Forthcoming in *Ergo*

The Many, the Few, and the Nature of Value

Daniel Muñoz
UNC Chapel Hill

John Taurek argues that, in a choice between saving the many or the few, the numbers should not count. Some object that this view clashes with the transitivity of 'better than'; others insist the clash can be avoided. I defend a middle ground: Taurek cannot have transitivity, but that doesn't doom his view, given a suitable conception of value. I then formalize and explore two conceptions: one context-sensitive, one multidimensional.

1. Introduction

When it comes to the good things in life, more is usually better. Who doesn't want to win more jackpots, strike more things off their bucket list, enjoy more happy decades? Flipped around, who wants *fewer* dear friends, fond memories, and lucky breaks?

No one, of course. There is thus something radically baffling about the core idea of John Taurek's "Should the Numbers Count?" Taurek (1977) grants that more is better when it comes to fungible trinkets and solo pleasures. But he denies that two happy lives are better than one, and insists that it would be a mistake to save the many rather than the few purely on the basis of numbers. The survival of five is not better than the survival of a single other; five people's headaches are not worse than the lone headache of a sixth. When goods are scattered over lives, "more" does not mean "better."

Taurek's view has provoked a number of objections (Halstead 2016; Hirose 2001; 2004; Kamm 1993; 2005: 4; 2007: 32, 51; Kavka 1979; Kumar 2001; Parfit 1978; Sanders 1988; Scanlon 1998: 232; Timmermann 2004; Woodward 1981). I am here to discuss one in particular, which is that Taurek's view cannot be squared with:

Transitivity

If $A \geq B \geq C$, then $A \geq C$.

I.e.: if A is at least as good as B, and B at least as good as C, then A is at least as good as C.

TRANSITIVITY can seem not just plausible but inevitable; it is hard to imagine even schematically how it could fail. The reason why is given by the truism I started with—that more is better (or rather, by something like the converse: that "better" means "more goodness"). If betterness depends on relative goodness, then A ≥ B ≥ C entails that A has at least as much goodness as B, which has at least as much goodness as C. It is natural to want to measure this with numbers, awarding a "goodness score" of 3 to A, 2 to B, and 1 to C. But this guarantees that A ≥ C, since A will have at least as much goodness. (The same follows when some or all of the scores are equal.)[1]

If TRANSITIVITY fails, it seems we cannot think of betterness as arising from goodness. No surprise, then, that Taurek's critics think his view is nonsense, supposing that it really does clash with TRANSITIVITY. Well, does it?

Several philosophers—most notably Weyma Lübbe (2008)—have independently argued *no* (Friedman 2002: Chapter 2; 2009: 8 fn. 8; Otsuka 2004: 420; Wasserman & Strudler 2003: 74). They think we can reconcile Taurek with TRANSITIVITY in the familiar cases. I argue that they are wrong about those cases, but even if they are right, their defense is incomplete (§§2–4). There are further cases where Taurek's view, even seen through Lübbe's lens, is stubbornly nontransitive.

If I am right, Taurek must give up TRANSITIVITY. Does that doom his view? Critics would accuse him of flubbing an axiom of axiology, but I think, in light of recent developments in ethics, that this may be too harsh (§7). At any rate, Taurekians are left

[1] This is Temkin's (2012: 229, 386) argument that the "Internal Aspects View" of value entails TRANSITIVITY.

with a challenging question, with which I conclude (§§5–7): if more isn't always better, what does that mean for the nature of value?

2. A Failure of TRANSITIVITY

Taurek's view is said to violate TRANSITIVITY. How?

Let's start with a clarification. When Taurek denies that five lives are better than one, he is thinking of a case that involves six people. He is not comparing the survival of a group to the survival of a single member. For even Taurek accepts:

> **PARETO**
>
> A > B if A is better for someone than B and worse for no one.[2]

For example, suppose I own a scarce drug that I could split into halves to save a group of two strangers. The drug has no other use; both strangers stand to live nice lives if spared; and no one has any special right to my supply. It would be better, Taurek thinks, to save both rather than saving one and trashing the rest of my drug. That much is secured by PARETO: double survival is worse for no one, and much better for the one who would

[2] Taurek's paper does not discuss PARETO, but there are several reasons for attributing the principle to him. First, he is reported to have endorsed PARETO in conversation (Kamm 1993: Chapter 5, n. 12), and his defenders are happy to go along with the reports (see, e.g., Lübbe 2008: 69). Second, PARETO enjoys bipartisan support, being accepted both by Taurek's nemeses (e.g., Kavka 1979: 292) and allies (e.g., Lübbe 2008). Finally, as I argue in §4, there is a way to derive PARETO from a deeper part of Taurek's ethics: his concern for people as individuals. (If I am indifferent between saving A & B, on the one hand, and saving only A on the other, that betrays a lack of concern for the welfare of a particular individual: B.) My sincere thanks to an anonymous referee for pressing me to say more about Taurek and PARETO.

otherwise have died.

If, however, the choice is between saving the two strangers or saving a third stranger, who needs a full dose to live, Taurek would not say that two beats one. 'More' isn't better here; the PARETO principle goes quiet. Instead, we turn to:

> **TRADEOFFS**
>
> If the only relevant difference between A and B is that A spares one group from each suffering a harm and B spares another (non-overlapping) group from each suffering a similar harm, then A ∼ B.

Where 'A ∼ B' means neither A > B nor B > A.[3] TRADEOFFS tells us that it is not better to save a big group from death rather than saving one other person from death. Although the one is outnumbered, when trading off harms, the numbers don't count.

With this in mind, consider three outcomes, where the survivors are indicated with boldfaced underlining.

A2: **<u>Aaron</u>**, **<u>Alex</u>**, Betty.

A1: **<u>Aaron</u>**, Alex, Betty.

B1: Aaron, Alex, **<u>Betty</u>**.

[3] I am happy to say 'A ∼ B' means that A is just as good as B—i.e., that A ≥ B and B ≥ A. This follows given two plausible assumptions, to which I could help myself: A ≥ A (REFLEXIVITY), and if A ≥ A and B ≥ B, then either A ≥ B or B ≥ A (COMPARATIVITY). For a defense of this latter principle, see Dorr, Nebel, and Zuehl (forthcoming). But cf. Chang (2002) on parity. (I use '∼' in a way that is neutral between Chang's notions of parity and equality.)

If the numbers count, we get a transitive ranking: A2 > A1, B1. The more the better. But Taurek's view seems to get us a nontransitivity. Given TRADEOFFS, A1 ~ B1 ~ A2, and therefore A1 ≥ B1 ≥ A2. But given PARETO, we cannot have A1 ≥ A2, since A2 > A1. That is a failure of the transitivity of '≥', and of '~'.[4]

Let me break this down. For Taurek, saving one stranger is as good as saving another (A1 ~ B1). Indeed, since the numbers don't count, saving one is as good as saving two others (A2 ~ B1). But Taurek agrees that it is better to save a group of two rather than a single member thereof (A2 > A1). This delivers the nontransitivity. Saving only Aaron is as good as saving only Betty, which is as good as saving Aaron plus Alex, but saving only Aaron is *not* as good as saving him and Alex.

The same argument is given in different keys. Kavka (1979: 291–93, 294 n. 7) says Taurek must give up either PARETO or TRANSITIVITY. For Hirose (2001), who draws on Kamm (1993: 85–87), TRANSITIVITY is assumed *sotto voce* in the case against TRADEOFFS.[5] Either way, the song is the same. Given TRANSITIVITY, Taurek is in trouble.

3. Lübbe's Loophole

Lübbe thinks that Taurek is not really in trouble, because there is not really any tension between his view and TRANSITIVITY.

---

[4] In particular, this violates what I call Transmission Over Ties: A2 > A1 ~ B1, but not A2 > B1. (See Muñoz 2021; Sen 2017: Chapter 1* calls this "PI-transitivity.") Kavka (1979: 291–93) focuses on the transitivity of '~', which is a weaker principle.

[5] Hirose's (2001: 341) argument quietly invokes PI-intransitivity—a weakening of TRANSITIVITY; see Footnote 4, above. From PARETO, we get A2 > A1. By a principle Hirose calls "impartiality," we get A1 ~ B1. And then by PI-intransitivity, we get A2 > B1, which means that TRADEOFFS is false and the numbers count. For more on this argument, see Hirose (2004: 68–69), Kamm (1993: 85; 2005: 4; 2007: 32, 51), Liao (2008: 450).

How could that be? Recall our outcomes (where the survivors are in bold and underlined, and the deceased are in normal font):

A2: **<u>Aaron</u>**, **<u>Alex</u>**, Betty.

A1: **<u>Aaron</u>**, Alex, Betty.

B1: Aaron, Alex, **<u>Betty</u>**.

Lübbe (2008: 80) thinks Taurek can transitively rank these: A2 > B1 > A1.

But why should B1 be better than A1? Isn't saving one just as good as saving another? Here's the key idea: saving A1 is not *just* saving Aaron. It is saving him while *gratuitously letting Alex die.* Aaron and Alex only need half a dose each to live. If you save only one, you are not making the most of your lifesaving drug. As Lübbe asks:

> how can we claim that [B1] and [A1] are morally equal when in choosing [A1] we decide deliberately to watch [Alex] die and waste a resource that could have been used to save her, while in choosing [B1] we do no such thing, since there is no resource left to save [Alex] when we save [Betty]? (2008: 80)

Nor is it just Lübbe. Wasserman and Strudler (2003: 74) argue for the same view from another angle (see also Friedman 2002: Chapter 2; 2009: 279, n. 8; Otsuka 2004: 420). They would say that Alex's death has a different "moral significance" in the context of A1 than it does in the context of B1. The failure to save Alex is a stronger mark against A1, since there it counts as "the gratuitous waste of a life" (Wasserman & Strudler 2003: 74).

The key fact, for these authors, is that you are needlessly, indefensibly letting someone die by choosing A1, whereas you are not by choosing B1. Hence: A1 > B1.

4. Closing the Loop

Lübbe, Wasserman, Strudler, Friedman, and Otsuka all claim that Taurek's view can be reconciled with TRANSITIVITY in certain cases, like the choice from {A1, A2, B1}. Their basic idea is simple. Even if two options allow the *same number of deaths*, one can be worse than the other if it involves more *gratuitous death*. If an option is gratuitously bad—like A1, which saves a subset of the people saved by A2—that itself makes the option decisively worse. I have three objections to this view.

First, gratuitous badness is not obviously a dealbreaker.[6] Let me illustrate with another case. Suppose 100 people, including Aaron, each need 1% of your drug to survive; Betty needs a whole dose. Your options are:

A100: <u>**Aaron**</u>, <u>**99 others**</u>, Betty.

A99: Aaron, <u>**99 others**</u>, Betty.

B1: Aaron, 99 others, <u>**Betty**</u>.

(Reminder: underlined boldface indicates survival.) On Lübbe's interpretation, Taurek would transitively rank these: A100 > B1 > A99. That is surprising. By choosing B1, you are allowing 100 people to die. By choosing A99, you are allowing only two to die. But for

[6] I think Kavka (1979: 292) makes this point, though Wasserman and Strudler (2003: 74) interpret him differently.

Lübbe, it is worse to let the two die, since one of the two deaths is *gratuitous*. This strikes me as odd—and not because I am counting numbers. It is puzzling to me why gratuitous death should be any worse than the usual kind.

To be clear, I agree that gratuitously bad options like A99 are *wrong*.[7] It is hard to justify gratuitous badness, and an option is wrong if it cannot be justified (Horton 2017: 96). Saving the 99, for example, cannot be justified over the alternative of saving the 99 along with Alex. There is no reason why, given that you are saving the 99, you should waste the last 1% of your drug and let Alex die.

I am just skeptical that wrong options (like A99) must always be worse than permissible ones (like B1). What makes A99 wrong isn't how it compares to B1, but to A100. So why should A99's wrongness affect how it compares to B1 pairwise? Some writers share my skepticism (see Pummer 2019). That said, we are outnumbered, so I won't lean on this point. Let's grant Lübbe her claim that, given the option of A2, B1 > A1, and I'll grant B1 > A99 in the case above.

My second objection is that, even if this proposal isn't problematic in itself, it may not fit *Taurek*. Taurek is famously wary of the concept of "good outcomes," unless it can be understood in terms of what is better or worse from particular people's points of view. In a well-known passage, he writes of a choice between saving David—someone he knows and likes—or saving five others:

[7] This claim is often made about instances of "suboptimal supererogation," such as the choice to rush into a burning building and save only one of the two kids inside. See Horton (2017), Parfit (1982: 131), Pummer (2016; 2019), Rulli (2020). I give my own take in Muñoz (2021).

> I do not wish to say in this situation that it is or would be a worse thing were these five persons to die and David to live than it is or would be were David to die and these five to continue to live. I do not wish to say this unless I am prepared to qualify it by explaining to whom or for whom or relative to what purpose it would be a worse thing. (1977: 304)

Now consider the idea that A1 (saving Aaron) is worse than B1 (saving Betty) because, although Alex dies either way, choosing A1 means her death is gratuitous. *For whom* is gratuitous death worse than a non-gratuitous death? It's not worse for the people uninvolved, or for the agent. That leaves only the person who dies—but their loss is the same no matter if it was avoidable.

The special badness of gratuitous death seems to only make sense as a kind of "impersonal" badness, irreducible to what is good and bad for particular individuals. Since Taurek is skeptical of impersonal badness, I think he would be skeptical of the principle that gratuitous death is particularly bad.

Someone might object here that, as I am reading Taurek, there is one case where he *does* accept impersonal badness: the PARETO principle, which says that an option is better if it is better for someone and worse for no one. I think Taurek can and should accept PARETO (in rescue cases where no one's rights are violated, and where harms are of the same size). For example, he should accept that A2 > A1. But isn't this an impersonal value judgment, since it compares two situations involving multiple people?

Not necessarily. *Some* philosophers believe PARETO because they think it's impersonally better to produce the greater and greater globs of utility—no matter whose.

But there is another route to PARETO, one that starts from a concern for people as individuals. If A2 is better for someone than A1, and worse for no one, then A2 will be preferable to anyone who cares about that "someone." In this case, concern for Alex—plus the fact that her survival is the only relevant difference between A1 and A2—is what makes A2 the better choice. In this way, PARETO can emerge from Taurek's concern for each individual personally rather than their sum. That is why I think Taurek should accept PARETO, though not the impersonal principle that gratuitous deaths are worse. And it is this principle about gratuitous death that Lübbe needs to rescue TRANSITIVITY.

I've just argued that the appeal to gratuitous badness is hard to square with Taurek's views, and that it may be questionable in itself. But I've saved my main objection for last. Even if we grant Lübbe (and the others) everything they say about the cases above, where one option involves gratuitous death, that won't be enough. Taurek still cannot have TRANSITIVITY in cases where *three* options involve gratuitous deaths.[8]

Suppose that three people—Aaron, Alex, and Alice—each need a third of your drug to survive, whereas Betty and Boris need only half each. You have five options:

A3: **<u>Aaron</u>**, **<u>Alex</u>**, **<u>Alice</u>**, Betty, Boris.

A2: **<u>Aaron</u>**, **<u>Alex</u>**, Alice, Betty, Boris.

A1: **<u>Aaron</u>**, Alex, Alice, Betty, Boris.

B2: Aaron, Alex, Alice, **<u>Betty</u>**, **<u>Boris</u>**.

B1: Aaron, Alex, Alice, **<u>Betty</u>**, Boris.

[8] Graham (2017: 131) uses a related pair of cases for a different purpose; he argues that Taurek is committed to a nontransitivity involving degrees of wrongness.

Given Tradeoffs, we know that A1 ~ B1 ~ A2, and so A1 ≥ B1 ≥ A2. Given Transitivity, we infer A1 ≥ A2. And yet, Pareto yields A2 > A1. Taurek's views here are inconsistent with Transitivity, and this time, there are no loopholes.

Let's unpack this. Why should Taurekians have to say that A1 ~ B1 and B1~ A2? Sure, both comparisons involve tradeoffs of like harms across different groups. But the Tradeoffs principle only kicks in when there is no other relevant difference. Isn't it relevant here, as before, that the options involve gratuitously letting people die?

But this is not a *difference* between A1 and B1. Both options involve wrongfully, gratuitously letting people die. To be sure, the numbers are different; A1 involves gratuitously letting two die (Alex and Alice), whereas B1 entails only the gratuitous death of Boris. But this can't matter for Taurekians—the numbers don't count![9]

The result is that A1 and B1 are, morally, a wash. They save different people and forsake different people. But by Taurek's lights they are symmetric in the ways that matter for moral betterness. Both A1 and B1 save a group (size: 1); both gratuitously forsake a group (sizes: 2 and 1, respectively); both non-gratuitously forsake a group (sizes: 2 and 3, respectively); and both enjoy the same deontic status (namely: *wrong*). By similar reasoning, we can show that A2 and B1 are a wash, as well.

And with that, we have all we need for a clash with Transitivity. From the argument above, we have A1 ≥ B1 and B1 ≥ A2. Transitivity would say that A1 > A2, but this can't be, given Pareto. (Again: A2 is better for Alex, worse for nobody.)

[9] This point still stands even if we are moderate Taurekians (see §7), who think big numbers count. We are dealing with one gratuitous death vs. two, not one vs. a billion.

5. Is Nontransitivity Nonsense?

Taurek's TRANSITIVITY problem, I've argued, is real and deep.[10] There is no loophole, no easy way out. Lübbe and others try to save TRANSITIVITY by insisting that gratuitously bad options belong at the bottom of the betterness ranking, which turns a nontransitive jumble like A2 > A1 ~ B1 ~ A2 into a neat ordering like A2 > A1 < B1 ~ A2, with A1 demoted below B1. But this move cannot work across the board, because it can't be used if all three options in the nontransitivity are gratuitously bad, as in the choice from {A1, A2, A3, B1, B2}.

What are Taurekians to do?

Their only option, I think, is to give up TRANSITIVITY. This will not be easy. Not only does TRANSITIVITY have plenty of defenders (e.g., Binmore & Voorhoeve 2003; Nebel 2018); it is so simple and natural that any departure from it can seem like nonsense. Recall the argument I gave earlier (due to Temkin 2012: 229, 386): it seems obvious that better options are those with more goodness, and that an outcome's goodness can be measured with a number, but given these assumptions, TRANSITIVITY follows straightaway. To resist the argument for TRANSITIVITY, we must lose a truism about the nature of value—which threatens to make nonsense of axiology.

We can make this "nonsense argument" precise. Let 'V' be a set of values $\{v_1, v_2, \ldots\}$ assigned to options using a value function V, which can be relative to comparisons. '$V_B(A)$'

---

[10] In this paper, I focus on harms of the same size. Otsuka (2004: section III) argues that Taurek's view entails cycles of '>' in a case where sizes of harm vary. But he assumes that a "serious harm" is worse than a harm that is "less serious to a nontrivial degree," even if they fall on different people (2004: 414). Taurek (1977: 302) would deny this; he does not think my death, e.g., is worse than the loss of your arm (as emphasized by Liao 2008: 452 n. 26; see also Doggett 2009; 2013; Setiya 2014). In light of this, I think it is still an open question whether Taurek should be worried about Otsuka-style cycles. For more on Otsuka's cycle, see Cohen (2014), Kamm (2005: 19–23), Meyer (2006).

gives the value of A when compared to B. To say that one value is higher than another—say, v_1 is higher than v_2—we write '$v_1 >> v_2$'.

The argument's premises are:

SCORES

A > B *iff* $V_B(A) >> V_A(B)$.

Informally: to be better is to have a higher value.

INTERNAL SCORING

$V_B(A) = V_C(A)$.

Informally: a thing's value stays the same no matter what it is compared to.

1D SCORING (STRONG)

$V = \{x: x \in \mathbf{R}\}$ and $v_i >> v_j$ *iff* $v_i > v_j$.

Informally: a value can be represented as a single real number.[11]

Which together entail TRANSITIVITY.

(Here is why. If TRANSITIVITY fails, then we can have A > B > C, though not A > C. Given SCORES, INTERNAL SCORING, and 1D SCORING, this would imply that there are three numbers with the same structure: a > b > c, though not a > c. But this is impossible: the '>'

[11] The argument would still work given the weaker claim that there is a homomorphism from values to numbers—i.e., there is *some* way to slap numbers on the values and swap out 'better than' for 'greater than'. I focus on the strong claim because it is more natural.

relation ('is greater than') is transitive.)

Do Taurekians have a response to the nonsense argument? Yes. We find it in the work of Alex Friedman (2009: 280–84)—who is also the first to make Lübbe's move (in his unpublished 2002: Chapter 2).[12] Friedman denies INTERNAL SCORING: he does not think we can measure how good a thing is with a fixed number, because he thinks a thing's goodness can depend on what we compare it to.[13]

The simplest way to fill out Friedman's view is to keep SCORES and 1D SCORING, which results in what I call the 1D ESSENTIALLY COMPARATIVE VIEW.[14] On this view, a thing's goodness is measured by a single number that may change as we swap out alternatives. For example, in my five-option case, if we compare A1 (saving Aaron) to A2 (saving Aaron and Alex), A1 will be quite bad, given that PARETO prefers A2; choosing A1 is gratuitously worse for Alex. In this context, we might give A1 a score of 1 and A2 a score of 2. But when we compare A1 to B1 (saving Betty), PARETO no longer matters; the two options save totally

[12] The literature on Taurek and transitivity is, unfortunately, a bit messy. It's often unclear who came up with which ideas. Even Friedman (2009: 281) misses a citation; his counterexample to Kavka (1979) is just like Parfit's (1982: 131) counterexample to a later time-slice of Kavka (1982).

[13] I am simplifying. Friedman switches between three views: (1) a thing's goodness depends on the pairwise alternative ("the degree of significance that different morally relevant factors have varies depending on the particular comparison being made" (2009: 281)); (2) betterness is menu-relative, i.e., whether A > B may depend on the presence of C ("examples . . . where addition of alternatives changes preference to indifference, or vice versa, are easy enough to imagine" (2009: 281)); and (3) betterness depends on multiple factors aggregated non-additively (see his formal model in 2009: 282–83). I will focus on Friedman's (1). Since I reject Lübbe's move, I don't think Taurekians need (2). I endorse (3), though not Friedman's formalism (see Footnote 26, below).

[14] The term "essentially comparative" is due to Temkin (1987; 2012). Its meaning is debated (Cusbert 2017; Handfield 2016; Huemer 2013: 323–25)—see especially Dancy (2005: 1) on "alternative complementarity" (goodness depends on the alternative) vs. the "provenance view" (goodness depends on past history). I will focus on complementarity.

different people, so no one's death is gratuitous; in this context, we might give both a score of 2—the same score we would give to A2 and B1 when compared pairwise. Thus A1 is just as good as B1, which is just as good as A2, and yet A1 can still be worse than A2, because A1's goodness varies depending on whether the alternative is PARETO-preferred.

6. Comparativity vs. Multidimensionality

With the 1D ESSENTIALLY COMPARATIVE VIEW, Taurek has a way out of the nonsense argument. The view, of course, may be open to objections. Some might insist that TRANSITIVITY is a self-evident fixed point, or that changing values are incoherent. These claims are difficult to adjudicate without wading into deep and murky waters. Thankfully, my complaint is less complex. The 1D ESSENTIALLY COMPARATIVE VIEW, whatever its merits, is unsuitable for *Taurek*.

The problem is not the comparativity; it is the one-dimensionality. There is something funny about using numbers for values when the numbers don't count.

Consider Taurek's take on a tradeoff between lives, like A1 vs. B1 in the choice from {A1, A2, A3, B1, B2}. These two options are equally good, because they save disjoint groups and are otherwise similar.[15] But they are not morally *indistinguishable*. A1 saves Aaron; B1 saves Betty. On Taurek's view, unlike a crude utilitarian's, this difference makes a moral difference: the value of Aaron's life and the value of Betty's are nonfungible.[16] The lives are,

[15] As a reminder, I am not using "equally good" in a way that contrasts with being "on a par." (See Footnote 3, above.). For Chang (2002), 'equally good' means "evaluatively indistinguishable." For me, it just means "comparable, and neither is better than the other"—the analogue of "indifference" as used by economists like Sen (2017).
[16] I say "crude" because a sophisticated utilitarian could treat different people's pains and pleasures as in a sense nonfungible, as Chappell (2015) nicely shows.

in their own ways, equally good. This is not to say that they are equivalent in the way a ten-dollar bill is the equivalent in monetary value of two five-dollar bills. Someone who prefers the ten-dollar bill is either confused or concerned with more than cash value. But we understand why someone might prefer to save Aaron over Betty, simply for Aaron's sake, even if neither is a close friend (see Taurek 1977: 300–301). In the same way, we think it is fine to prefer a career in the arts over an equally rewarding career in journalism; when faced with equipollent and plural values, it is fine to have a favorite.[17]

The problem with Friedman's proposal, in this simple 1D form, is that it has no role for nonfungibility—the deep idea underlying Taurek's views.[18] The 1D proposal can allow for different good-*makers* (PARETO, respect for rights . . .), but it only uses fungible *values*. Things that are equally good have one and the same value, because they are given one and the same number. This all strikes me as more of a formal trick than a genuine expression of Taurek's ethics.

Can we do better? Well, let's think about what we are after. We still need to reject a premise of the nonsense argument—either SCORES, INTERNAL SCORING, or 1D SCORING. It is hard to imagine ethics without SCORES. (*Value* theory without *values?*) But it is also hard to pair Taurek's view with 1D SCORING, which quantifies the value of life so abstractly.

The natural solution, I think, is to give up 1D SCORING. Taurek can have internal values; he just needs a fancier view of what values *are*. Rather than being single numbers,

---

[17] When it is permissible to prefer either of two options over the other, Rabinowicz (2008: 30) calls them "on a par." (Cf. Chang's notion, cited in Footnotes 3 and 15, above.)
[18] See also Kamm's (1993; 2005: 3) discussion of "sobjectivity," her term for the way in which Taurek mixes the subjective perspectives of different people into an objective judgment of goodness.

values might have multiple dimensions—at a minimum, one per person. For example, in a choice between:

A: **<u>Aaron</u>**, Betty.

B: Aaron, **<u>Betty</u>**.

We might assign the following 2D values: [19]

V(A): (1, 0).

V(B): (0, 1).

These values are clearly, in some sense, distinguishable: A has a higher value in the first dimension, which measures how good the outcome is for Aaron, while B does better in the Betty dimension. Now we have a way to represent the nonfungible value of the two lives. But we aren't done yet. We still need to know how the dimensions *aggregate*—we need a rule telling us whether A is better than B given their 2D goodness scores.

As it turns out, a simple rule will do the trick, at least in cases where PARETO and TRADEOFFS are the only principles in play. (No rights violations, no varying sizes of harm, etc.) In such cases, ignoring Lübbe's move, Taurek's view is equivalent to:

---

[19] Whenever we are dealing with internal values, not ones that change depending on the alternative, I will omit the subscript from our value function V.

PARETO EXTENSION

A > B *iff* A is better for someone than B and worse for no one.

This is a principle about the relative value of options (cf. Sen 2017: 119). It says: A ~ B *unless* PARETO says otherwise; it "extends" PARETO by making options equally good whenever PARETO does not tell in favor of one over the other.

PARETO EXTENSION is not itself a rule telling us how to aggregate different dimensions of value. But we can give an analogous rule that does just that, with the different dimensions playing the role of the different people. (The analogy is especially tight here since the dimensions correspond to people.) Formally, the rule is:

PARETO EXTENSION (SCORING)

$(x_1, x_2, \ldots, x_n) >> (y_1, y_2, \ldots, y_n)$ *iff* for some x_i, $x_i > y_i$, and for no y_j, $y_j > x_j$.

Informally: a value is higher just if it is better in one way, worse in none.[20]

The 'iff' is essential. This rule doesn't just say that a value is higher if it outranks in one dimension and isn't outranked in any; it also says that this is the *only* way for a value to be higher. If two values have the same scores across the board, or if each outranks the other in a dimension, neither value is higher; they are equal.

Let's see the PARETO EXTENSION in action. In the simple A vs. B case, it says A ~ B,

[20] I am assuming that two values are equal (though not necessarily indistinguishable) if neither is higher than the other (see Footnote 3, above), and, purely for simplicity, I assume that goodness along any given dimension can be measured with a single number.

since each option is better in one respect. So far, so good. In the choice from {A1, A2, B1}, we will need 3D scores. Let the first dimension measure how good things are for Aaron; the second, for Alex; and the third, for Betty, setting the value of survival again at a nice simple '1'. The scores are:

A2: (1, 1, 0).

A1: (1, 0, 0).

B1: (0, 0, 1).

This rule also gives us Taurek's nontransitivity: A2 > A1 ∼ B1 ∼ A2. Indeed, the nontransitivity persists even in the five-option case, where we add dimensions for Alice and Boris. The scores are:

A3: (1, 1, 1, 0, 0).

A2: (1, 1, 0, 0, 0).

A1: (1, 0, 0, 0, 0).

B2: (0, 0, 0, 1, 1).

B1: (0, 0, 0, 1, 0).

And again, the PARETO EXTENSION delivers the nontransitivity: A2 > A1 ∼ B1 ∼ A2. The extra dimensions don't change the fact that A2 is a pure improvement on A1, whereas each of the As is somehow better than each of the Bs and vice versa.

We finally have it: a response to the nonsense argument. Taurek can accept SCORES and Internal Scoring, but instead of 1D Scoring, he can allow for multidimensional scores, combined using the Pareto Extension rule. And this rule really is crucial; there are other rules that do *not* deliver Taurek's view even given multidimensional scores. For suppose we say that one value is higher just in case its dimensions add up to a higher sum. Or more formally:

Addition (Scoring)

$(x_1, x_2, \ldots, x_n) >> (y_1, y_2, \ldots, y_n)$ *iff* $\sum_{i=1}^{n} x_n > \sum_{i=1}^{n} y_n$.

Informally: higher values are ones whose dimensions sum to a higher number.

This view pays lip service to nonfungibility. But it doesn't let nonfungible values have any effect on betterness; it "counts the numbers" and treats two lives as greater than one, other things equal. We could have gotten all the same betterness judgments by replacing the multidimensional scores with sums, going back to 1D Scoring. The Pareto Extension, by contrast, is an essentially multidimensional rule; it delivers a nontransitivity that cannot be replicated with single fixed numbers.

(There is a subtler 1D way to model Taurek's view, which is at least worth a mention. Suppose we keep Scores and Internal Scoring and think of values not as single numbers but as *intervals*—ranges like [0, 1]. We then say that A > B *iff* every number in A's interval is higher than every number in B's (Gert 2004: 505; see also Chang 2005). The use of intervals suggests that we are dealing with value judgments that are "imprecise" (see Parfit 2011 on "imprecise equality"), which sounds vaguely anti-additive, and it does allow us to model some nontransitivities. But intervallic modeling clearly won't work for Taurek.

First, it doesn't have any real role for nonfungibility. Second, it fails spectacularly in cases with four or more options, where Taurek's PARETO EXTENSION can violate the:

INTERVAL ORDER PROPERTY

If A > B and C > D, then either A > D or C > B.

Provably, no case that violates this property can be modeled using intervals and the rule I gave above (see Fishburn 1970: 20–23; Rabinowicz 2008: 33 n. 23). And yet, it is easy to get violations from the PARETO EXTENSION; consider our choice from {A3, A2, A1, B2, B1}, in which the rule gives us A2 > A1 ~ B2 > B1 ~ A2. Intervals on a line are better than points, but still a poor substitute for truly multiple dimensions.)

7. Conclusion

I have argued that Taurek's nontransitivity cannot be cut out from his view, as his defenders hoped, but I have also argued that the nontransitivity is not sheer nonsense, as his critics allege.

There are two main ways to formalize Taurek's nontransitive view: we could have simple values that change with context (the 1D ESSENTIALLY COMPARATIVE VIEW) or complex values that combine in some way subtler than mere addition. I think complex values are truer to Taurek. His nontransitivity arises from his concern for the nonfungible value of human life, which can be expressed formally by comparing multidimensional values with a non-additive rule—the PARETO EXTENSION.[21]

This may sound a bit squishy to the tough-minded maximizer. But for Taurek, who denies that good things must add up to something better, it is only natural that values should be at their core something marvelously uncountable—or, at least, something subtler than single numbers.

That said, Taurek's view is still *strange*. I have made no attempt to argue otherwise. But by working through the TRANSITIVITY objection, I hope we can now better understand the strangeness—where it comes from, where it isn't so bad, and where it can be fixed.

First, note that Taurek's nontransitivity is of a mild variety. We have not seen any spicy violations of the transitivity of '>', nor any dreaded cycles, where A > B > . . . > A, leaving the agent in a dilemma where every option loses to something. Taurek just has a nontransitive '∼' and '≥'. In particular, he allows cases where betterness does not transmit over ties: A2 > A1 ∼ B1, but not A2 > B1.[22]

Second, such cases are hardly unique to Taurek; they seem to arise in other cases of

[21] See Doggett (2009: 8–14) for a Taurekian discussion of non-additive rules for how to weigh reasons for action, as opposed to dimensions of value. (Some, like Lübbe 2008: 74 and Lee 2017: 5, think this amounts to the same thing). It is not trivial to extend Taurek's view to cases where sizes of harm vary and some acts violate rights. Certainly, Taurek will need something more than the PARETO EXTENSION, if only because he thinks it better to prevent a giant harm to one rather than a nuisance to another (1977: 302).

[22] See the citations in Footnote 4, above. Also, since Taurek's view rejects cycles of betterness, he does not have to give up Sen's (2017: Chapter 1*) attractive Property α (sometimes called "the independence of irrelevant alternatives"), which says that a permissible option cannot be made wrong by taking other options off the menu. He only has to give up the less attractive Property β, which says that if x and y are both permissible, adding options cannot make only one of them wrong. (As adding A2 to {A1, B1} might make only A1 wrong.) Property β is equivalent, given minimal conditions, to PI-intransitivity: if A > B ∼ C, then A > C (Sen 2017: 66). For more on Sen's properties in ethics, see Muñoz (2021: 708–12).

conflicting values. Here is a familiar kind of example.[23] Suppose I am choosing between coffee (C) and tea (T). Each has its advantages—the dark roast's smoothness, the oolong's freshness—but neither is better overall: C ∼ T. Now suppose we add a third option: coffee at a *slightly* nicer temperature (C+). Surely C+ > C, since it is better in one way and worse in none. But it is hardly obvious that C+ > T. After all, the tea still has its advantages over the improved coffee. We might say that the '∼' relation is, in this case, stable over small improvements. That violates transitivity: C+ > C ∼ T, but not C+ > T. (So: C ≥ T ≥ C+, but not C ≥ C+.) Some ethicists, like Chang (2002) and Hare (2010), have come to embrace such judgments. I do not think their view is particularly extreme. If that is right, and their view is less wacky than Taurek's, there must be something *else* in Taurek, something besides the nontransitivity, that grounds the wackiness.

I believe the true source of strangeness is not that Taurek has a nontransitive view, or that he multiplies dimensions. The problem is that he is so *uncompromising* in how he refuses to trade off one dimension against the rest. Even if X saves a hundred lives and Y saves only a single other, Taurek would not conclude that X is better, other things equal.[24] For Taurek, '∼' is not just stable over small improvements; it is stable over what seem like arbitrarily massive improvements. (For instance: A1 ∼ B1, and A100 massively improves on A1—saving 99 more lives!—yet A100 ∼ B1.) This is what makes his view so extreme:

[23] The authoritative treatment is Chang's (2002: 667–73) discussion of the "small improvement argument;" see her many citations, and see Hare (2010) on "insensitivity to mild sweetening."

[24] Contrast this with the beverages: if I slightly improved the coffee along a bunch of different dimensions, eventually the improved coffee could well be better than the tea overall, even if it remains worse in terms of freshness.

the numbers *never* count, no matter how enormous.[25]

If Taurek compromised by counting numbers when dealing with big differences (like A100 vs. B1), his view would be moderate, like the popular Chang/Hare view of "small improvement" cases. Taurek could achieve this by supplementing PARETO (well, really, the analogous rule for comparing values) with:

MODERATE TRADEOFFS (SCORING)

$(x_1, x_2, \ldots, x_n) >> (y_1, y_2, \ldots, y_n)$ if $\sum_{i=1}^{n} x_n > m + \sum_{i=1}^{n} y_n$

Informally: a value is higher if its sum is greater by a certain margin.[26]

Where this says that one value can be higher than another, despite losing along one dimension, if it wins by a sufficient amount along all dimensions combined.

The result is a Paretian view on which the numbers count when they aren't close. This is Taurek Lite—his view minus one bit that makes it extreme.[27] I expect that many philosophers will be more open to this kind of view than Taurek's own. Then again, the

[25] Put formally, Taurek's dimensions of value form a LIBERUM VETO OLIGARCHY: if A is better than B along all dimensions, A is better overall, and if A is better along *any* dimension, then A is at least as good as B overall. No dimension's dissent can ever be overruled. The formal concept of an oligarchy comes from Gibbard (2014), who notes that the PARETO EXTENSION leads to an oligarchy of all, and explores connections between oligarchies, Arrow's Impossibility Theorem (Arrow 1951), and varieties of transitivity.

[26] This principle—though *very* crude, with its single fixed margin—has more expressive power than Friedman's (2009: 282–83) formalism, which only allows us to compare one respect in which A is good to one respect in which B is good.

[27] Anscombe (1967), in an important precursor to Taurek's paper, takes another kind of moderate anti-counting position. She denies that you must save five strangers rather than one, but she allows that it would be intelligible to save the many for the reason that they are more. For another moderate view, see Setiya (2014).

things that make Taurek's original view unpopular also make it singularly fascinating, especially to the formal ethicist.

In this paper, I have tried not to take sides, but only to understand Taurek rigorously—on his own terms. If Taurek's real concern is the nonfungible value of human life, his nontransitivity might be a feature, not a bug, and the extreme parts of his view might be the icing, not the cake.

Acknowledgments

My thanks to the editors of *Ergo* and several anonymous referees for speedy and insightful comments; this was one of the best experiences I have ever had with an academic journal. For discussion, I thank Brian Hedden, Zach Barnett, Theron Pummer, Toby Handfield, Peter Graham, and Kerah Gordon-Solmon. Thanks also to Kieran Setiya, who sparked my interest in Taurek, and special thanks to Tyler Doggett, whom I 'know and like', for encouragement, inspiration, and—best of all—detailed written comments.

References

Anscombe, G. E. M. (1967). Who Is Wronged? Philippa Foot on Double Effect: One Point. *The Oxford Review*, *5*, 16–17.

Arrow, Kenneth (1951). *Social Choice and Individual Values*. Wiley.

Binmore, Ken and Alex Voorhoeve (2003). Defending Transitivity Against Zeno's Paradox. *Philosophy and Public Affairs*, *31*(3), 272–79.

Chang, Ruth (2002). The Possibility of Parity. *Ethics*, *112*(4), 659–88.

Chang, Ruth (2005). Parity, Interval Value, and Choice. *Ethics*, *115*(2), 331–50.

Chappell, Richard Yetter (2015). Value Receptacles. *Noûs*, *49*(2), 322–32.

Cohen, Yishai (2014). Don't Count on Taurek: Vindicating the Case for the Numbers Counting. *Res Publica*, *20*(3), 245–61.

Cusbert, John (2017). Acting on Essentially Comparative Goodness. *Thought: A Journal of Philosophy*, *6*(2), 73–83.

Dancy, Jonathan (2005). Essentially Comparative Concepts. *Journal of Ethics and Social Philosophy*, *1*(2), 1–16.

Doggett, Tyler (2009). What Is Wrong with Kamm's and Scanlon's Arguments Against Taurek. *Journal of Ethics and Social Philosophy*, *3*(3), 1–16.

Doggett, Tyler (2013). Saving the Few. *Noûs*, *47*(2), 302–15.

Dorr, Cian, Jacob M. Nebel, and Jake Zuehl (forthcoming). The Case for Comparability. *Noûs*.

Fishburn, Peter C. (1970). *Utility Theory for Decision Making*. John Wiley & Sons.

Friedman, Alex (2002). *Minimizing Harm: Three Problems in Moral Theory* [unpublished doctoral dissertation. MIT.

Friedman, Alex (2009). Intransitive Ethics. *Journal of Moral Philosophy*, *6*(3), 277–97.

Gert, Joshua (2004). Value and Parity. *Ethics*, *114*(3), 492–510.

Gibbard, Allan F. (2014). Social Choice and the Arrow Conditions. *Economics and Philosophy*, *30*(3), 269–84.

Graham, Peter A. (2017). The Numbers Count. *Thought: A Journal of Philosophy*, *6*(2), 129–34.

Halstead, John (2016). The Numbers Always Count. *Ethics*, *126*(3), 789–802.

Handfield, Toby (2016). Essentially Comparative Value Does Not Threaten Transitivity. *Thought: A Journal of Philosophy*, *5*(1), 3–12.

Hare, Caspar (2010). Take the Sugar. *Analysis*, *70*(2), 237–47.

Hirose, Iwao (2001). Saving the Greater Number Without Combining Claims. *Analysis*, *61*(4), 341–42.

Hirose, Iwao (2004). Aggregation and Numbers. *Utilitas*, *16*(1), 62–79.

Horton, Joe (2017). The All or Nothing Problem. *Journal of Philosophy*, *114*(2), 94–104.

Huemer, Michael (2013). Transitivity, Comparative Value, and the Methods of Ethics. *Ethics*, *123*(2), 318–45.

Kamm, Frances M. (1993). *Morality, Mortality: Volume I: Death and Whom to Save From It*. Oxford University Press.

Kamm, Frances M. (2005). Aggregation and Two Moral Methods. *Utilitas*, *17*(1), 1–23.

Kamm, Frances M. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford University Press.

Kavka, Gregory S. (1979). The Numbers Should Count. *Philosophical Studies*, *36*(3), 285–94.

Kavka, Gregory S. (1982). The Paradox of Future Individuals. *Philosophy and Public Affairs*, *11*(2), 93–112.

Kumar, Rahul (2001). Contractualism on Saving the Many. *Analysis*, *61*(2), 165–70.

Lee, Dong-Kyung (2017). Against Hirose's Argument for Saving the Greater Number. *Journal of Ethics and Social Philosophy*, *11*(2), 1–8.

Liao, S. Matthew (2008). Who Is Afraid of Numbers? *Utilitas*, *20*(4), 447–61.

Lübbe, Weyma (2008). Taurek's No Worse Claim. *Philosophy and Public Affairs*, *36*(1), 69–85.

Meyer, Kirsten (2006). How to Be Consistent without Saving the Greater Number. *Philosophy and Public Affairs*, *34*(2), 136–46.

Muñoz, Daniel (2021). Three Paradoxes of Supererogation. *Noûs*, *55*(3), 699–716.

Nebel, Jacob M. (2018). The Good, the Bad, and the Transitivity of Better Than. *Noûs*, *52*(4), 874–99.

Otsuka, Michael (2004). Skepticism about Saving the Greater Number. *Philosophy and Public Affairs*, *32*(4), 413–26.

Parfit, Derek (1978). Innumerate Ethics. *Philosophy and Public Affairs*, *7*(4), 285–301.

Parfit, Derek (1982). Future Generations: Further Problems. *Philosophy and Public Affairs*, *11*(2), 113–72.

Parfit, Derek (2011). *On What Matters*. Oxford University Press.

Pummer, Theron (2016). Whether and Where to Give. *Philosophy and Public Affairs*, *44*(1), 77–95.

Pummer, Theron (2019). All or Nothing, but If Not All, Next Best or Nothing. *Journal of Philosophy*, *116*(5), 278–91.

Rabinowicz, Wlodek (2008). Value Relations. *Theoria*, *74*(1), 18–49.

Rulli, Tina (2020). Conditional Obligations. *Social Theory and Practice*, *46*(2), 365–90.

Sanders, John T. (1988). Why the Numbers Should Sometimes Count. *Philosophy and Public Affairs*, *17*(1), 3–14.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Sen, Amartya (2017). *Collective Choice and Social Welfare: An Expanded Edition*. Harvard University Press.

Setiya, Kieran (2014). Love and the Value of a Life. *Philosophical Review*, *123*(3), 251–80.

Taurek, John (1977). Should the Numbers Count? *Philosophy and Public Affairs*, *6*(4), 293–316.

Temkin, Larry S. (1987). Intransitivity and the Mere Addition Paradox. *Philosophy and Public Affairs*, *16*(2), 138–87.

Temkin, Larry S. (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford University Press.

Timmermann, Jens (2004). The Individualist Lottery: How People Count, but Not Their Numbers. *Analysis*, *64*(2), 106–12.

Wasserman, David and Alan Strudler (2003). Can a Nonconsequentialist Count Lives? *Philosophy and Public Affairs*, *31*(1), 71–94.

Woodward, James F. (1981). Why the Numbers Count. *Southern Journal of Philosophy*, *19*(4), 531–40.