking the welfare of each person, and aggregating those levels of welfare.

That is, we should count the people and their welfare rather than, for instance, person-stages, or parts of people. Count the Happy People will be particularly tempting to philosophers who, with Slote 1982 and Velleman 1981, believe that the value of a life cannot be reduced to the values of its individual parts.[3]

We are simplifying a little by restricting our focus to people: in a less simplistic account, the welfare of animals that do not have person status, or other entities besides animals and people, may play a role as well. Again for simplicity, we will neglect these potential components of overall

3 Some consequentialists hold that only valuable states matter: JJC Smart, for example, famously claimed "A utilitarian has to think of people rather as buckets into which happiness can be poured" (Smart 1991 p 361). Even Smart-style consequentialists have to decide how to aggregate across buckets: and "counting the happy buckets" can be contrasted with other measures, including the analogue of measuring the welfare in each bucket-stage. (Other consequentialists will reject Smart's attitude: there is nothing in consequentialism per se that holds that the value of people derives from their states.)

welfare, but not because we are indifferent to the plight of kittens or other-worldly world-spirits.

Together with Maximising Consequentialism and Count the Happy People, Multiple Occupancy yields strange moral consequences in a variety of cases.

3. Many-Headed Utility Monsters

Suppose a group, consisting of six person-stages at the time of the story, is deliberating about the best way to divide a delicious cherry pie. Each member of the group would derive considerable enjoyment from eating 1/6 of the pie. More pie would give more enjoyment, but, due to the diminishing marginal utility of pie, each subsequent slice would make only half the difference of the previous slice. (Suppose eating no pie provided 0 utiles and eating 1/6 of the pie provided 32 utiles. Then eating 1/3 of the pie would provide 48 utiles, eating 1/2 of the pie would provide 56 utiles, and so on.) Assume that the pie consumption has no relevant long-term consequences beyond immediate pleasure: nobody is going to gain enlightenment through delicious experience, or contract diabetes, no matter how much of this one pie they eat. Since the members of the group are good Maximising Consequentialists, they agree to divide the pie six ways.

“But wait!” one of the group members pipes up, “I have a fission appointment scheduled for next week, when the doctors will divide me into one hundred separate persons. (Don't be misled by figures of speech, like `divide' and `I'; thanks to my future fission, you are currently speaking with a hundred overlapping persons who will one day lead separate lives.) Are any of you planning to undergo fission? No, I thought not. Then you should give your pie to me. By foregoing one slice of pie, you can provide extra happiness for a hundred people!”

What should the other five do, as good Maximising Consequentialists? They seem bound to give

up their pie, at least if they are also committed to Count the Happy Persons. Giving up the pie means a net loss of 160 utiles for the five—more than enough to outweigh the 31-utile gain that would result from giving the whole pie to one person, but tiny in comparison to the 3100-utile gain that would result in 100 people each getting to enjoy an entire pie. Those are the calculations relevant for Total Maximising Consequentialism. If instead they are Average Maximising Consequentialists, seeking to create the highest average wellbeing, the verdict for pie-distribution is the same. 3100/105 (=29.52) easily outweighs 16, the average if the pie is split six ways.

Notice that the verdict here isn't based on any future benefits that might accrue to the hundred after fission. Even if the pie is digested and forgotten by the time the fission takes place, the consumption of the pie still brings enjoyment to 100 different people.

What goes for pie goes for other goods. This seems counter-intuitive: there seems a good sense that, at the time, six people are splitting a pie that they all deserve equally, but one gets it all and the others get nothing.

This case is like one that Robert Nozick mentions (Nozick 1974 p 41) as a problem for consequentialism: the *utility monster*. A utility monster derives much more good from resources than the rest of us: since a utility monster would derive a hundred times more utility from 1/6 of the pie, the consequentialist verdict seems to be that the entire pie should be given to the utility monster. This, Nozick argues, violates intuitions we have about morally appropriate distributions.

In our case the verdict that the "person" to undergo fission deserves more utility (or desire-satisfaction, or resources, or what-have-you) is not because any one person is privileged over another: rather, where it appears to be 6 people sharing a pie, it is in fact 105: and the 100

outweigh the 5, in a way familiar to consequentialists. Still, insofar as we are inclined to think that future fission should not entitle a stage to more pie now, as opposed to otherwise similar stages now, the pie distribution apparently endorsed by the combination of Maximising Consequentialism, Multiple Occupancy, and Count the Happy People seems on the face of it unfair, immoral and absurd.[4]

4. Infinite Overlap?

Things get even worse when we consider David Lewis's Methuselah, who lives for 969 years, undergoing psychological change at the same rate as an ordinary person (Lewis 1976 pp 30). It takes exactly 137 years for Methuselah's memories, ambitions, and personality traits to change completely, so that by the time Methuselah is 200, nothing is left of the 63-year old who preceded him. Some views of personal identity require that the R-relation be transitive, but setting these views aside, we can plausibly suppose that two stages of Methuselah are R-related if and only if they are within 137 years of each other. Thus, every 137-year long aggregate of Methuselah-stages will form a distinct person, since it will be a maximal aggregate of R-related person-stages.

Consider now a stage of Methuselah located at midnight on his 300th birthday. This person-stage is part of many distinct, 137-year long persons: the one stretching from year 250 to year 387, the one stretching from year 193 to year 310... and so on. Thus, whatever happens to the 300th-birthday stage of Methuselah happens to many people. If space and time are infinitely subdivisible, it

4 Zimmerman 2002 and Olson 2010 raise related problems for ontologies where each person overlaps one or more non-people, and where those non-people have features we consider morally important in people. Zimmerman considers the ontology of Baker 2000, who posits that each person is constituted by a person-shaped hunk of matter; Olson focuses on a four-dimensionalist ontology, which posits long-lived temporal parts of people, and aggregates of those parts, which are not themselves people. Both authors frame the problem as one of how to discount the interests of the apparently person-like entities as derivative of the interests of people. In our problems, solutions that discount the interests of one or more entity are off the table. Our Multiple Occupancy principle says that one or more *persons* can overlap at a time—not merely that one or more *entities that superficially resemble persons* can overlap at a time. Unlike Zimmerman and Olson's cases, our cases are symmetric, so that all the overlapping persons involved have an equal claim to moral consideration.

happens to infinitely many people.

This is already enough to cause trouble for Total Maximising Consequentialism. Infinite populations paralyse Total Maximising Consequentialism even in the face of simple questions. For instance, if you can make Methuselah's 300th-birthday stage happy by bringing him a surprise birthday cake, and the cake costs you nothing, should you bring it? On the one hand, it seems that you should: you will provide infinitely many people each with the same finite increase in wellbeing, and therefore generate an infinite increase in wellbeing. On the other hand, it seems that you needn't. If all of Methuselah's lives are already worth living, the people who overlap Methuselah's 300th stage already have infinite wellbeing in aggregate. If you were to bring the cake, they would still have infinite wellbeing in aggregate, and therefore, collectively, they would be no better off than before.

The trouble is that traditional decision theory is not equipped to aggregate finite utilities distributed across infinitely many locations. However, there is a growing literature in infinite decision theory. How does this literature fare when against our monstrous Methuselah? Let us set aside hard problems involving uncertainty and variation in the domain of persons and begin with a simple question. Suppose we hold fixed Methuselah's lifespan and the degree of R-relatedness between his stages. Then how can we maximise the collective welfare of all the overlapping persons who compose Methuselah?

In the cake example, there is a simple answer. Since bringing the cake makes everyone better off, and makes no one worse off, a weak dominance principle gives you a reason to bring the cake. But now consider a variant on the cake example. Suppose that eating the cake will bring enjoyment to anyone who eats it, but will cause severe indigestion the next morning. The indigestion is twice as

bad as the cake is good: anyone who eats the cake will gain 10 utiles, but anyone who experiences the indigestion will lose 20 utiles. Now should you bring the cake? Notice that some of the guests at Methuselah's 300th birthday party will perish before morning, and these guests will enjoy the benefits of the cake without the costs. Notice that other people will turn up too late for cake, but in time to suffer indigestion. So bringing the cake will benefit some guests and harm others, and the dominance principle falls silent.

A proposal by Vallentyne and Kagan (1997) for extending the dominance principle in infinite cases appeals to the convergence properties of various infinite sums. (For similar proposals, see (Lauwers and Vallentyne 2004), (Atsumi 1965), and (Von Weizsäcker 1965).) Consider two worlds, w1 and w2, containing exactly the same locations—in this case, persons who are part of the 969-year continuant object Methuselah. Begin with a sequence of finite sets of locations such that each set is a superset of the previous one. (Intuitively, each set contains all the individuals in the previous set and a few nearby ones.) For each *S* in the sequence, there will be a number measuring the total wellbeing of the members of *S* in w1, and another number measuring the total wellbeing of the members of the set in w2. The difference *dS* between these two numbers measures how much better off members of *S* are collectively in w1 than in w2. If for all suitable sequences, the *dS*es converge to a positive limit as the sets expand, then w1 is better than w2.

Some authors criticise Vallentyne and Kagan's version of the proposal on the grounds that it grants too much significance to the spatiotemporal locations of persons—rearrange the persons in a world, and you are liable to get different verdicts about how that world compares to others (Hamkins and Montero 2000; Mulgan 2002). But in the case of Methuselah, there is a much more serious problem. If time is continuous, countably infinite sums will be useless, since the guest list at Methuselah's birthday will be uncountably long. Furthermore, the most natural way of lining up the

guests, by age, will produce a dense ordering, destroying any hope of a weakly converging sum (as if the crowded guest list was not already bad enough). Infinite decision theory does not even get started.

The prospects are slightly brighter for Average Maximising Consequentialism. We can take an average of infinitely many—even continuously many—points, so long as there is some measure over them. In Methuselah's case, we might again line the birthday guests up from oldest to youngest. We might then assign the measure according to anniversary of `birth', so that the group of persons created between t and t' years old and the group of persons created between s and s' years old have equal measure whenever $|t-t'| = |s-s'|$. Alternatively, we might line the guests up according to anniversary of `death', so that the group of persons destroyed between t and t' years after Methuselah's birthday and the group of persons destroyed between s and s' years after Methuselah's birthday have equal measure.

In Lewis's example of Methuselah, both ways of measuring come to the same thing. But in other versions of the example, they needn't. Methuselah's rate of psychological change might slow down as a constant function c of age, so that persons created within n years of each other are destroyed within e^{cn} years of each other. If there is more than one possible measure over the different Methuselahs, how are we to fix on one?

Even if we can fix on a single, natural measure for the Methuselahs, there are more troubles for Average Maximising Consequentialism. Suppose two short-lived guests, Bill and Ted, attend Methuselah's birthday party. Bill and Ted have a natural lifespan of threescore and ten years apiece—a time span too short to allow the type of change the Methuselah undergoes. All the Bill-stages are R-interrelated, and all the Ted-stages are R-interrelated, so that Bill and Ted comprise

only one person each. Since Bill and Ted are only two persons among infinitely many, it seems that the set {Bill, Ted} ought to have measure zero. After all, all the finite sets of persons who share a body with Methuselah have measure zero. Why should Bill and Ted get special treatment?

But if the set {Bill, Ted} has measure zero, then Bill and Ted's welfare does not count at all toward the average wellbeing. Bringing Methuselah a diamond on his birthday would bring joy to a set of persons with positive measure, and so increases the average wellbeing. If the diamond is available only at a terrible price to Bill and Ted—years of miserable, under-paid toil in the mines—no matter. So long as Methuselah gets even a tiny amount of joy from the diamond on his 300th birthday, this enjoyment will make up for any amount of torture we can inflict on Bill and Ted.

On some theories of the R-relation, Methuselah cases will be ruled out. Let person-stages *a* and *b* be *forwards connected* just in case the links of memory, anticipation, physical similarity, and psychological similarity which ground the R-relation obtain between *a* and *b*, and a is earlier than *b*. And let *a* be *backwards connected* to *b* iff *b* is forwards connected to *a*. One might then say that the *forwards R-relation* was the transitive closure of forwards connectedness, the *backwards R-relation* was the transitive closure of backwards connectedness, and two stages were R-related iff one was forwards R-related to the other or backwards R-related to the other. Imposing this kind of transitivity on the forwards and backwards R-relation rules Methuselah's case out as impossible. Methuselah's 10-year old stages are forwards R-related to his 110-year old stages, and his 110-year old stages are forwards R-related to his 210-year old stages; thus, it cannot be the case that, as stipulated in the example, the forwards R-relation fails to obtain.

Denying the possibility of Methuselah does not save maximising consequentialism from paradoxes flowing from infinite multiple occupancy, however. Methuselah makes the point about infinite

multiple occupancy nicely, but we can have infinite multiple occupancy even if the metaphysics of personal identity somehow rules out this case. Infinite fission seems possible. We do our first teletransportation or brain hemisphere transplants after one minute. Our second, to each of the products, after a further half minute (perhaps after some quick regeneration if we are producing fission through brain transplants). Our third splitting, to each of the 4 products of the previous procedure, after a further quarter of a minute... at the end of two minutes, we have infinitely many copies. (We leave the logistical details of where to store everyone and all the equipment to the imaginative reader—feel free to use spaceships heading in different directions in infinite space, alternate dimensions, or whatever other thought-experiment technology you need.) If Anna is about to go through such a Zeno-fission series, she will cohabit with countably infinitely many others: not quite as many as Methuselah, but enough for a serious utility-monster problem. On most ways of summing utility, keeping Bill and Ted in miserable suffering to confer a trivial benefit on Anna (and her co-occupiers) might well end up the maximising thing to do. (More exotic cases of fission into more-than-countably-many people can be dreamed up, but we will resist the temptation to elaborate.)[5]

5. A New Repugnant Conclusion

In addition to many-headed utility monsters, which create obligations concerning the distribution of resources based on fixed facts about fission, we can create a version of Derek Parfit's Repugnant Conclusion (Parfit 1984 chapter 17). In Parfit's example, Total Maximising Consequentialists are

5 There will also be problems of massive overlap, and plausibly infinite overlap, if "many" solutions to the problem of the many are endorsed: that is, if we respond to the puzzle that there are many person-like aggregates where I am with some claim to be my parts by saying there are in fact many people where I am, differing slightly in their parts. (See Lewis 1993.) There are many versions of the problem of the many, including modal ones (perhaps there are as many people here as there are functions from worlds to occupied locations of objects in those worlds that map the actual world to this location). If we give a "many" response to the modal problem of the many, there will be a high infinity of multiple occupiers whenever we ordinarily think there is one person. Even temporal and spatial problems of the many can motivate the view that there are continuum-many people wherever I am. Moral views that have trouble with infinities of people, especially multiple occupancy people, have a number of hostages to metaphysical fortune.

forced to conclude that a world with a few extremely happy people is worse than a world with a great many people, all of whose lives are barely worth living. The Parfit example has terrifying consequences for population ethics: if the Total Maximising Consequentialist is right, then we are obligated to have more children even in cases where doing so will make everyone else worse off. In our example, if Total Maximising Consequentialism is right, then we are sometimes obligated to promote fission—even at significant expense to human happiness.

Suppose you are nearing the end of a full and happy life; your doctor has determined that you have a week to live. The hospital where you spend your dying days also houses a fission machine, once considered promising in anti-ageing research. Sadly, that research program was a failure: instead of prolonging life, the machine merely caused chronic pain in all of its subjects. To be more precise: products of fission enjoyed the same life span that the original person would have enjoyed, had no fission occurred, but experienced more pain than the original person would have experienced, had no fission occurred. Every subsequent pass through the machine resulted in slightly more severe chronic pain for the resulting people.

Your doctor, the former P.I. of the fission research program, projects that you have one more week to live. She would like to put you through the machine, just for old times' sake. But before she does, she needs to clear her decision with the hospital's ethics committee. What should the ethics committee conclude?[6]

Let us add a few more details. If fission occurs, the two resulting people will each be worse off than the person who would have existed, had fission never occurred. However, they will not be too much worse off; the bad of the pain outweighs less than half the good of the good of their pre-

6 If you think that the fact that this is your *doctor* should make a difference, due to special duties doctors bear to patients, change the example so that the researcher has no special personal connection to the patient.

fission lives. Similarly for subsequent fissions: the fission products will be a bit more than half as well off as the alternative individual who never undergoes fission.

If the ethics committee are Total Maximising Consequentialists, they should conclude that the doctor is morally obligated to put you through fission. Furthermore, once she's done that, she is morally obligated to repeat the procedure on each of the resulting people. She is obligated to create the procedure as much as she can, up to the point where the extra people perish in such agony that the pain of their torture nearly outweighs the accumulated good of the rest of their lives.

Parfit's Repugnant Conclusion is bad for Total Maximising Consequentialism. But in one respect, our Repugnant Conclusion may even be worse. In Parfit's example, there are some people for whom the supposedly better world—the one full of people whose lives are barely worth living is—is arguably better than the alternative. Those people exist with a life of positive utility rather than fail to exist, and so arguably do better existing than not. In our example, every person is worse off in the supposedly better world than they would have been in the supposedly worse world with no fission. For the one, unsplit person in the world without fission is plausibly a counterpart of each of the fission products in the world with plenty of fission: if this is the way to understand the case, each of them would have been much better off if no fission had occurred, since each of them would have been the one, unfissioned, person.

6. How Should Maximising Consequentialists Respond?

While the problems we have raised can be avoided by rejecting any of Multiple Occupancy, Maximising Consequentialism, and Count the Happy People, we will focus in this section on what the most appealing responses to the problem might be for those interested in maintaining Maximising Consequentialism. Not that we think that this premise is compulsory either: but since

problems similar to those that arise for maximising consequentialists will beset many other ethical theories if they accept Multiple Occupancy and Count the Happy People, we think the most interesting ways to avoid the sorts of problems we raise require questioning the other two assumptions.

The intuitive problem with many-headed utility monsters is that fission leads us to double-count overlapping persons. Therefore, the Maximising Consequentialist might seek solace from Lewis's solution to a problem involving counting. We describe fission as a situation in which *one* person becomes *two*. In the pie case, where there are six time-slices dividing up the pie, it seems that the pie is divided among six people and not among 105. How is the defender of Multiple Occupancy to account for this data?

Lewis argues that at any given time t, we do not individuate persons according to identity, but by the weaker relation *identity-at-t*—roughly, total overlap of temporal parts. Two persons, P1 and P2 are identical at t iff each has a stage at t, and all and only P1-stages located at t are P2-stages located at t. Since all the fission products have a single stage at the time when the pie is divided, we count them as only one person, rather than one hundred. In order to determine how many people there are at any time t, we count according to identity at t.

Might the Maximising Consequentialist use Lewis's method of counting for weighing up harms and benefits? We fear not. Lewis's method is useful for counting at a time, and we can use it to determine how much aggregate wellbeing there is at any given time. But presumably, the Maximising Consequentialist wants to make the world better—not better at this or that time, but better simpliciter. To focus on making the world better only at a time seems like a hopelessly parochial attitude with no place in a fundamental moral theory. A Maximising Consequentialist

should surely care about acting to promote the good across all times she can affect, not some privileged moment. And if we need a measure of welfare across times and not just at a time, a proposal that only tells us how to measure welfare at a time will not be sufficient.

A more promising option for the Maximising Consequentialist is to reject Count The Happy People, and instead adopt

Measure the Happy Person-Stages: The aggregate of wellbeing in the world is given by taking the welfare of each person-stage, and aggregating those levels of welfare.

Since time may turn out to be infinitely subdivisible, a continuous measure rather than a simple count is in order. (Measure the Happy Person-Stages also seems to us like the natural way to extend the proposal of the previous paragraph to a measure over worlds, not just over a single given time: though it does not rely on the Lewisian picture of identity-at-a-time for its appeal.)

The advocate of Measure the Happy Person-Stages will need to address an objection raised by Slote 1982 and Velleman 1991. These authors argue that one person can be better off than another, even when every person-stage of one can be mapped to an equally well-off person-stage of the other.[7] Velleman imagines two fictional lives: one that starts off with a destitute childhood and improves steadily to a fulfilled and contented twilight, and the other that starts off with an idyllic childhood and declines steadily to miserable old age. From the standpoint of the wellbeing of person-stages, the second life is just the first life run backwards. Yet someone who lives the first life is better off, all things considered, than someone who lives the second life: at least according to Velleman.

7 This somewhat simplistic statement of the claim presupposes that each person is composed of finitely many person-stages. In full generality, the claim ought is that one person can be better off than another even if for every level of wellbeing w, the measure of person-stages with wellbeing w belonging to one person is equal to the measure of person-stages with wellbeing w belonging to the other person.

Therefore, it seems that the wellbeing of a person cannot be reduced to the wellbeing of the stages that compose the person.

In responding to this objection, the advocate of Measure the Happy Person-Stages has two options. The first is to reject the apparent data as somehow misleading. Perhaps we have trouble truly imagining the two cases as they are described in the example, and tend to fill the first story in so that the person-stages in the first story are happier, on average, than in the second story. Or perhaps our tendency to prefer happy endings over sad ones fails to track any distinction in wellbeing. The second option for the advocate of Measure the Happy Person-stages is to claim that wellbeing is extrinsic, so that the wellbeing of one person-stage is entangled with the wellbeing of other person-stages R-related to it. For instance, a version of this view might posit that person-stages have intrinsic levels of proto-wellbeing, and a person-stage's wellbeing is an aggregate of the amount of proto-wellbeing it experiences together with the amount of proto-wellbeing experienced by forward R-related stages (perhaps with some discounting). This view may be bound up with our ability to Measure The Happy Person-Stages at all, since, as Sider 1996 and Hawley 2000 point out, most of the properties we attribute to persons are had by stages extrinsically if at all. It is not entirely clear how such a theory would interact with the Braddon-Mitchell and West view, which posits multiple R-relations, but there are nonetheless potential views in the vicinity whose details might be worked out.

Another option for Maximising Consequentialists is to reject Multiple Occupancy. We see three basic options here. First, Maximising Consequentialists might deny that there are persisting persons at all, instead preferring talk of person-stages and counterpart relations. Such a Maximising Consequentialist will also have to give up on Count the Happy Persons, and instead opt for Measure the Happy Person-Stages. Second, Maximising Consequentialists might claim that in puzzle cases

that appear to involve fission, continuant persons are destroyed and replaced with other (groups of) continuant persons. It is simple to see how this works in the case of fission, but more difficult to see how it would work in the case of Braddon-Mitchell and West's `transfiguration'. A third option is to adopt some more complicated view of identity across time, so that identity need not always be one-one (see, for example, the view developed in Gallois 1998). The cost here is that it is no longer clear how to Count the Happy Persons.

Yet another option for Maximising Consequentialists is to bite the bullet about our conclusions. There are a few arguments for doing so. Our maximising consequentialist could wax lyrical about the equal value of every person, whether common-sense morality recognises this or not. Just as Bentham and Mill argued to a sceptical public that every person mattered, slave or free, African or European, female or male, contemporary consequentialists can argue that temporary multiple occupancy is no bar to moral status as people: and a clearheaded following-through of this status leads to surprising conclusions that should nonetheless be accepted, here as in those other cases. Or she could complain about the strangeness of the cases: when we consult our intuitions about strange situations, very different from actual cases, perhaps it should not be surprising that the verdicts are themselves sometimes strange. That response does depend on there not being many actual cases of fission or fusion: if every gain or loss of parts results in an extra cohabiting person, as might be suggested by Dion and Theon cases, then actual multiple occupancy might be widespread, in which case paradoxical ethical consequences become much more pressing.

The other possible response, of course, is to give up Maximising Consequentialism in favour of some other moral theory, such as Kantianism or virtue ethics. Opponents of Maximising Consequentialism will doubtless see our results as more grist for their mill—worse even than ordinary utility monsters and repugnant conclusions. We think this is a fine response for Kantians

and virtue ethicists. (Though utility monsters potentially cause trouble for these theories too once they want to allow some role for beneficence, at least if they are tempted to allow that helping more people is better than helping fewer.) However, rejecting Maximising Consequentialism seems a strange response for philosophers otherwise tempted by Maximising Consequentialism—why swallow ordinary utility monsters and repugnant conclusions, only to strain at our metaphysically souped-up version? Maximising Consequentialists would do better to change their beliefs about exactly what ought to be maximised, or accept our conclusions.[8]

Rachael Briggs
Australian National University, Griffith University
Rachael.Briggs@anu.edu.au

Daniel Nolan
Australian National University
Daniel.Nolan@anu.edu.au

8 Thanks to David Braddon-Mitchell, Carrie Jenkins and an audience at the University of Kentucky for discussion.

References

Atsumi, H. 1965. "Neoclassical growth and the efficient program of capital accumulation". *Review of Economic Studies* 32: 127-36.

Baker, L. R. 2000. *Persons and Bodies: A Constitution View*. Cambridge University Press, Cambridge.

Braddon-Mitchell, D. and K. Miller. 2004. "How to Be a Conventional Person". *The Monist* 87.4: 457-474.

Braddon-Mitchell, D. and C. West. 2001. "Temporal Phase Pluralism". *Philosophy and Phenomenological Research* 62.1: 59-83.

Eklund, M. 2004. "Personal Identity, Concerns, and Indeterminacy". *The Monist* 87.4: 489-511.

Gallois, A. 1998. *Occasions of Identity*. Clarendon Press, Oxford.

Hamkins, J. D. and B. Montero. 2000. "With infinite utility, more needn't be better". *Australasian Journal of Philosophy* 78: 231-40.

Hawley, K. 2001. *How Things Persist*. Oxford University Press, Oxford.

Jeske, D. 1993. "Persons, compensation, and utilitarianism". *Philosophical Review* 102.4: 541-575.

Lewis, D. 1976. "Survival and Identity. Survival and Identity. In A.O. Rorty (ed.), *The Identities of Persons*. University of California Press, Berkeley.

Lewis, D. 1993. "Many, but Almost One" in Campbell, K., Bacon, J. and Reinhardt, L. (eds.) *Ontology, Causality and Mind: Essays on the Philosophy of D.M. Armstrong*. Cambridge University Press, Cambridge, pp 23-37

Langford, S. 2007. "How to Defend the Cohabitation Theory". *Philosophical Quarterly* 57.227: 212-224.

Mills, E. 1993. "Dividing without Reducing: Bodily Fission and Personal Identity". Mind 102.405: 37-51.

Mulgan, T. 2002. "Transcending the Infinite Utility Debate". *Australasian Journal of Philosophy* 82: 164-177

Noonan, H. 2003. *Personal Identity.* Routledge, London.

Nozick, R. 1974. *Anarchy, State and Utopia*. Blackwell, Oxford.

Olson, E. T. 2010. "Ethics and the generous ontology". *Theoretical Medicine and Bioethics* 31.4: 259-270.

Parfit, D. 1984. *Reasons and Persons*. Clarendon, Oxford.

Perry, J. 2002. *Identity, Personal Identity and the Self*. Hackett, Indianapolis.

Robinson, D. 1985. "Can Amoebae Divide Without Multiplying?". *Australasian Journal of Philosophy* 63.3: 299-319.

Shoemaker, D. 2007. "Personal Identity and Practical Concerns". Mind 116.462: 317-357.

Sider, T. 1996. "All the world's a stage". *Australasian Journal of Philosophy* 74.3: 433-453.

Slote, M. 1982. "Goods and Lives". *Pacific Philosophical Quarterly* 63: 311-26.

Smart J.J.C. "Utilitarianism and Punishment". *Israel Law Review* 25: 360-375.

Unger, P. 1992. *Identity, Consciousness, and Value*. Oxford University Press, Oxford.

Vallentyne, P. and Kagan, S. 1997. "Infinite Value and Finitely Additive Value Theory". *Journal of Philosophy* 94.1 :5-26.

Velleman, J. D. 1991. "Well-Being and Time," *Pacific Philosophical Quarterly*, 72.1: 48-77.

Von Weizsäcker, C. C. 1965. "Existence of optimal programs of accumulation for an infinite time horizon". *Review of Economic Studies* 32: 85-104.

Williams, B. 1981. "Persons, Character, and Morality". In *Moral Luck*. Cambridge University Press, Cambridge, pp 1-19.

Zimmerman, D. 2002. "Persons and bodies: Constitution without mereology?" *Philosophy and Phenomenological Research* 64.3: 599-606.