

MIT Open Access Articles

Should We Wish Well to All?

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Hare, Caspar. "Should We Wish Well to All?" *Philosophical Review* 125, 4 (October 2016): 451–472 © 2016 Cornell University

As Published: <http://dx.doi.org/10.1215/00318108-3624764>

Publisher: Duke University Press

Persistent URL: <http://hdl.handle.net/1721.1/115207>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Should We Wish Well to All?

1. Introduction

Some theories of the rights and wrongs of individual action (henceforth just *moral theories*) have a curious feature. They say that sometimes it is wrong for you to act as you would if you were reasonable and moved solely by individual concern for each one of the people affected by your actions. In short: they generate conflicts between *reasonable beneficence* and *morality*.

I want to do three things here. First I want to present some examples of moral theories that generate conflicts between reasonable beneficence and morality (this is the business of Sections 2 and 3). Second I want to present an argument to the conclusion that all such theories are mistaken (Section 4). Third I want to suggest that this raises some very interesting further questions about the nature of beneficence (Sections 5 and 6).

2. Socially Progressive Moral Theories and Reasonable Beneficence

Our first example involves moral theories that direct you to pay special attention to the less-well-off (henceforth *progressivist moral theories*, for short). Consider a very simple, very extreme progressivist moral theory, one that says that your single moral duty is to improve

¹ Many thanks to two anonymous reviewers, and to Michael Caie, Stephen Darwall, Toby Handfield, Richard Holton, Seth Lazar, Anna Mahtani, Michael Otsuka, Wlodek Rabinowicz, Agustin Rayo, Kieran Setiya, Larry Temkin, Judy Thomson and Alex Voorhoeve for most helpful comments. More generally, thanks to audiences at the London School of Economics, SUNY Buffalo, Vanderbilt University, the Philosophy Mountain Workshop, and the University of Colorado Boulder. And special thanks to participants at the 2015 Ethics and Decision Theory Workshop at the Australian National University.

the condition of the worst-off people affected by your actions (call it *extreme progressivism*).

And consider a case:

At the Doctor

You, the doctor, have twenty six patients – Abigail, Benny, Cofi, ..., and Zeke. You know them well. You and the twenty six know that one of the twenty six has been infected by *Maleficitus*, a very nasty disease, while the remaining twenty five have been infected by *Tolerabalitus*, a tolerably nasty disease. It is too early to guess who has what, but you and they know that now is your only opportunity to administer drugs, and that you have only two drugs at your disposal. One provides mild relief for *Maleficitus* sufferers, temporarily taking the edge off some of the symptoms. The other totally cures *Tolerabalitus*. But the drugs interact dangerously. Nobody wants to be taking both.

What should you do? If your single moral duty is to improve the condition of your worst-off patient then you should give them all the *Maleficitus*-relieving drug. That way none of them has to suffer through unrelieved *Maleficitus*. But each patient, insofar as he or she is reasonably tolerant of risk, does not want you to do that! Each patient, insofar as he or she is reasonably tolerant of risk, would take a 25/26 chance of cured-*Tolerabalitus* and a 1/26 chance of unrelieved-*Maleficitus*, over a 25/26 chance of uncured-*Tolerabalitus* and a 1/26 chance of (slightly) relieved-*Maleficitus*. And for you, the doctor, insofar as your concern for others involves reasonable toleration of risk to them, your concern for Abigail recommends giving out the *Tolerabalitus*-curing drug, and your concern towards Benny recommends giving out the *Tolerabalitus*-curing drug, ... and so on for all your patients.

So in this case extreme progressivism says that is wrong to act as you would if you were reasonable and moved only by individual concern for each of your patients. It generates a conflict between reasonable beneficence and morality.²

Extreme progressivism is a very uncompromising theory. Only the worst off matter. Few real philosophers accept that. Do less extreme, more plausible versions of progressivism generate conflicts between reasonable beneficence and morality? This is a tricky question to answer. Any progressivist theory specifies a certain way in which you should pay special attention to the less well off. It avoids the conflict if, in situations in which you are uncertain about who stands to gain or lose from your action (situations like At the Doctor), it would not be unreasonable for you, when acting out of concern for a particular person, to pay special attention in just the same way to possibilities in which things go less well for that person.³ So, in effect, making a judgment about whether a progressivist theory generates the

² The example has played an important role in discussions of progressivism. In brief: Parfit (1997) distinguished two kinds of progressivist theory, *welfare egalitarianism*, which says that equality of welfare is in itself good, and *welfare prioritarianism*, which says that incremental benefits to worse-off people have greater moral significance than equivalent incremental benefits to better off people. He then pointed out that welfare prioritarianism, unlike welfare egalitarianism, is not subject to a ‘leveling down objection’ – it does not say, implausibly, that we can make things in a morally significant way better by making everybody worse off). Rabinowicz (2001) then pointed out that prioritarianism is subject to a seemingly-similar-in-spirit objection: Though it does not say that sometimes we should act contrary to the interests of all, it does say that, in cases like At the Doctor, we should act contrary to the expected interests of all. There has been much subsequent discussion about whether this is a bad feature of the view. See Otsuka and Voorhoeve (2009), Parfit (2012), Otsuka (2012) and Otsuka (2015).

³ So, for example, a welfare prioritarian can avoid the conflict if she says that 1) acting morally consists in acting so as to maximize expected total priority-adjusted well-being, and that 2) acting from reasonable concern for a person consists in acting so as to maximize her expected priority-adjusted well-being. (Though there is then a question about whether she should really count as a ‘prioritarian’. If these ‘priority adjusted well-being levels’ matter both morally and prudentially then it is unclear why not to call them ‘well-being levels’ and be done with it.) A welfare prioritarian faces the conflict if she says that incremental benefits to the less-well-off matter much more, morally, than incremental benefits to the better-off, and that it would be unreasonable to adjust so

conflict involves making a judgment about just how risk-averse you can be, when acting out of concern for a particular person, without being unreasonable. But I know of no very convincing general proposals about just how risk-averse you can be on a person's behalf, without being unreasonable. We may all agree that it is unreasonable to be perfectly risk-averse on a person's behalf, to care only about minimizing the badness of the worst thing that can happen to them. I do not keep my children indoors, swathed in Kevlar. I would think you *nutty* if you did. But is it unreasonable to be moderately risk-averse on a person's behalf? If so, exactly what counts as 'moderate'? Many philosophers who think carefully about risk maintain a judicious silence on questions like this.⁴

Our next example will be cleaner. It involves a theory that is plausible enough to have been endorsed by many real philosophers. We do not need to make any assumptions about the reasonability of risk-aversion in order to see that this theory generates conflicts between reasonable beneficence and morality.

3. Constraints and Reasonable Beneficence

The single most famous and well-worn case from all of twentieth century normative ethics:

Footbridge

While loitering on a footbridge above a trolley track, you see that a trolley is heading towards a group of five people, immobile on the track. You can prevent their deaths,

heavily for well-being level in assessing risks to yourself or people you care about. In short: she faces the conflict when she is more *left-wing* than she is *personally risk-averse*.

⁴ Notably Lara Buchak in Buchak (2013). In Buchak (manuscript.) she outlines some restrictions on reasonable risk aversion.

but only by pushing a person, presently beside you, down onto the track, into the path of the trolley.⁵

As every sentient creature in the world now knows, the large majority of people who are asked about this case say that you ought not to push. Their judgment is endorsed by moral theories (henceforth *constraint-against-bad-killing* theories, for short) that say that, for a certain, bad kind of killing, we ought not to kill in this way, even when killing in this way brings about what utilitarians would call ‘the greater good’. What is the ‘certain, bad kind of killing’? Roughly: you kill in the bad way when the death of the person you kill *causes* the greater good to come about (getting less rough about this is a major, unfinished task in moral theory⁶).

We can easily fill in the details of the Footbridge case so as to show that constraint-against-bad-killing theories generate conflicts between reasonable beneficence and morality.

Footbridge With Suitcases

The six people involved here are Alexia, Benny, Cate, Dora, Emilio and Frederica. You know them all and care about them all. Why are they all so immobile? Why can’t the five just move off the track? Why will the one allow you to push him or her? Because the six people are trapped in six suitcases, one person per suitcase. The suitcases were shuffled a few moments ago. You have no idea who is where. Nobody has any idea who is where.

Constraint-against-bad-killing theories continue to say that you should not push (for this remains the bad kind of killing – the death of the one will cause the saving of the five), but

⁵ Thomson (1985).

⁶ The classic challenge is to precisify the theory so as to get the result that it is wrong to kill in Judy Thomson’s Footbridge case, okay to kill in Judy Thomson’s Switch and Loop cases. I should note that Thomson herself now rejects this challenge. Her final view, in Thomson (2008), is that all killing in cases like this is forbidden.

now each of the six wants you to push! Each of the six would take a $5/6$ chance of living and a $1/6$ chance of being crushed by a trolley (and saving five people by being crushed), over a $5/6$ chance of being crushed by a trolley (and saving nobody by being crushed) and a $1/6$ chance of living. And for you, insofar as you care about them, your concern for Alexia tells in favor of pushing, your concern for Benny tells in favor of pushing,... and so on for all of them.

Note that this time, in saying that each person, insofar as she has reasonable self-concern, will want you to push, and that you, insofar as you care about each person, will want to push, we are making no assumptions about the extent to which it is reasonable to be risk-averse on your own behalf and on behalf of the people you care about.

4. The Argument From Composition

So what? Is it a problem for a theory that it says that sometimes you should act against reasonable concern for all people involved? What should we think about cases like At the Doctor and Footbridge with Suitcases? Well, in both these cases my first inclination is to say that it is morally fine to act out of reasonable concern for all people involved. Go ahead – give out the Tolerabailitus-curing pills, push the suitcase on the track. Moral theories that say otherwise are mistaken.

Some people have shared my first inclination, notably John Harsanyi.⁷ But, oddly, I find that many contemporary philosophers do not. They say that the important question is not whether I can cast my action as an expression of concern for all, given what I now know (in now common terms: whether it is *ex-ante justifiable to all*), but whether I could cast my

⁷ See Harsanyi 1955 and Harsanyi (1976).

action as an expression of concern for all, if I knew everything about its consequences (in now common terms: whether it is *ex-post justifiable to all*)⁸.

And they have further things to say in support of their view. Derek Parfit, writing about progressivist moral theories and cases like At the Doctor, suggests that if we find the idea that it is morally fine to act out of reasonable concern for all appealing, it is because we are making a particular mistake. It might be tempting to think “Of course you should do what everybody involved reasonably wants you to do. It is in everybody’s interests that you do that. And we are never morally obliged to act against everybody’s interests.” But that is not right. Though all of your patients reasonably want to be given the Tolerabalitus-curing drug, it does not follow that it is in all of their interests to be given the Tolerabalitus-curing drug. One of them has Maleficitus. It is in his or her interests that you give out the Maleficitus-relieving drug. And something similar could be said about Footbridge With Suitcases. It might be tempting to think “Of course you should push. It is in everybody’s interests that you push.” But that is not right. Though all of the entrapped people reasonably want you to push, one of them is on the footbridge. It is in his or her interests that you hold back.⁹

And Marc Fleurbaey and Alex Voorhoeve, also writing about progressivist moral theories and cases like At the Doctor, suggest that if we think that ex ante justifiability is what matters then we will be liable to predictable changes of mind in response to new information. So you, the Doctor, not knowing who has what disease, will want to give everyone the Tolerabalitus-curing drug, but you will know that you would not want to do this if

⁸ These terms are entrenched, but a bit misleading. The intended contrast is not between what I can say *beforehand* and what I can say *afterwards*. It is between what I can say, *given what I know*, and what I could say *if I knew more*.

⁹ Parfit (2012).

you knew who had what disease. You will then have an aversion to information that is highly relevant to the problem at hand. Such aversions are irrational.¹⁰

I am not persuaded by either of these arguments. So far as mistakes go, of course somebody *might* mistakenly assume that it is in everybody's interests that, e.g., you push. But I do not think that I am assuming that. I think that I am well aware that there is someone in the suitcase, and that he or she will die if you push. And yet the thought that you should push remains appealing to me. And, so far as aversions to information go, I do not think that aversions of the sort Fleurbaey and Voorhoeve describe need be irrational. In cases like this, where things of incommensurable value (like people's lives) are at stake, it makes perfect sense to resist the judgments of your better informed self.¹¹

So is there anything to be said on the other side, beyond just insisting that it is obvious and right?¹² Happily I think there is. I think there are at least three strong arguments to the conclusion that theories like extreme progressivism and constraint-against-bad-killings are mistaken, that you ought to act out of reasonable concern for all in some cases in which these theories say otherwise.

Two of the arguments I am not going to talk about in any detail here. The first is the argument from *presumed consent*. Roughly: Suppose you consult the people involved (talking to

¹⁰ Fleurbaey and Voorhoeve (2013).

¹¹ See Hare (2010) and Hare (2013) Chapter 3.

¹² In Harsanyi's (1976) axiomatization of utilitarianism, he proposes an axiom that effectively ensures that morality and beneficence will not come apart: '*Axiom 3: Pareto Optimality*. Suppose that at least one individual $j(j=1, \dots, n)$ personally prefers alternative A to alternative B, and that no individual has an opposite personal preference. Then individual i will morally prefer alternative A over alternative B.' Glossing the axiom, he says only: 'Axiom 3 is a very weak and hardly objectionable moral postulate.' But, as John Broome pointed out in Broome (1987), once we understand that 'is personally preferred to' and 'is morally preferred to' play the role in Harsanyi's theory that 'is better for an individual than' and 'is better than' play in standard utilitarian theory, and once we understand that the 'alternatives' to which he refers are uncertainty-weighted prospects, not outcomes, this axiom no longer appears weak or 'hardly objectionable'. It is the sort of thing that needs to be argued for.

them face to face in the doctor case, via mobile phone in the footbridge case). What will they say? The patients will all say they want you to give out the Tolerabalitus-curing drug. The people trapped in suitcases will all say that they want you to push. Wouldn't it be okay for you to push then? If so, wouldn't it be okay for you to push in our case, on grounds that they would request it if asked?¹³

The second is the argument from *dirty hands*. Roughly: in a case like Footbridge With Suitcases, if you willingly refrain from pushing and you are rational then you must prefer outcomes in which everybody, including you, is worse off over outcomes in which everybody, including you, is better off. But it is indecent to have preferences like that.¹⁴

The third argument, the argument I want to talk about in detail here, has to do with *composition*. The basic idea: Break the action we looking at up into proper sub-actions, each of which only influences the interests of one person. Then argue that you ought to do each of these sub-actions, appealing to the idea that, if you know that an action affects only one person's interests, then it is morally permissible to act out of concern for that person. Then infer, by way of way of ought-agglomeration, that you ought to do all of the sub-actions.

Now for the detail. Let's begin with a warm up. Here is an argument to the conclusion that in a certain case (not exactly Footbridge With Suitcases) you ought to kill one person by pushing him or her under a trolley, so as to prevent five people from being crushed by trolleys. I will try to set it out as clearly as I can, in austere premise-and-conclusion form. If you reject the conclusion then it will be interesting to see which premise you reject.

¹³ Frances Kamm explores this sort of argument in Chapter 11 of Kamm (2001).

¹⁴ I discuss this in Hare (2013) Chapter 4.

Six Tracks

You see six tracks. On each track is a trolley, heading toward a pile of five suitcases. Above each track is a footbridge, on which is perched one suitcase. You can stop each trolley, but only by pushing the suitcase on the footbridge above it into its path. You learn that Alexia is in a suitcase in the vicinity of the first track (henceforth *Alexia's track*) – either in the suitcase on the footbridge (henceforth *Alexia's footbridge*) or in one of the five suitcases on the track. The remaining five suitcases in the vicinity of Alexia's track are full of sand. Likewise, Benny is in a suitcase in the vicinity of the second track (henceforth *Benny's track*), with the remaining five suitcases in the vicinity of the Benny's track filled with sand... and so on for Cate, Dora, Emilio and Frederico. Furthermore you learn that one and only one of them, you have no idea who, is on a footbridge.

The first premise:

P1 In Six Tracks, irrespective of what you do later (with respect to pushing the other five suitcases), you ought to push the suitcase on Alexia's footbridge down onto Alexia's track. And, irrespective of what you did earlier and will do later (with respect to pushing the other five suitcases), you ought to push the suitcase on Benny's footbridge down onto Benny's track... and so on.

(A clarificatory aside: I am working with a *subjectivist* notion of oughtness here – according to which what you ought to do is intimately connected with your beliefs about what will happen if you do one thing or another. This may make you pause. Some philosophers insist that the term 'ought' is ambiguous between this subjective reading and an objective reading – according to which what you ought to do is intimately connected with

what will happen if you do one thing or another. If you are one such person then please substitute 'subjectively ought' for 'ought' in all that follows. Some other philosophers say that 'ought' only supports an objective reading. Instead of saying 'you subjectively ought to do this' we should say 'it is rational for you to do this', 'it is sensible for you to do this', or, when things of moral significance are at stake, 'if you were decent then you would do this.' If you are one such person then please make these substitutions in all that follows.)

Clarificatory aside over, why accept **P1**? Well, let's attend (without loss of generality) to that first suitcase, the one on Alexia's footbridge. You know that pushing it will affect Alexia and nobody else. If you push it then there's a 1/6 chance, from both your and her point of view, of your killing her and a 5/6 chance of your saving her life. Those odds sound good to you and her. Why not push?

A reply: 'Because then you run a risk of killing her!'

True, but this exactly the sort of thing that a brave surgeon does. She operates on her patient for his own sake, knowing that there is a small chance that she will kill him and a larger chance that she will save his life. We have reached the point where I have no further argument, but am I happy to resort to strong language. It is *absurd* to say that it is never appropriate to run a risk of killing a person when we intervene solely for that person's own sake.

A more sophisticated reply: 'What matters is keeping your confidence that you kill somebody at some point acceptably low. If you knew that you weren't going to push any of the other suitcases, then you could push now while keeping your confidence that you kill somebody at some point acceptably low. Then you ought to do it. But once you know that you have pushed in the past or that you will push in the future, you cannot push now without allowing that confidence to rise unacceptably high. In the extreme, once you know

that have pushed all five other suitcases, then you cannot push now without becoming *near certain* that you kill somebody at some point. That makes it okay to refrain from pushing.’

But again, the brave surgeon does not just operate on one patient. She operates on many patients. And as she operates on more and more, her confidence that she kills somebody at some point rises and rises. Does that change her moral obligations to her later patients? The suggestion seems to be that she should get her medical degree and perform one or two risky operations. Then she can retire, leaving future risky operations to be performed by new recruits whose slates are clean.¹⁵ Again, I am happy to say that this is *absurd*.

If you are not yet convinced then please ask yourself what you would do if Alexia were someone you cared about. Really. Imagine she is your daughter, or mother, or sister, or old friend. You know that she is in one of the six suitcases in the vicinity of the first track. You (and she) think it 5/6 likely that she is in one of the five suitcases on the track, in the path of the trolley. You know there is nobody but you and Alexia anywhere near the track... What would you do? I would push. And though I would feel vexed about it, though I would chew my fingers in horror at the thought that I might be killing her, I would not feel guilty about it. This is just between me and her, and I am acting out of concern for her interests as best I can. That is what I ought to do.

The second premise is a general principle concerning the relation between the deontic status of actions and the deontic status of their proper parts. Call it *Weak*

Agglomeration:

¹⁵ Thanks to Frank Jackson for making a close variant of this point to me.

P2 For any composite action $A_1 \dots A_n$, if irrespective of whether you do A_2 through A_n , you ought to do A_1 , and... and irrespective of whether you do $A_1 \dots A_{n-1}$, you ought to do A_n , then you ought to do $A_1 \dots A_n$.

Why accept **P2**? I think it compelling because I think that it is the job of a moral theory to give you a package of advice, and a package of advice that is *followable* in the following sense: there is something that you can do such that, if you were to do it, then you would have done everything that the theory says you ought to do. Any theory that violates Weak Agglomeration does not give us a package of advice that is followable in this sense. To that extent it is defective.

(Another clarificatory aside: The ‘Weak’ is in there to distinguish it from a principle that is controversial in the theory of practical rationality, *Strong Agglomeration*:

For any composite action $A_1 \dots A_n$, if you ought to do A_1 , and... and you ought to do A_n , then you ought to do $A_1 \dots A_n$.

Philosophers who call themselves *actualists* reject Strong Agglomeration.¹⁶ They say that it fails in cases where $A_1 A_2$ is the best composite action you can do, but doing either A_1 -but-not- A_2 or A_2 -but-not- A_1 is disastrously bad (worse than doing neither A_1 nor A_2), and you know that, although you can do $A_1 A_2$, as a matter of fact if you do A_1 then you will not do A_2 . In such cases you ought to not-do- A_1 , and you ought to not-do- A_2 , but you ought to do $A_1 A_2$.

¹⁶ The canonic statement of actualism is in Jackson and Pargetter (1986).

But actualists (fortunately for them, I think) need not reject Weak Agglomeration, nor need they reject the idea that their job is to give a followable package of advice. In cases where Strong Agglomeration fails, according to the actualist, although you ought to do A_1A_2 , and ought to not-do- A_1 , there is something (A_1A_2) such that if you were to do it then you would have done everything that you ought to have done – if you were to do A_1A_2 then it would be the case that you ought to do A_1 , because it would be the case that if you were to do A_1 then you would do A_1A_2 .)

Aside over, the conclusion follows:

C In Six Tracks you ought to push all the suitcases – knowingly killing one person and saving five.

Fine, but Six Tracks is not exactly like Footbridge with Suitcases, the case we started with. And maybe one difference is important. In Six Tracks the death of the person you kill is not a cause of the saving of the five. If the person had not been killed then the five would still have been saved. Maybe this is not the bad, constrained-against kind of killing. So maybe a constraint-against-bad-killing theorist can concede that is okay to kill in this way.

Finding a case in which the action of killing-the-one-and-saving-the-five decomposes into independently performable sub-actions, each of which affects one and only one known person (as in Six Tracks), but in which the killing of the one causes the saving of the five (as in Footbridge With Suitcases), is a little trickier. But it can be done. Here is an example:

Two Parallel Tracks – Six Buttons

You see two tracks, running in parallel. On each track is a trolley, heading toward a pile of five suitcases. Above each track is a footbridge, on which is

perched one suitcase. You learn that Alexia, Benny, Cate, Dora, Emilo and Frederica are in the six suitcases in the vicinity of the first track, but you have no idea who is where. And you learn that the six suitcases in the vicinity of the second track are full of sand. You cannot stop the first-track-trolley, but you can press any combination of six buttons – marked ‘Move Alexia’, ‘Move Benny’... and so on. By pressing the ‘Move Alexia’ button you move Alexia from her position on the first track to the corresponding position on the second track (though you don’t see the move). By pressing the ‘Move Benny’ button you move Benny from his position on the first track to the corresponding position on the second track (though you don’t see the move)... and so on. By pressing any of the buttons you push the suitcase on the footbridge on the second track, and thereby stop the second trolley (if you press more than one button then this is symmetrically over-determined by your button-pressing).

In this case, by pressing all six buttons you move everybody from the first track to their corresponding position on the second track, and push the person on the bridge into the path of the second trolley, killing him or her in such a way that his or her death (or something very close to it – the impact of the trolley against his or her body) causes the trolley to stop and the five to be saved. This appears to be the bad, constrained-against kind of killing, but the argument from composition can be applied to this case too:

P1 In Two Parallel Tracks – Six Buttons, irrespective of what you do later (with respect to button-pushing), you ought to press the ‘Move Alexia’ button. And, irrespective of what you did earlier and will do later (with respect to button-pushing), you ought to press the ‘Move Benny’ button... and so on.

- P2** For any composite action $A_1 \dots A_n$, if irrespective of whether you do A_2 through A_n , you ought to do A_1 , and... and irrespective of whether you do $A_1 \dots A_{n-1}$, you ought to do A_n , then you ought to do $A_1 \dots A_n$.
- C** In Two Parallel Tracks – Six Buttons you ought to press all the buttons – knowingly moving all six people to the lower track, and pushing one of them into the path of the trolley in such a way that his or her death causes the saving of the five.

And the motivation for **P1** is much as before. You know that pressing (e.g.) the ‘Move Alexia’ button will affect Alexia’s survival and nobody else’s. If you press it then there’s a 1/6 chance (from your and her point of view) of your killing her and a 5/6 chance of your saving her life. Why not press?

A reply: maybe, if you are pressing no other buttons, then it you ought to press the ‘Move Alexia’ button. But suppose that you are pressing other buttons. Then, if you press the ‘Move Alexia’ button, there’s a 1/6 chance of your killing her *in the bad way* – killing her in such a way that her death causes the saving of the five. That changes things. It is never true that you ought to risk killing someone in the bad way.

One might say that, but I find it very hard to resist this now-familiar thought: because you know that pressing the ‘move Alexia’ button has no impact on the well-being of anybody other than Alexia, the decision to press or not is in a certain way *private* – it is between you and her. She would take the 1/6 chance of being killed and 5/6 chance of being saved. It doesn’t matter to her whether the being-killed involves being-killed-in-the-bad-way (her death saves lives) or being-killed-in-the-okay-way (her death saves no lives). If

anything, she would reasonably prefer that her death saved lives.¹⁷ And, insofar as you care about her, you should feel the same, and act accordingly.

Imagine, again, that Alexia is your daughter, or sister, or mother, or old friend.

Would you press the ‘move Alexia’ button? Again I would press. Again I would feel worried but not guilty. I would be moving her for her own sake, and only for her own sake.

We can call this thought, the thought that I am repeatedly appealing to here, *defeasible privacy*: When you know that an action will affect the well-being of only one person, and you know who that person is, then it is morally appropriate, in the absence of powerful defeating considerations, to act out of reasonable concern for that person. What would be a powerful defeating consideration? Maybe something like ‘this person has done horrific things and does not deserve to reap the rewards of her doing those horrific things’ would do the job. Maybe then it would be fine not to act out of concern for her. But, the thought is, when you know that only one person has something at stake in your action, and you know who she is, there is a strong moral presumption in favor of acting out of reasonable concern for her. Any defeaters must be powerful. Considerations like, ‘if I do this then there is a chance that I will harm her in the bad way’ or ‘the worst thing that could happen to her if I do this is worse than the worst thing that could happen to her if I do not do this’ are much too weak.

That’s the argument. It applies just as well to extreme progressivism and our original At the Doctor case, just so long as you, the doctor, are not giving out the drugs all in one go, but to each patient individually. It applies quite generally to cases in which there is a beneficent-to-all action (one recommended by individual reasonable concern for each person

¹⁷ This observation was made by Derek Parfit: “...when someone else could act in some way that would both kill us but also save several other people’s lives, we would have no strong reason to prefer being killed as a side-effect of the saving of these people’s lives rather than as a means. It would be in one way better to be killed as a means, since our death would then at least do some good.” Parfit (2011) p. 365.

affected by your action) that decomposes into independently performable sub-actions, each of which affects one and only one, known person, and each of which is recommended by reasonable concern for that person. In such cases you ought to take the beneficent-to-all action. Moral theories that say otherwise (e.g. some forms of progressivism and standard forms of constraint-based deontology) are mistaken.

Stepping back from the details, we can think of the argument as offering a solution to hard problem in moral theory. Hard problems in moral theory arise when there are prima facie appealing sets of moral norms that turn out, on careful reflection, to be inconsistent. In this case there is a prima facie appealing set of moral norms governing our interactions with groups of people, expressed by platitudes like *be fair, attend to the neediest first, don't leave anyone behind, don't use some to benefit others*. And there is a quite different, also prima facie appealing set of moral norms governing our one-on-one interactions with individual people, expressed by platitudes like *be guided by concern for the person's well being, act as she would reasonably want you to act if she knew what you knew*. But interact with enough individuals and you have interacted with a group! To avoid inconsistency one set of norms must be restricted. I say it is the first set.

5. Nearby Cases

Well and good, but just how far must the group-interaction norms be restricted? The argument from composition applies when a beneficent-to-all action decomposes into independently performable sub-actions, each of which affects one and only one, known person, and each of which is recommended by reasonable concern for that person. What should we say about nearby cases that do not satisfy this intricate condition? Do (e.g.) the progressivist and deontological moral theories that fail when the condition is satisfied hold

up when the condition is not satisfied? These are the questions I will address in the remainder of this paper. I will look at some nearby cases in this section. What I say about them will be more speculative than what I have said thus far, but I hope that it will allow me to frame an important and interesting problem.

One nearby case involves a beneficent-to-all action that does not decompose in the right way. An example:

Two Parallel Tracks – One Big Button

Your situation is exactly as in Two Parallel Tracks, but this time there is just one big button. Pressing it will have the same effect as pressing all six buttons in Two Parallel Tracks.

Can we apply the argument from composition here? Well, in this case maybe we can say that, by pressing the big button, you do six things – you move Alexia, you move Benny... etc. And maybe we can still ask whether you ought to move Alexia, whether you ought to move Benny... etc. But in this case moving Alexia involves moving them all, so, in order to decide whether you ought to move Alexia, we need immediately to take a view about whether you ought to move them all. The claim that you ought to move Alexia cannot then serve as a premise in a persuasive argument to the conclusion that you ought to move them all. The argument from composition is not going to work.

Nonetheless my inclination is to think that you ought to press in this case too. In Two Parallel Tracks – Six Buttons you, the morally contentious person, press all six buttons because you want to move all six people from their position on the upper track to the corresponding position on the lower track. Why does it matter, morally, whether you get what you want by pressing six buttons or one button? This seems to me to be too delicate a thing to support so much moral weight.

To dramatize the point: Imagine that, in this case, you start off believing that the big button is not wired to anything. Then you are told that pressing it will move Alexia from her position on the upper track to the corresponding position on the lower track. You welcome this news. You decide to press the button. And then you learn that it will do something else too: it will move Benny from his position on the upper track to the corresponding position on the lower track. Isn't this further good news? After all, if the opportunity arose to move Benny by pressing another button then you would take it. And you would take it because you want to move Alexia and Benny – which is exactly what pressing the big button does. Why should you care about whether you get what you want by pressing multiple buttons or one button?

I do not yet see a very convincing reply to this.

Another nearby case is one in which they, but not you, know precisely whom you are in a position to kill, whom you are in a position to save. An example:

Two Parallel Tracks – Six Buttons – Knowing Victim

Your situation is exactly as in 'Two Parallel Tracks – Six Buttons', except this time each of the six knows his or her place on the track.

What should you do now? I say that, although you cannot anymore take solace in the thought that each one of them wants you to press, the argument from composition still applies. Ought you press, e.g., the 'Move Alexia' button? There's a 5/6 chance that Alexia is thinking "Please, please, press that button – it will save my life and affect nobody else." There's a 1/6 chance that she is thinking "Please, please, don't press that button – it will kill me and affect nobody else." Pressing remains the thing to do.

The really interesting nearby case is one in which you know precisely whom you are in a position to save and precisely whom you are in a position to kill. What should you do now? Obviously, if the action decomposes into independently performable sub-actions, as in Two Tracks – Six Buttons, and you know which sub-action affects which person, then you should save the five but leave the one alone. But what if the action does not decompose, or you don't know which sub-action affects which person? Consider:

Two Parallel Tracks – One Big Button – Known Victim

Your situation is exactly as in Two Parallel Tracks – One Big Button, except this time you know that Alexia is the person on the upper bridge, the person you are in a position to kill.

Two Parallel Tracks – Six Buttons – Known Victim – Unknown Buttons

Your situation is exactly as in Two Parallel Tracks – Six Buttons, except this time you know that Alexia is the person on the upper bridge, and this time you do not know which button moves which person. You just know that by pressing all the buttons you will kill Alexia and save Benny, Cate, Dora, Emilio and Frederica.

In these cases pressing is no longer beneficent-to-all. What ought you to do?

The argument from composition will not establish that, still, you ought to press. But this may not be a failing in the argument, because it may be that the deontic facts have changed – it may be that it matters whether you know who is who.

For thinking that it is no longer the case that you ought to press: Now, because you know so very much about Alexia and the rest, you have strong reasons not to press. These are considerations to do with Alexia's mother and father, to do with her career, to do with her funny impressions of old philosophers etc. And you have strong reasons to press. These are considerations to do with Benny's son, Cate's unfinished sonata etc. The reasons are

incommensurable. Faced with such incommensurability you need a tiebreaker, and a policy of non-intervention in cases like this is as good a tiebreaker as any.

For thinking that you ought *not* to press: By pressing you express the belief that your reasons not to press (to do with Alexia's family, her career, her funny ways etc.) are outweighed by your reasons to press. This is a mistake. Alexia is not fungible in this way. And it is a mistake that diminishes us all.

If the second way of thinking is right then classical deontological theories remain sound in their application to cases in which you know who is who. Is it right? Let's put this question to one side. (I have view about it. I am one of those people who thinks that, by pressing, you need not be taking any more or less of a stand on the relative strength of your reasons than by not-pressing. But many people do not share my view. This dispute is old and deadlocked. I doubt we will settle it here.) Let's just adopt the first, anti-aggregative way of thinking. Once you know who is who, it is at least no longer the case that you ought to press.

This gives rise to a curious problem.

6. What is it to Know Your Victim?

Look hard at the Two Parallel Tracks – Six Buttons – Known Victim – Unknown Button case. Why doesn't the argument from composition apply to this case too? In this case you may dub the person who will be moved by your pressing the first button, whoever he or she is, 'Oney'. And you may dub the person who will be moved by your pressing the second button, whoever he or she is, 'Twoey'... and so on. In this case, you know that pressing button one affects only Oney, and, by pressing button one, you raise your confidence that

Oney will survive from 1/6 to 5/6.¹⁸ So it would seem, by defeasible privacy, that you ought to press button one. Likewise, you know that pressing button two affects only Twoey, and, by pressing button two you raise your confidence that Twoey will survive from 1/6 to 5/6. So it would seem, by defeasible privacy, that you ought to press button two... And so it would seem, by weak agglomeration, that you ought to press all the buttons.

The natural answer is that, in this case, for each button, you now have a strong reason not to press it. Yes, pressing button one raises your confidence that Oney will survive, but pressing button one also reduces your confidence that Alexia will survive. That matters.

But why can't the same be said about Two Parallel Tracks – Six Buttons? In that case you may dub the person on the upper bridge, whoever he or she is, 'Bridgely'. Yes, pressing button one raises your confidence that Alexia will survive, but pressing button one reduces your confidence that Bridgely will survive. Why doesn't that matter in just the same way? You don't know whether Bridgely is Alexia, Benny, Cate, Dora, Emilio or Frederica. But isn't Bridgely, in any case, a human being, with dignity, and worthy of respect?¹⁹

To answer this question in a full and satisfying way we need an explanation of why the cases differ. And we need an explanation that can be applied quite generally. I have said that when you know nothing about the person you are in a position to kill, beyond that he or

¹⁸ Putting this thought in a precise way is tricky. On one understanding of propositions, supposing that Oney is (e.g.) Alexia, the proposition that Oney will survive is the proposition that Alexia will survive, so it cannot be that you have greater confidence in the proposition that Oney will survive than in the proposition that Alexia will survive. We must put the thought in terms of modes of presentation, or metalinguistically (you are more confident in the truth of the sentence 'Oney will survive' than in the truth of the sentence 'Alexia will survive'). On other understandings of propositions this is not so. We can say there are two propositions in which you have different levels of confidence. For present purposes I don't think it matters which way we go.

¹⁹ Elizabeth Harman asks a question very like this in her (2015) review of my (2013) book.

she is one of the six, you ought to kill. When you learn that the person you are in a position to kill is on a bridge, still you ought to kill. When you learn that the person you are in a position to kill is Alexia, and you know a very great deal about Alexia, it is not the case that you ought to kill. What triggers the change? Consider:

Two Parallel Tracks – Six Buttons – Poet Victim

You learn that the person on the footbridge has a secret talent for balladic poetry.

Ought you to press now?

Two Parallel Tracks – Six Buttons – ‘Bugaboo’ Victim

You learn that the person on the footbridge was affectionately nicknamed ‘Bugaboo’ by his or her mother.

Ought you to press now?

Two Parallel Tracks – Six Buttons – Fleshy-Eared Victim

You notice a crack in the suitcase on the footbridge. Sticking out of the crack is the occupant’s red, fleshy ear. A bead of sweat drops from the helix into the cavum conchae.

Ought you to press now?

Rather than answering these questions directly, let me offer a way of thinking about them.

I said earlier that Defeasible Privacy applies when you know who the person you are in position to affect is. A natural proposal, then, is this: It ceases to be true that you ought to press at the point at which you know who your victim, the person on the footbridge, is. In

Two Parallel Tracks – Six Buttons – Known Victim – Unknown Button you know who your

victim, Alexia, is. In Two Parallel Tracks – Six Buttons you do not know who your victim, Bridget, is. What we need is a general theory of what it is to *know who somebody is*. Happily, philosophers of language have worked for many years on this project (typically by way of addressing the question: What is it to be capable of entertaining ‘singular thoughts’ about a person?) So maybe our strategy should be to borrow their best theory, apply it case-by-case and thereby find out what your obligations are.

I do not think this strategy very promising. For one thing, philosophers of language have not arrived at anything like consensus on what it is to know who somebody is. For another thing, it is not clear to me that the considerations to which philosophers of language have been sensitive in accepting or rejecting this or that theory of singular thought have had to do with the moral significance of singular thought. So it is not clear to me that their best theory would go very far towards explaining why it matters so much, morally, whether you know who your victim is. This talk of ‘knowing who the person you are in a position to affect is’ should be regarded as shorthand for something with explanatory force.

Here is a strategy that I find more promising. We start with this proposal: it ceases to be true that you ought to press when you have reasons to press, reasons not to press, and you no longer have most reason to press. What we then need to do is to think about your reasons to press, and about whether they are decisive.

In the background, I am thinking of reasons for acting a certain way as considerations that count in favor of your acting that way. I am thinking of a reason as yours when you are aware of it, and are in a position to be motivated by it. I am thinking that sometimes you may have reasons to do one thing, reasons to do another, inconsistent, thing. I am thinking that in some such cases there is most reason to do one or the other thing, in other cases there is not. I am thinking that the latter is typical of cases in which there are

incommensurable values at stake – roughly when there are ways in which it is good if you do the one thing, very different ways in which it is good if you do the other thing, and no fact of the matter about it being overall better if you do either.

When you learn more about your victim you may thereby gain reasons not to press. Suppose you have drawn up ‘Pro-Pressing’ and ‘Con-Pressing’ columns. Now suppose you learn about the balladic poetry. That gives you something new to write in the ‘Con-Pressing’ column: ‘If I press then an important poetic voice will be stifled’. Now suppose you see the sweaty ear. That gives you something further to write: ‘If I press then that ear will soon be blue.’ As you learn more and more, the reasons against pressing mount up, and after a while (any proposal as to exactly when will have an unseemly precision to it) your reasons to press are no longer decisive.

But why doesn’t the knowledge that pressing will kill *Bridget* already give you powerful reasons not to press? After all, you know that, whoever Bridget is, it will be in certain ways worse if Bridget dies and the others live, than if Bridget lives and the others die. Why can’t you dub these ways of being better ‘Bridget-values’ and think to yourself ‘Pressing will be Bridget-worse than not pressing’?

Of course you can think that to yourself, but I say that you still do not have a reason not to press. To have a reason to do something you must do more than recognize that there is a way in which it is better if you do it. You must be in a position to recognize the way of being better, and thereby be moved to pursue it.

Consider a less vexed case, in which the values in play are of little moral significance:

Oprah or Opera?

In one envelope I place a ticket to Oprah (*thrilling, a once-in-a-lifetime opportunity to commune with the popular zeitgeist*). In another envelope I place a ticket to the opera

(*sublime, uplifting*). I shuffle the two envelopes, place them before you and allow you to pick one.

In this case, though you know that there is a way in which the first envelope is better than the second envelope (the opportunity it gives you is either more thrilling or more sublime), I say you have no reason to pick the first envelope over the second envelope. The consideration ‘There is a way in which the first envelope is better’ is not a reason you have. If you pick the first over the second, and I ask you what moved you to do that, you can’t very well reply: ‘I was moved by the thought that there is a way in which it is better if I pick it.’ You knew that the very same thing could be said for picking the second over the first – there is a way in which it is better if you pick it! You do not have a reason until you recognize the way in which it is better if you pick the first envelope. And just dubbing it ‘lefty-value’ will not help you recognize it. You need to open the envelopes.²⁰

Likewise, in Two Parallel Tracks – Six Buttons, the consideration ‘There are ways in which it is better if I don’t press’, though true, is not a reason you have not to press. You know the very same thing can be said for pressing – there are ways (more ways, in fact) in which it is better if you press! You do not have reasons until you recognize the ways in which it is better if you don’t press. And just dubbing them ‘Bridgey-values’ will not help you recognize them. You need to learn more about Bridgey.

²⁰ This is a background assumption of the most-reason argument for prospectist decision theory in Hare (2010) and Hare (2013) Chapter 3.

7. Wrapping Up

In sum: Many moral theories tell us to act against reasonable concern for all in some cases. The argument from composition shows that such theories are mistaken. Wish well to all, and act on your wishing.

References

- Broome, John. 1987. "Utilitarianism and Expected Utility." *Journal of Philosophy* 84 (8): 405-422.
- Buchak, Lara. 2013. *Risk and Rationality*, Oxford: Oxford University Press.
- Buchak, Lara. (Manuscript). "Taking Risks for Others and Aggregating Risks."
- Fleurbaey, Marc, and Alex Voorhoeve. 2013. "Decide as you Would with Full Information! An Argument Against Ex Ante Pareto." In *Inequalities in Health: Concepts, Measures and Ethics*, ed. Nir Eyal, Samia Hurst, Ole Norheim and Dan Wikler, 113-128. Oxford: Oxford University Press.
- Greene, Joshua, R.B. Somerville, L.E. Nystrom, J.M. Darley, J.D. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293: 2105-2108.
- Hare, Caspar. 2010. "Take the Sugar." *Analysis* 70 (2): 237-247.
- Hare, Caspar. 2013. *The Limits of Kindness*. Oxford: Oxford University Press.
- Harman, Elizabeth. 2015. "Review: Caspar Hare, *The Limits of Kindness*." Oxford: Oxford University Press. *Ethics* 125 (3): 868-872.
- Harsanyi, John. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63: 309-321.
- Harsanyi, John. "1977. "Morality and the Theory of Rational Behavior." *Social Research* 44: 623-656.
- Jackson, Frank, and Robert Pargetter. 1986. "Oughts, Options and Actualism." *Philosophical Review* 95 (3): 233-255.
- Jackson, Frank, and Michael Smith. 2006. "Absolutist Moral Theories and Uncertainty." *Journal of Philosophy* 103 (6): 267-283.
- Kamm, Frances. 2001. *Morality, Mortality Volume II*, Oxford: Oxford University Press.

- Otsuka, Michael. 2015. "Prioritarianism and the Measure of Utility." *The Journal of Political Philosophy* 23 (1): 1-22.
- Otsuka, Michael. 2012. "Prioritarianism and the Separateness of Persons." *Utilitas* 24 (3): 365-380.
- Otsuka, Michael, and Alex Voorhoeve. 2009. "Why it Matters that Some Are Worse Off Than Others: An Argument Against the Priority View." *Philosophy and Public Affairs* 37 (2): 171-199.
- Rabinowicz, Wlodek. 2002. "Prioritarianism for Prospects." *Utilitas* 14 (1): 2-21.
- Parfit, Derek. 1997. "Equality and Priority." *Ratio* 10 (3): 202-221.
- Parfit, Derek. 2011. *On What Matters, Volume One*. Oxford: Oxford University Press.
- Parfit, Derek. 2012. "Another Defense of the Priority View." *Utilitas* 24 (3): 399-440.
- Thomson, Judith. 1985. "The Trolley Problem." *The Yale Law Journal* 94 (6): 1395-1415.
- Thomson, Judith. 2008. "Turning the Trolley." *Philosophy and Public Affairs* 36 (4): 359-374.