Against Anonymous Pareto

ERAN FISH

The Hebrew University of Jerusalem

The principle known as 'anonymous Pareto' has it that an alternative *A* is better than another, *B*, in case it is (strictly, non-anonymously) Pareto superior to either *B* or a permutation of it. It is an attractive idea, offering to apply Pareto-based judgments to a broader range of cases while preserving some of the intuitive appeal of the standard, more familiar principle. This essay considers some ways in which anonymous Pareto is defended and argues against each separately, as well as in more general lines. It suggests that the reasons in light of which people find strict Pareto so compelling are the reasons for doubting the anonymous variation of that principle.

INTRODUCTION

It would probably be better to give John a large benefit rather than a smaller one, if we could do so without making anyone worse-off. Now, a small benefit to John seems to be as good as a small benefit to his equally deserving, similarly situated fellow citizen, *Paul*. And so, if a large benefit to John is better than a small benefit to John, and a small benefit to John is as good as a small benefit to Paul, then it follows that a large benefit to John is also better than a small benefit to Paul, doesn't it?

This way of reasoning has led some to what is sometimes labeled *the anonymous Pareto principle* (AP).[1] On this principle, a distribution *A* is said to be better than distribution *B* in case it is strictly – that is, non-anonymously – Pareto superior to either *B* or a permutation of *B*. A permutation of *B* is a distribution which follows the same pattern as *B*, except that its relevant benefits befall different individuals.[2] For example, a distribution of the form (2, 1) is a permutation of (1, 2): the distribution of benefits in both exhibits the same pattern, with the worst-off having one and the better-off two. It is just that the actual person occupying each position is different. And so, on this principle, (1, 3) is better than (2, 1) because it is strictly Pareto superior to a permutation of it, namely (1, 2). By contrast, (0, 4) is not AP superior

[1] Some recognize AP as the 'Suppes-Sen grading principle'. See: A. Sen, *Collective Choice and Social Welfare* (San Francisco, 1970); Patrick Suppes, 'Some Formal Model of Grading Principles', *Synthese* 6 (1966), pp. 284–306.

[2] See e.g. Michael Otsuka, 'Saving Lives Moral Theory, and the Claims of Individuals', *Philosophy & Public Affairs* 34(2) (2006), pp. 109–35, at 122; David McCarthy, 'Distributive Equality', *Mind* 124 (2015), pp. 1045–1109, at 1059.

Utilitas Vol. 31, No. 1, March 2019
doi:10.1017/S0953820818000134

to (2, 1), since there is no permutation of (1, 2) to which (0, 4) is strictly Pareto superior.[3]

Another way of thinking about this is as a Pareto principle which applies to *positions* in a distribution, or to placeholders, rather than to the particular individuals who are filling them.[4] If for example things go better for the best-off and not worse for any other position, this is to be considered an AP improvement. That is so even if the actual person who was the best-off has changed her position and is now having less than before. A distribution may be anonymously Pareto better without being *strictly* Pareto better.

It is an appealing principle, for two main reasons. One is that it offers a more workable version of the strict, non-anonymous Pareto principle. The strict Pareto principle, according to which a state X is better than another state Y if it is better for at least one person and not worse for anyone, is very rarely applicable. If we carried out only those policies that do not leave anyone harmed, we would likely end up doing nothing at all. If the Pareto principle is to remain a relevant guide to policy, it seems necessary to view it as applying to positions or classes of people, rather than to individual persons. Second, AP holds great promise as a theoretical tool as well. It is a way of deciding among competing interests without trade-offs or aggregation. By appealing to AP, we are able to say that the greater good of some is morally preferable to the lesser good of others without committing to such notions as impersonal or interpersonally additive good. The principle makes use of relatively modest assumptions that seem acceptable to consequentialists and non-consequentialists alike: for example, the assumptions that a strictly Pareto superior state is morally preferable, and that one person's good counts as much as another's.[5] Once these

[3] This feature distinguishes AP from ex-ante Pareto, with which it is sometimes conflated. What is AP superior is also ex-ante Pareto superior, but not vice versa.

[4] Jonathan Dancy, for instance, believes that an equally valid version of the Pareto principle is one that 'holds that one alternative is better than another if in it some better slots are occupied and no worse ones'. See Jonathan Dancy, 'Essentially Comparative Concepts', *Journal of Ethics & Social Philosophy* 1(2) (2005), pp. 1–15, at 9. For a discussion of this idea at work, see Nir Eyal and Alex Voorhoeve, 'Inequalities in HIV Care: Chances Versus Outcomes', *The American Journal of Bioethics* 11 (12) (2001), pp. 42–4.

[5] Though not everyone, most people seem to accept that a change that is good for some and bad for no one, is – probably, or very often, or often enough – a change for the better (some even find it 'hard to see how anyone could resist such a principle'. See J. Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford, 1986), p. 147. There might be exceptions, of course. There might be cases in which the status quo is so unjust that a Pareto improvement would seem morally neutral, or even strike us as a change for the worse. But as a general rule, (strict) Pareto improvements are taken to be non-controversially desirable: improvements that carry no moral cost; a case of all reasons counting in favour of the change in question and none against.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

assumptions are accepted, we seem to be compelled to reach certain moral judgments we might otherwise consider controversial. Thus, the principle has been used, for example, to explain why it is better to save A and B over saving just C. Saving A and B is strictly Pareto superior to a permutation of saving just C – say, saving just A – and is therefore AP superior.[6] AP has also been employed in the context of the non-identity problem. A distribution of the form (2, Ω) – in which one person is enjoying some level of welfare, 2, and the other does not exist – is better than (Ω, 1), because it is better than its permutation, (1, Ω).[7] AP purports to allow us – indeed, commit us – to make these judgements without appealing to utilitarianism, prioritarianism, or other good-maximizing theories.

The idea, then, is highly attractive. AP promises to improve on the more familiar, impracticable strict Pareto principle, to provide significant philosophical results using uncostly theoretical resources, and to bypass some of the more intractable disputes among moral philosophers. But how defensible is it? In what follows I shall explore three of the justifications some philosophers have raised in its defence. One is the claim that if strict Pareto, anonymity and transitivity are assumed, AP must follow. The second is that since people might not have a claim to any particular position in the distribution of goods, the Pareto principle may apply to positions, rather than to individual persons. The third, which has been titled *morphing*, is an argument that seeks to demonstrate that Pareto betterness obtains even when the identities of the persons involved are replaced. I will argue against each of these justifications separately (on sections I–III), as well as in more general lines (section IV). I will suggest that the reasons in light of which people find strict Pareto so compelling are the reasons for doubting the anonymous variation of that principle.

I. ANONYMITY

As its name suggests, the most common and straightforward way of motivating AP is by appealing to the notion of anonymity or impartiality. AP, it is argued, is a conjunction of strict Pareto and the idea that morality is impartial between individual persons: if two alternatives are alike in every respect, except that one benefits individual *i* over *j* and the other benefits *j* over *i*, then they are equally desirable.[8] The argument is as simple as it is powerful.

[6] Iwao Hirose, 'Saving the Greater Number Without Combining Claims', *Analysis* 61 (2001), pp. 341–2; F. M. Kamm, *Morality, Morality,* vol. 1 (New York, 1993), p. 85.

[7] J. Broome, *Weighing Lives* (New York, 2004), p. 136.

[8] See e.g. Hirose, 'Saving the Greater Number Without Combining Claims', p. 341.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

Suppose that only one of two patients, A and B, can be spared premature death. Both patients are of the same age and physiologically similar, and neither is any more deserving to live than the other. If we choose alternative (i), A would live another twenty years and B would die. If we choose (ii), B would live another twenty years and A would die. There is also an alternative (iii), in which B lives another *thirty* years and A dies. The alternatives are presented in the following table:

	(i)	(ii)	(iii)
Patient A	Lives 20 years	Dies	Dies
Patient B	Dies	Lives 20 years	Lives 30 years

By anonymity, (i) is as good as (ii). By strict Pareto, (iii) is better than (ii). Assuming transitivity, we should conclude that (iii) is better than (i). This conclusion – that we should favor the patient who is expected to live longer – is one that is not immediately intuitive, or even attractive. To some, particularly non-consequentialists, disfavouring a patient simply because we could do better for somebody else might seem grossly unfair. Some of us strongly believe that patient A still has some moral claim to being treated. Indeed, some of us may believe that the case for treating her is every bit as strong as the case for treating patient B. Yet to deny this conclusion, it appears, we would need to deny either that thirty years of life are better than just twenty, or that twenty years of one patient's life are somehow preferable to twenty years of the other's. The conclusion seems to force itself on those who wish to oppose it.

I shall argue, however, that this conclusion *can* be resisted. And what is more, it is possible to resist it without having to deny strict Pareto or the equal worth of lives. As is often the case, the problem is in the details.

To see where the difficulty is, it is important first to get a clear idea of the exact sense of betterness and equivalence the argument employs. Consider the equivalence between alternatives (i) and (ii). In what sense, precisely, is giving twenty years of life to A as good as giving twenty years to B? After all, these are not equally good for any of the patients concerned. Nor are they Pareto equal (they are in fact Pareto incomparable). It might seem natural to say that the acts are equal in that they result in equally good states of affairs. Twenty life years for either patient would be an equally good outcome in an impersonal, consequentialist sense. Equal from the point of view of the universe, so to speak.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

This may or may not be true, but if that were the sense of goodness we had to assume for the argument to work, then Pareto betterness – anonymous or otherwise – would become redundant. If this were the notion of value with which we begun, we could simply judge that B's living another thirty years is a better outcome than A living twenty. We wouldn't need to be told in addition that thirty years for B is better than a permutation of twenty years for A and so forth. AP is attractive, remember, in that it promises to make such judgements available to those who are not committed consequentialists already. It is meant to do that by showing how these judgments follow from the modest assumptions most people already accept. The sense of goodness the argument presupposes cannot therefore be that of impersonal goodness, of the kind used by goodness-maximizing moral theorists.

If we are to avoid trivializing the argument, or begging the question against the unpersuaded, we should consider the choice between (i) and (ii) as a choice between possible acts, rather than states of affairs, and evaluate these alternatives as such. The problem, however, is that once we view the alternatives in this way, the argument seems to lose much of its sting.

There are some alternative, equally plausible, and *non*-consequentialist ways to think of the equivalence between (i) and (ii). For example, (i) and (ii) may be equivalent in that giving twenty years of life to A is as permissible as giving twenty years to B. If that is the case, it is not immediately clear what follows from the fact that thirty years to B is better than twenty to B. If I am permitted to select either x or y, the fact that some third option, $y+$, is preferable to y does not automatically render x impermissible. Some other substantive claim would need to be added, such as the claim that we are not permitted to select the lesser alternative. Without it, the availability of $y+$ may not have any bearing on the normative status of x at all. Giving thirty years to B is not 'more permissible' than giving twenty to A, whatever this might mean.[9] Actions are either permissible or they are not, and there is no reason to assume that saving A ceases to be simply because the option of letting B live longer becomes available.[10]

[9] To say that it is permissible to do x rather than y is not to imply that there must be some aspect in which x outranks y (see also: L. Temkin, *Rethinking the Good* (New York, 2012), p. 196).

[10] Some have interpreted the somewhat elusive notion of parity as a relation of equal permissibility: two alternatives, x and y, are on a par iff it is (i) permissible to prefer x to y and (ii) permissible to prefer y to x (Wlodek Rabinowicz, 'Broome and the Intuition of Neutrality', *Philosophical Issues* 19 (2009), pp. 389–411, at 402). If equally permissible actions are on a par, their relation may not be transitive (Ruth Chang, 'The Possibility of Parity', *Ethics* 112 (2002), pp. 659–88).

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

Or suppose, alternatively, that (i) and (ii) are equally obligatory – that is, suppose it is as obligatory to save A as it is obligatory to save B. If that is so, saving A might remain obligatory even if saving B would do more good. This obligation need not simply dissolve when a better alternative presents itself (again, unless some further assumption is introduced). Whether it dissolves or remains standing depends on the content of that obligation. Let me explain.

Suppose that we ought to save patient A as much as we ought to save B because, and only because, we ought to save the patient who can live the longest, regardless of who it is, and either patient is expected to live another twenty years if rescued. If that were the case, any option that would guarantee twenty-one years or more, for either patient, would be obligatory.[11] It would be the option that satisfies our duty. Saving twenty years of life when more could be saved would be impermissible.

But now suppose, alternatively, that our obligations are more specific. Suppose that we ought to save A as much as we ought to save B because we owe it to A as much as we owe it to B. In that case, our duty is not just to save life in general, but rather to save A's particular life as well as B's. If *that* is the content of our obligation, saving A remains something we ought to do even if it is possible to do more for B. Saving B would mean saving more life, but it would not be A's life. Our obligation to save A's life would remain undischarged no matter how much more we can do for B. The choice between A and B continues to be an unresolved conflict of obligations.

If that is correct, the upshot is that we do not *have* to accept the argument from anonymity. The fact that two conflicting alternatives are equally compelling need not entail that whatever is preferable to the one is also preferable to the other. It is at least sometimes the case that an alternative x and its permutation y are equally desirable, and that z is strictly Pareto better than y, and yet there is no clear way to settle the choice between z and x. The argument does work, albeit somewhat trivially, if the equivalence in question consists in the equal value of the respective outcomes, and that alone is the morally relevant characteristic of either alternative. It might not work in case the equivalence refers to the choiceworthiness of the actions themselves.

[11] Many people feel uncomfortable accepting that trivial differences, such as the difference between twenty years and twenty-one, should tip the balance in cases such as this. Some explain that such a minor difference is an 'irrelevant utility' (see, e.g. Kamm, *Morality, Morality*, p. 146). Others think that a trivial improvement is not enough to outweigh the badness of treating the other patient unfairly (see I. Hirose, *Moral Aggregation* (New York, 2014), ch. 8). The merits of these lines of argument should better be debated elsewhere. I am inclined to agree that given certain assumptions, twenty-one years might not be morally preferable to twenty. But my own reason for thinking so is different, as I explain below.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

But there is more to be said. It is not only that the argument *might* fail. It seems to me that once we consider its full implications, we should rather hope that it does.

For suppose that the argument does work, and thirty years of B's life are preferable to twenty of A's. Are those thirty years of B's life preferable to twenty of A's *as much* as they are preferable to twenty of B's? And if not, why not?

Supporters of AP, particularly those who are more formally inclined, might protest that the argument from anonymity is only committed to claims about 'better than', not about 'how much better'. Some think that transitivity in general can only yield an ordinal not cardinal ranking of alternatives.[12] Clearly, in cases in which all we know is that X is better than Y and Y is better than Z, the most we are entitled to infer is that X is better than Z. Nothing about the extent to which it is better follows from transitivity. The case at hand, however, is different. If one option, A, is said to be better than B, and B is said to be as good as C, then, assuming a single dimension of betterness, A should stand in the same relation to B as it does to C. If Jane is in the market for a new car, and finds car x to be as good as car y in every respect, then she would have to find car z, which is preferable to y, to be preferable to y exactly as much as it is to x. She would have to be just as happy to receive z instead of y as she would to receive z instead of x. Why wouldn't she? If the relation between x, y, and z is indeed transitive, there is no reason why this shouldn't be the case. Accordingly, if the relation between the alternatives in the example above is transitive, as the argument assumes, we should be expecting the same thing. We should be expecting that thirty years of B's life would be as preferable to twenty of A's life as they are preferable to twenty of B's.

But if this in fact follows from the argument, it is a consequence that makes its antecedent very hard to accept. If we could make things better for B at no cost to anyone else, then, other things equal, we would have an extremely compelling, arguably conclusive reason to do so. It would be a case of all moral reasons pointing towards helping patient B to live longer, without any reasons counting against it. By contrast, helping B to live longer while A could be helped instead, even if not as much, is a harder choice to make. Patient A's avoidable premature death gives us at least some reason, however weak, against preferring B.

The extent to which thirty life years to B are morally preferable to just twenty life years to B is, therefore, *greater* than the extent to which thirty years to B are morally preferable to twenty years to A. Choosing

[12] I am grateful to Ittay Nissan-Rozen for pressing me on this point.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

to treat B over A involves a moral cost that merely extending B's life does not. In preferring B over A, we condemn A to dying when instead she could live. The other choice – between a state in which B lives longer and A dies, and a state in which B lives less long and A dies – does not involve such moral cost, as patient A's fate is the same no matter what we choose.[13]

Nothing I have said so far suggests that favouring the patient who would live longer could not be justified on some other grounds. But if it could, we should expect that such a preference be justified on an all-things-considered basis, after having given patient A's loss some moral weight. And this may suggest, I think, something fundamentally problematic about the argument from anonymity, and perhaps about the very idea of AP. For what the argument does is precisely to equate hard moral decisions – decisions that typically involve real moral costs – to trivial choices between a strictly Pareto superior option and its inferior alternative. This can only be true if the losses to those who are made worse-off do not matter, or are somehow made good by someone else's gain. Indeed, an AP improvement can resemble a strict Pareto improvement – that is, can resemble a change that is good for some and worse for no one – only if we give individual losses no weight at all. That, if nothing else, is reason enough to be suspicious.

II. DO PEOPLE HAVE A CLAIM TO A PLACE IN THE DISTRIBUTION OF BENEFITS?

Now, some supporters of AP might not be moved by the problem I have just criticized: I have taken it for granted that individual losses must carry a certain moral weight. But what if they do not? Consider the following case. Suppose that we could increase doctors' annual pay without adversely affecting any other sector in society. However, if we choose to do so, admission to medical school will become more competitive, and some people who would otherwise become doctors will have to settle for some lesser trades. Increasing doctors' wages will have made these people worse-off, and so it does not count as a strict Pareto improvement. Yet it seems to be as good as one. 'Doctors', whoever these may turn out to be, are made better-off, while everyone else – again, whoever that may be – is not harmed. Intuitively, this seems to be all that matters in this case.

According to a second argument for AP, this intuition is justified by the fact that even though the change is bad for some, their loss is not

[13] For an excellent discussion of a similar point in a slightly different context, see Weyma Lübbe, 'Taurek's No Worse Claim', *Philosophy and Public Affairs* 36 (2008), pp. 69–85.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

a loss that counts morally. A doctor's career is not anyone's to lose, and therefore no one has a justified complaint against not having it. Nothing is lost, morally speaking, by shifting from a state of affairs in which a particular person is employed as a doctor to one in which somebody else is. This might not be a case of 'better for some and not worse for anyone else', but it may well be a case of 'better for some and not *objectionably* worse for anyone else'. All moral reasons seem to favour the AP outcome, as if it were strictly Pareto superior.

Abstracting away from this example, the general idea here is that people might not have a moral claim to any particular place in the distribution of benefits prior to its being set up. At least some of the things we come to enjoy under a certain social arrangement are not owed to us in some deep, pre-institutional sense.[14] We are entitled to having them only in virtue of that social arrangement, and we have a claim to having that arrangement in place only in so far as it is morally warranted. If it is morally preferable to have a different arrangement instead, we have no legitimate complaint in case that second arrangement entitles us to less than we would have under the first one. We do not have a valid complaint, that is, in case we do not happen to inhabit the position we hoped to. We are not made *objectionably* worse-off.

If this is correct, then what makes a given arrangement better than another is not its relative effect on this or that individual person, but the better distribution, or better positions it offers to whoever will come to occupy them. Accordingly, the Pareto principle should no longer be understood in its strict sense, as a rule that concerns the well-being of specific individuals. It should be taken to apply to the levels of well-being that are associated with positions: to how well the 'worst-off', the 'second worst-off' or the 'best-off' are doing, whoever these turn out to be.[15] A change that is better for at least one place in the distribution and worse for none is a change for the better. That is so even if some of the particular people who turn out to be occupying these positions fare worse than before.

This is quite plausible, as far as it goes. If you and I were to agree on an acceptable distribution of some benefit to which none of us has a prior claim, an AP optimal arrangement would be a safe choice. There is no compelling reason why, in that case, we shouldn't prefer

[14] The idea that people might not have a pre-institutional claim to what they have is famously associated with Rawls (J. Rawls, *A Theory of Justice*, (Cambridge, MA, 1971), pp. 103–4), though AP cannot be defended on strictly Rawlsian grounds (for more on this see n. 16 below).

[15] This argument is due to Alex Voorhoeve ('Should Losses Count? A Critique of the Complaint Model', *LSE Choice Group Working Papers* 2 (2006), sect. 4). I am unsure to what extent he himself is still committed to it, but I find it very well worth considering.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

a distribution such as (3, 1) to (1, 2). Wherever we turn out to be located within this distribution, the first pattern is at least as good as the second for each of us.[16] However, the assumption that none of us has a prior claim is something that needs to be discussed. We may agree that individual losses do not always count, and when they do not, AP becomes plausible. The question is whether individual losses *ever* count morally. If the answer is 'no', AP is fully vindicated. But is there a reason to believe that the answer is 'no'?

We may not reach a consensus on what moral claims people have, or whether they have any at all. These are highly contested questions. What is clear, however, is that proponents of AP cannot simply *assume* that losses never count. In a sense, the notion that individual losses are unobjectionable is what AP is supposed to establish. It is used precisely in order to persuade those who believe that losses do matter. As was mentioned earlier, part of the motivation behind AP is to offer a principle that everyone, including non-consequentialists who believe in individual moral claims, could accept. If we all agreed that nobody is owed anything, there wouldn't be a need for AP to begin with.

Intuitively, many of us think that at least some goods *are* owed to people, even before a distribution has been put in place. This is the case when a person has an interest at stake which is weighty enough to hold others under a duty. The interest in living is a paradigmatic case. It is common to assume that the weight of that interest generates *pro tanto* negative duties to refrain from taking a life, as well as some *pro tanto* positive duties, such as the duty to assist a person in grave danger if one can do so within reasonable means. We think that a patient is wronged, or has a justified complaint, if we choose not to treat her, and therefore that failing to save a life is a loss that should count. Proponents of AP need not think otherwise. They may agree, in principle, that there are various things that people are owed.[17]

[16] Perhaps this is not as obvious. If our choice is guided by a Rawlsian maximin principle, we might not prefer the AP superior outcome. The distribution (3, 1) is AP superior to (1, 2), but it may not be better according to maximin, it might even be worse. Since the increased inequality involved in shifting from (1, 2) to (3, 1) does not work to the advantage of the least well-off, the change might be unjust (Rawls, *A Theory of Justice*, p. 151). By comparison, if what we are after is maximizing expected utility, any AP superior distribution would be considered better, but, of course, not vice versa. For instance, we would prefer (3, 0) to (1, 1). Hence, even in the absence of pre-institutional claims, we might either not prefer a Pareto superior distribution – for example, if it increases inequalities in a way that does not benefit the worst-off; or we might prefer a Pareto superior option only incidentally, as it is one among a number of ways to maximize our expected utility.

[17] What if one of the things that people are owed is equal consideration? Does that undermine the case against AP? The answer is 'no'. As I hope to have shown in section I, the fact that people's good counts equally does not entail that whatever is better than benefiting one of them is also better than benefiting the other. Moreover, it

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

More plausibly, the argument for AP rests on the following thought: when considered in isolation, it may be true that every single person is owed all sorts of benefits, such as, for example, having her life saved. But the contexts in which AP is employed are those in which we need to decide among multiple, competing demands. If we cannot save all the lives that need saving, and if 'ought' implies 'can', then it is not true that we *ought* to save all these lives. Hence, if either A or B can be rescued though not both of them together, *neither* patient has a claim to be the one who is helped. Under these circumstances, the life we will not save will not be a life we *ought* to save, and therefore the loss is not a loss that should count. As no individual claims are involved, we should focus on the assistance itself rather than on *who* is being assisted. It would be right to decide in a way that is better for whoever is rescued and not worse for whoever is not, even if that means that one of the individual patients will be adversely affected by such a choice, or so the argument would go.

This line of reasoning is particularly common among those who are sceptical about the significance of moral claims. To these sceptics, mutually exclusive claims, none of which has privileged status, are as good as no claims at all. But this is a mistake. First, as Jeremy Waldron has once argued, the fact that our duties towards different persons are not *compossible* does not mean that we are unable to meet each one of them.[18] In rescue cases such as the one we have been considering, the choice is hard precisely because we *can* rescue each of the patients but cannot rescue both. Therefore, even though 'ought' implies 'can', moral claims need not dissolve once they are in conflict. Each life that is lost is a loss that could have been prevented.

Second, this sort of scepticism only appears persuasive given a particularly narrow view of what we owe to people, and of what moral claims in fact are. On this narrow view, a moral claim such as the claim to a life-saving treatment corresponds to the single duty to provide this treatment. But as is often recognized, claims may generate an array of different duties. We might not always be able to satisfy everyone's claim to medical treatment. But there may well be other things we ought to do. We may owe compensation to people whose claim we failed to satisfy, for example.[19] We might have a duty to make treatment more available, at the expense of other interests.[20] We would also be

is important to stress, in this context, that people have a right to *due*, as well as equal, consideration. Their interest ought to be appropriately addressed, in absolute not just comparative terms. I thank the anonymous reviewer for this point.

[18] Jeremy Waldron, 'Rights in Conflict', *Ethics* 99 (1989), pp. 503–19, at 506.

[19] Joel Feinberg, 'Voluntary Euthanasia and the Inalienable Right to Life', *Philosophy and Public Affairs* 7.2 (1978), pp. 93–123.

[20] Waldron, 'Rights in Conflict', p. 512.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

required to treat everyone should this suddenly become possible – a requirement we might not have if all patients had no moral claim, as the sceptics argue. The fact that we are not able to satisfy every claim of every individual might mean that each person is owed less than she otherwise would. But it does not mean that she is not owed anything. Something of moral significance often *is* lost when our policies are not strictly Pareto better.

In light of these considerations, I submit that the present argument does not vindicate AP. In those choices in which a morally significant loss is believed to be at stake, the argument does not give us a reason to disregard it and shift our focus from individual persons to positions. However, this argument for AP is nevertheless significant. It is important because it shows why AP may be valid *in some cases*, namely those that do not involve individual moral claims. This is a desirable result for AP's proponents *and* opponents.

I have said earlier that one important motivation for adopting AP is to render strict Pareto more practicable as a guide to policy. We rarely face a decision that is not bound to be bad for someone. The different policies among which we choose often benefit different particular individuals. The way to save the Pareto principle from irrelevance is to allow it to apply to types or classes of people, rather than to individual persons. It would be an undesirable consequence for the case against AP if it turned out to imply, in fact, that strict Pareto should be abandoned as well, even if for pragmatic reasons.[21] The discussion in this section has shown that there are contexts in which applying AP can be justified, and in which it could be a useful principle for guiding policy-making. At the beginning of this section we have seen an example of the sort of decisions for which this is true: paradigmatically, these would be policies that apply to classes or positions to which individual moral claim are yet to be established.

III. MORPHING

The third and last argument in defense of AP I would like to consider is Caspar Hare's rather ingenious idea of *morphing*.[22] This argument is designed to address a broad range of problems in moral philosophy. For convenience and brevity, I will limit my discussion to its most basic form.

Suppose that we are to decide, as before, between (i) an outcome in which patient A is saved and lives another twenty years while patient B dies, and (ii) an outcome in which B is saved and lives

[21] I am grateful to the anonymous reviewer for suggesting this point to me.

[22] Caspar Hare, *The Limits of Kindness* (Oxford, 2013).

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

another *thirty* years, while A dies. Outcome (ii) is of course the AP superior alternative, though it is not strictly Pareto superior. In order to establish that this is the better choice, according to the argument, all we need to accept is strict Pareto, transitivity – and some relatively mild assumption about personal identity – namely, that it is only *somewhat* fragile. Let me explain.

Patients A and B are distinct persons. That is to say, a person who is like A in every respect cannot be said to be B. However, A could be ever so slightly different and still be A. What is essential to his identity is *somewhat fragile*: a person who would be very different would no longer be A, but not every difference is sufficient. There could conceivably be several slight variations of a person that could all be properly considered to be A.

Now, we can imagine a sequence of possible worlds, stretching from a world in which (i) is the case, to one in which (ii) is. In each of the intermediary worlds, each patient is ever so slightly different from her counterpart person in the preceding world, until A ceases to be himself and becomes B, and B, similarly, becomes A. Finally, in each possible world the surviving patient lives a little longer than the one in the previous link in the chain. The resulting picture is the following:

W_1 …	W_2 …	$W_{n\text{-}1}$ …	W_n
Lives 20 years $_{\text{Patient A}}$	Lives longer $_{\text{A slightly different A}}$	Lives longer still $_{\text{A slightly different B}}$	Lives 30 years $_{\text{B}}$
Dies $_{\text{Patient B}}$	Dies$_{\text{Slightly different B}}$	Dies $_{\text{A slightly different A}}$	Dies $_{\text{A}}$

If we accept strict Pareto, we should prefer each world to the one to its left. At each step, one person is better-off while the other's fate is the same. If we also accept transitivity, so the argument goes, we should conclude that the last link in the chain is preferable to the first: that is, W_n is preferable to W_1.

To avoid a potential misunderstanding, it is important to look closely at what gets the argument going. One might wonder as follows: A and B are either the same person or they are not. If they are, then it is easy to see how the strict Pareto dominance relation persists all the way from W_1 to W_n. But if that is the case, then the argument turns out to be unnecessary. We could simply say that W_n strictly Pareto dominates W_1 right from the start. If, however, A and B are distinct persons, it seems mysterious that strict Pareto dominance should hold through each and every link in the chain. After all, there must be some pair of neighbouring worlds that are Pareto incomparable – namely, the last world in which the living patient is still A, and the first in which it is sufficiently B.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

So on a more coherent reading of Hare's argument, there isn't such a pair of worlds. Instead, what we have to assume is some intermediary world, in which each patient is sufficiently like A *and* sufficiently like B. That is, along the way leading from A to B (B to A), there is a person who can be considered both A and B at the same time. That person can be said to be better-off than the nearest A, and worse-off than the nearest B:

W_1	W_2	W_3
Lives longer$_{A}$	Lives longer still$_{A\&B}$	Lives even longer$_{B}$

This reading allows strict Pareto dominance to hold between each of two neighboring worlds. Each of the worlds is better for one person who exists in the world preceding it and no worse for anyone else. But note, however, that this reading has a quite significant implication. Hare does not say that identity is intransitive.[23] Yet if this reading is indeed the correct one then what we are asked to assume is something rather very close to it. A is thought to be sufficiently the same as the counterpart who is like both A and B, and that person is thought to be sufficiently the same as B, yet B is by no means the same person as A. I do not know whether this, in and of itself, is a problem for the argument. It could be objected that morphing is purporting to solve a problem in moral philosophy by creating one for metaphysicians, but I am willing to assume that this can be made sense of. My own objection to the argument is different: what makes each world preferable to its predecessor is that between each pair of worlds there is one specific person for whom things are better. But if it turns out that personal identity is not transitive, this preferability relation need not be transitive either.

Transitivity, we know, applies to those things that can be ordered across a single dimension, or a common linear scale.[24] For example, the relation 'taller than' is transitive because we are able to rank objects on a single scale from the shortest to the tallest, such that anything taller than one thing must be taller than anything ranking below it. That is, there is no way of being taller than some object without thereby being also taller than whatever is shorter than that.

[23] Hare's argument does not commit him to an intransitivity of identities in the strictest sense, because on his account each world is populated by a *counterpart* of the person in the world preceding it, rather than by the same person. He considers numerical identity to be 'paradigmatically transitive' (Hare, *The Limits of Kindness*, p. 129).

[24] Temkin, *Rethinking the Good*, ch. 6.

But the relation 'being preferable to' is different. There are multiple ways in which one thing can be preferable to another. Career *x* might be preferable to career *y* because it pays better, and *y* to *z* because it is more fulfilling. In this case, there is no compelling reason to think that *x* is preferable to *z*. That is not because of some failure of transitivity, but simply because this is not the kind of case to which transitivity is supposed to apply in the first place.

Now, it would seem as though the sort of preferability the argument employs *is* a single-dimensional relation. After all, each possible world is preferable to its predecessor in virtue of what might be thought to be the same thing – i.e. 'being strictly Pareto superior'. Strict Pareto dominance does seem, intuitively, like a transitive relation. When all individual persons remain the same (and personal identity is transitive), there is no reason to think otherwise. Any distribution that is strictly Pareto superior to some state of affairs S would be strictly Pareto superior to whatever is strictly Pareto inferior to S.

But as a matter of fact there is a sense in which the relation 'strictly Pareto superior to' is quite unlike relations that are paradigmatically transitive. Consider this: the state of affairs (1, 1) could be strictly Pareto improved to either (2, 1) or (1, 2). Both these improvements are strictly Pareto superior to the same state of affairs, (1, 1), possibly to the same extent, and yet they are not (Pareto) comparable to each other. That is, neither of them is strictly Pareto superior to the other, nor are they equal: there might be further improvements that are strictly Pareto superior to one without being strictly Pareto superior to the other. This would not be possible if strict Pareto dominance were single-dimensioned in the way that, say, tallness is. If two persons are both taller than a third, their height cannot be incomparable. This is just what it means to have a single scale of tallness. The incomparability of (1, 2) and (2, 1) makes sense only if we grant that strict Pareto dominance is not, in and of itself, a single dimension of betterness.

Now consider the relation 'being preferable to' in Hare's morphing sequence. A relation of strict Pareto dominance does indeed hold between each pair of neighbouring worlds. But this relation, as we have just seen, need not be one-dimensional. And indeed it isn't: up to a certain point, each world is superior to its predecessor in virtue of being better for A, whereas after that point the worlds are getting better in virtue of being better for B. These are two distinct dimensions of being preferable, even if both can be described as 'being strictly Pareto superior'. We have no special reason, as far as I can see, to expect transitivity to obtain.

To sum up, we considered two ways of reading the argument for morphing. On the first reading, somewhere along the sequence there is

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press

a pair of neighbouring worlds that are inhabited by different individual persons, namely the last one in which the surviving patient is A (and not B), and the first in which it is B (and not A). The problem with this reading is that strict Pareto dominance cannot hold between these two worlds. The second reading assumes personal identity to be (in effect) intransitive. This reading allows strict Pareto dominance to obtain throughout, but at the cost of dropping what makes strict Pareto dominance transitive. In either case, morphing fails to support the anonymous Pareto principle.

IV. PARETO, ANONYMITY, AND ANONYMOUS PARETO

I have addressed but three arguments in defence of AP. There may well be others still. The discussion so far has not shown AP to be an incorrect criterion as a matter of principle, and I am not sure whether this could be shown. However, there is something more general to be said about what these arguments are trying to achieve. They seek to establish judgements of betterness that are identity-insensitive by requiring us to accept both the standard, strict Pareto principle and certain additional assumptions, that motivate moral anonymity. Technically speaking, that is a valid move: accepting strict Pareto is not incompatible with accepting these other assumptions. Yet there is nevertheless a certain sense of disharmony between strict Pareto and anonymity. While the two are consistent with each other, their underlying motivations are not. I should elaborate on this point.

Anonymity is the view that the desirability of decisions or states of affairs is insensitive to how particular individuals are affected. On an anonymity-based moral theory, the mere fact that some particular person, *p*, would fare better in state X than in state Y, counts neither in favour of X nor against Y. Strict Pareto, on the other hand, would be at least partly sensitive to this fact, in the sense that a policy that harms a particular individual cannot be strictly Pareto better.

As I said, this does not yet amount to a contradiction between the two. Strict Pareto merely states that if – though not only if – state X is better than state Y for at least one person and not worse for anyone, then X is better. It does not say that X couldn't be better than Y in some other way as well. If X is in fact worse than Y for someone, or not better for anyone, it could still be better in some anonymous or impersonal sense, without the strict Pareto principle being contradicted. Strict Pareto and anonymity are logically independent, in that each of them may be true even if the other is false.[25] This observation may license the shift from the strict Pareto principle to its impersonal, anonymous variation. One could accept Pareto without committing oneself to caring about individual persons.

However, while there is no outright contradiction in accepting both strict Pareto and anonymity, the two are nevertheless at odds in the following sense: to many of us, strict Pareto seems so plausible, and so much easier to accept than many other moral principles, precisely because we do not readily accept anonymity. Strict Pareto's great appeal is due to the intuition – whether this intuition is justified or not – that it *does* matter morally how particular individuals are affected.[26] Without this intuition, we would not find it as significant as we do that a policy is good for some *without being bad for others*. The fact that no one is affected for the worse is thought to be important in so far as individual losses are believed to count against a policy that causes them.

Some of the proponents of AP may not fully appreciate this point. They attempt to show that a commitment to an impersonalist variation of the Pareto principle can be found in what we already believe when we accept Pareto in its non-anonymous version. But in so far as strict Pareto is intuitive and easy to accept, it is precisely because the impersonalist character of AP is not. Of course, anonymity can be argued for, as demonstrated by the arguments we considered. But theorists who are attracted to the notion that the fate of particular individuals should not matter would do well to argue for that view directly. Combining it with the strict Pareto principle does not strengthen the case for it. If anything, it might weaken whatever appeal anonymity has. To hold that strict Pareto is non-trivial – that it is not superfluous – is to maintain that it matters how individual persons are affected. It cannot be offered as support for anonymous, impersonalist views.[27]

eran.fish@mail.huji.ac.il

[25] A similar point has been made with regard to the compatibility of the Pareto principle and the person-affecting view, or the view according to which a state of affairs can only be good or bad if it is good or bad for someone. See Iwao Hirose, 'Review of Nils Holtug's *Persons, Interests, and Justice*', *Economics and Philosophy* 28 (2012), pp. 98–102, at 101; Bertil Tungodden, 'The Value of Equality', *Economics and Philosophy* 19 (2003), pp. 1–44, at 15.

[26] For a discussion of this point see Larry Temkin, 'Equality, Priority, or What?', *Economics and Philosophy* 19 (2003), pp. 61–87.

[27] I am sincerely grateful to Daniel Attas, Shlomi Segall, David Enoch, Ittay Nissan-Rozen, Ofer Malcai and the anonymous reviewers for their very helpful comments.

https://doi.org/10.1017/S0953820818000134 Published online by Cambridge University Press