



ARTICLE

# Sources of transitivity

Daniel Muñoz 

The University of North Carolina at Chapel Hill, 240 E Cameron Ave Chapel Hill, NC 27514, USA  
Email: [Daniel.Munoz@unc.edu](mailto:Daniel.Munoz@unc.edu)

(Received 3 March 2021; revised 30 January 2022; accepted 19 March 2022; first published online 13 June 2022)

## Abstract

Why should ‘better than’ be transitive? The leading answer in ethics is that values do not change with context. But this cannot be the entire source of transitivity, I argue, since transitivity can fail even if values never change, so long as they are complex, with multiple dimensions combined non-additively. I conclude by exploring a new hypothesis: that all alleged cases of nontransitive betterness, such as Parfit’s Repugnant Conclusion, can and should be modelled as the result of complexity, not context-relativity.

**Keywords:** Transitivity; value theory; multidimensionality; intrinsic value; Condorcet’s Paradox

## 1. Introduction

A hot cup of coffee beats a lukewarm mug, which beats no coffee at all. It seems to follow that hot coffee is better than no coffee. More generally, where ‘>’ means ‘is better than’, it seems:

TRANSITIVITY

If  $A > B$  and  $B > C$ , then  $A > C$ .

Almost every philosopher who discusses this principle endorses it. Many think it is undeniable, even trivial.<sup>1</sup>

But what *makes* TRANSITIVITY so hard to deny? The leading answer in ethics, due to Larry Temkin, is that TRANSITIVITY has its source in our naïve conception of goodness: the *Internal Aspects View*. On this view, the truth of ‘ $A > B$ ’ depends on how  $A$ ’s goodness compares with  $B$ ’s, and a thing’s goodness just depends on its internal features, not on what it is compared with.<sup>2</sup> Temkin (2012: 388)

<sup>1</sup>For some important defences, see Binmore and Voorhoeve (2003) and Nebel (2018). For challenges to transitivity, see Rachels (1998), Friedman (2009) and Temkin (2012).

<sup>2</sup>In economics and decision theory, the leading argument for transitivity, as a requirement on rational preferences, is that nontransitive preferences can be ‘money pumped’. To illustrate, suppose I get  $C$  if I do nothing, and I prefer  $A$  to  $B$  to  $C$  to  $A$ . Then I should be willing to pay money if someone offers me  $B$  instead of  $C$  – and I should pay, again, to trade up from  $B$  to  $A$ , and once more to trade up from  $A$  to  $C$ . But then I will have paid good money to end up where I started. For an early example of such an argument, see Davidson *et al.* (1955).

sometimes describes the process in the language of ‘scores’. First, we ponder A on its own and award it a goodness score. Next, we do the same for B; finally, we compare scores: if A’s is higher,  $A > B$ .

I argue that the Internal Aspects View cannot be the full source of TRANSITIVITY’s appeal. The view does not even *entail* TRANSITIVITY – not unless the ranking of value ‘scores’ is itself transitive, which the view does not itself guarantee.

How could nontransitive rankings emerge from internal scores?<sup>3</sup> The answer is that scores could be complex – involving multiple dimensions combined in some way subtler than mere addition. I illustrate this with an adaption of Condorcet’s Paradox, in which majority rule leads to cycles of group preference. In the adapted version – familiar to economists and decision theorists – ‘majority rule’ is applied to dimensions of value rather than individual voters.<sup>4</sup> I conclude that if we find TRANSITIVITY obvious, that is not just because we accept the Internal Aspects View; we must also be rejecting views with complex aggregation rules like majority rule. With the right complexity, even internal scores give rise to nontransitivity.

This conclusion is at odds with Temkin’s own. He thinks TRANSITIVITY failures can only be understood on his *Essentially Comparative View*, which says that values can change as we compare an option to different alternatives.<sup>5</sup> This may be one way to get TRANSITIVITY failures. But I argue that a multidimensional Internal Aspects View is a second, neglected cause of nontransitivity (§§2–4).<sup>6</sup> I then consider whether it might be the *only* cause (§§5–8), and show how certain comparative effects can flow from an Internal Aspects View (§9). If I am right, the transitivity debate should be reconceived: it is not just about context, but also complexity, and ethicists should be drawing inspiration from economists’ insights into aggregation.

## 2. Condorcet’s Paradox

Let’s start with a classic case. Alan, Betty and Chico want snacks, but they have different tastes. Their preferences are as follows, where ‘A’ stands for the outcome where everyone gets apples, ‘B’ bananas and ‘C’ cherries:

<sup>3</sup>I use ‘nontransitive’ instead of ‘intransitive’ because some people use the latter to mean ‘never transitive’, instead of ‘not always transitive’ (Humberstone 2016: 369); see Temkin (2012: Ch. 11) for a third use.

<sup>4</sup>Here are just a few important discussions of majority rule for values. May (1954: 6–7) considers a version of majority rule where the ‘voters’ are replaced with dimensions of preference; Hurley (1985: 505) touches on majority rule for ethical values; and Arrow and Reynaud (1986) explore the viability of majority rule as a multicriterial decision rule for decision-makers in industry. My thanks to the referees for pointing out the relevance of Arrow and Reynaud’s work.

<sup>5</sup>Temkin (2012: Ch. 11, fn. 32) says ‘there might be room in conceptual space’ for an Internal Aspects View without TRANSITIVITY, but he sets the possibility aside; as he points out, no one has given even the ‘broadest sketch’ of how to formalize such a view (Temkin 2014a: 142).

<sup>6</sup>Katz (2015: 420–21) agrees; see fn. 9 below. See also Persson (2014), along with Temkin’s (2014a) persuasive reply. Some doubt that essential comparativity clashes with TRANSITIVITY (Handfield 2016; Klockseim 2016: 1315; Cusbert 2017: 76–77). I interpret the essentially comparative view so that it *does* clash with TRANSITIVITY, but will not have space to compare interpretations.

Alan:  $A > B > C$   
 Betty:  $B > C > A$   
 Chico:  $C > A > B$ <sup>7</sup>

Their preferences, moreover, are transitive. Alan prefers apples to cherries, Betty prefers bananas to apples, and Chico prefers cherries to bananas. All that matters in this case is giving people what they want. We can only choose one snack, however, so there are just three outcomes: apples for all, bananas for all, cherries for all. Which is best?

That depends on how we settle conflicts. Apples are best for Alan, bananas for Betty, and cherries for Chico. We want to know: which is best *overall*?

The answer is easy if we can just sum up individual desires, measured with a shared cardinal unit, to arrive at the group's satisfaction level. For example, if Alan's preferences are vastly stronger, the best outcome is for Alan to get his apples, as in:

	Apples	Bananas	Cherries
Alan	10	5	0
Betty	0	2	1
Chico	1	0	2
Group	11	7	3

Here we resolve the conflict by adding desires and comparing sums, arriving at a tidy transitive ranking:  $A > B > C$ .

But consider a non-additive resolution:

#### MAJORITY RULE

$A > B$  if the majority of people prefer A to B.

A foundational result in social choice theory is *Condorcet's Paradox*: MAJORITY RULE can violate TRANSITIVITY even when individual preferences are transitive (Arrow 1963; Sen 2017).<sup>8</sup> Here is how. Suppose Alan, Betty and Chico vote, pair by pair, on their favourite snacks. MAJORITY RULE will 'rank' the outcomes:  $A > B > C > A$ . Why? Because two prefer apples to bananas; two prefer bananas to cherries, and two prefer cherries to apples.

The group's preferences violate not just TRANSITIVITY, but even:

#### ACYCLICITY

If  $A > B > \dots > Z$ , then it is not the case that  $Z > A$ .

<sup>7</sup>I rely on context to clarify when '>' means 'better than', 'greater than', or 'preferred to'.

<sup>8</sup>Condorcet's Paradox is named after the Marquis de Condorcet, a pioneer in social choice. Condorcet made several groundbreaking contributions; for example, he showed that the Borda Rule violates the Independence of Irrelevant Alternatives; see Baker (1975), cited in Katz (2011: 3–4).

Which says that there are no cycles of '>'. It is rather shocking that MAJORITY RULE should violate such a basic principle. Hence Condorcet's *Paradox*.

I am not, however, here to talk about rules for voting. I am interested in MAJORITY RULE as a principle for how values combine in a single agent's choice.<sup>9</sup> We have three outcomes, none of which is best in every respect: A is best for Alan, B for Betty and C for Chico. We have three dimensions of value, three ways in which an outcome can be good – one for each individual. We can think of MAJORITY RULE as a method for aggregating dimensions:  $A > B$  if A is better than B for *most* people. This leads to violations of TRANSITIVITY and even ACYCLICITY for betterness:  $A > B > C > A$ .

Of course, cycles aside, MAJORITY RULE is a problematic moral principle. One problem is that doesn't factor in anything besides utility – what about justice, knowledge and other dimensions of value? But this is fixable. We simply generalize MAJORITY RULE to say that A is better than B if it is better along most dimensions, whatever they may be.

The real problem with MAJORITY RULE in morality is that it only cares about whether A beats B along most dimensions, not by how much.<sup>10</sup> For example, even in the case where Alan's preferences are by far the most intense, MAJORITY RULE will *not* say that it is best overall for Alan to get his apples – not even if we multiply the intensity of his desires by a million, and dull down Betty's and Chico's – because the majority prefer cherries. This alone makes MAJORITY RULE fishy as a principle of value.<sup>11</sup>

Thankfully, I am not *advocating* for this principle, only using it for illustration.<sup>12</sup> My point is that MAJORITY RULE can lead to cycles of betterness even given the Internal Aspects View, which means that this view cannot really underwrite TRANSITIVITY.

### 3. Internal aspects and 1D scores

Now we need to show that MAJORITY RULE fits the Internal Aspects View. Let's start by getting clear on what the view really is.

---

<sup>9</sup>In this part of the paper (§§2–4), I am not trying to establish that MAJORITY RULE can be applied to moral values – that much is already well-known (see fn. 4, above). My point is that a moral MAJORITY RULE gives us a crisp example of how TRANSITIVITY could arise *without essentially comparative values*. The only other person to make this point, as far as I know, is Katz (2015: 421), who notes that MAJORITY RULE can lead to cycles without comparativity: 'It is not in this case the varying nature of the criteria, however, that generates the cycle. Instead it inheres in the very method of aggregation.' I agree. My arguments in §§2–4 can be read as a defence and exploration of Katz's neglected insight. (I am grateful to the referees for helping me frame this part of the paper.)

<sup>10</sup>Here I agree with Temkin (2015a: 486–487).

<sup>11</sup>Does the fact that MAJORITY RULE ignores intensity of preference make it a bad *voting* rule? This question has been debated since Condorcet's time (Gehrlein 2006: 2–3). Regardless, MAJORITY RULE is much more reasonable as a rule for aggregating voters rather than a principle for combining values. (But see fn. 12, below, for a point in favour of MAJORITY RULE for values.)

<sup>12</sup>We might prefer MAJORITY RULE to ADDITION if we reject interpersonal utility comparisons, as does Arrow: 'it seems to make no sense to add the utility of one individual, a psychic magnitude in his mind, with the utility of another individual' (Arrow 1963: 11).

Temkin (2012: 229) believes that the debate over TRANSITIVITY turns on a deeper conflict between two views of value. The Internal Aspects View says that a thing's value just depends on what the thing is like in itself, and that we compare values to determine which things are better. The Essentially Comparative View says that some aspects of a thing's value may depend on what the alternatives are; a certain factor might make A better than B without affecting whether A is better than C. The Internal Aspects View leads to TRANSITIVITY; the Essentially Comparative View leads us away.

That, anyway, is the breezy version. When we look closer, it is less obvious what Temkin takes to be essential to the Internal Aspects View. Consider this gloss:

On [the Internal Aspects View], each outcome is composed of a multitude of features. Together, these features and the relations between them will determine the value of that outcome from an impartial perspective. This value, which could (perhaps) be accurately represented on a linear scale measuring the value of outcomes, is *fixed solely by the outcome's internal features*, and hence will be unchanged as long as the outcome's internal features are themselves unchanged. This value has a special significance – it is, as it were, the real value or true value or, perhaps, the intrinsic value of the outcome. Correspondingly, this value will be the key feature determining how the outcome compares to other outcomes, whose values will likewise be determined solely by *their* internal features and will also be accurately representable on the same linear scale. (Temkin 2012: 229, emphasis original; see also Temkin 1987: 158–159)

There are three claims here. First, betterness is determined by comparing values. Second, an outcome's value cannot change as we compare it to different alternatives. Third, values can ('perhaps') be ranked along a linear scale, like notches on a ruler.<sup>13</sup>

Let's call this combination of claims the 1D INTERNAL ASPECTS VIEW. Put precisely, the view combines the following:

#### SCORES

$$A > B \text{ iff } V_B(A) >> V_A(B).$$

*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.*

<sup>13</sup>There are two ways to interpret this third claim: (1) we *can* assign numbers to values, and (2) the numbers we assign to values are cardinal measurements, i.e. unique up to positive linear (or affine) transformation. 'Linear scale' suggests cardinality. But I set aside this stronger claim since the weak one in the text is enough to secure TRANSITIVITY.

## 1D SCORING

There is a function  $f: V \rightarrow \mathbf{R}$  such that for any  $v_i, v_j \in V$ :  $v_i \gg v_j$  iff  $f(v_i) > f(v_j)$ .

*Informally: values can be ranked like numbers on a line.*

Let me explain the symbols.  $V$  is a function that tells us the value of an outcome, all things considered, in a context of comparison. For example, ' $V_B(A)$ ' denotes  $A$ 's value when compared with  $B$ . (When context doesn't matter, I write ' $V(A)$ '.) To say one value is higher, we use ' $>>$ '. For example: ' $V_B(A) \gg V_A(B)$ ' means that  $A$ 's value is higher than  $B$ 's when comparing the pair. The function  $f$  in 1D SCORING maps values (in the set  $V$ ) to real numbers (in the set  $\mathbf{R}$ );  $f(v_i) > f(v_j)$  means that  $v_i$ 's number is greater.<sup>14</sup>

SCORES tells us that  $A$  is better than  $B$  just if, when comparing the pair,  $A$ 's value is higher overall; INTERNAL SCORES tells us that  $A$  has the same value no matter what it is compared with; and 1D SCORING tells us that values can be ranked like numbers on a line.<sup>15</sup>

Do these claims entail TRANSITIVITY, as Temkin says?<sup>16</sup> Yes. Suppose TRANSITIVITY fails:  $A > B > C$  but not  $A > C$ . Given SCORES and INTERNAL SCORING, the options' values must have the same structure:  $V(A) \gg V(B) \gg V(C)$ , but not  $V(A) \gg V(C)$ . Given 1D SCORING, there must be a way to plug in numbers for these values and substitute 'is greater than' for ' $>>$ '. But this can't be done. 'Is greater than' is transitive.

So the 1D INTERNAL ASPECTS VIEW entails TRANSITIVITY, and the entailment crucially depends on 1D SCORING. But why should 1D SCORING be essential to the Internal Aspects View? It does not follow from the core idea: betterness depends on internal values. (This may be why Temkin introduces 1D SCORING with a parenthetical 'perhaps'.) Nor does 1D SCORING appear in Temkin's most explicit statement of the view, where SCORES and INTERNAL SCORING are front and centre.<sup>17</sup> It seems to me that these latter two principles alone suffice for an Internal Aspects View. Such a view, at any rate, could not be essentially comparative, since values would stay fixed.<sup>18</sup>

We need a name for this kind of view, and so, with apologies to Temkin, I will call it the 'INTERNAL ASPECTS VIEW'.<sup>19</sup> This label will cover any view that rejects essential comparativity by embracing SCORES and INTERNAL SCORING. So defined, the

<sup>14</sup>For simplicity, I assume that  $A$  is better than  $B$  if  $A$  wins in a pairwise comparison (ignoring any effect from  $C$ 's presence; see Cusbert 2017: 78–81).

<sup>15</sup>INTERNAL SCORES does not say that values arise from intrinsic properties – only that they stay constant across comparisons. Constancy is what matters for TRANSITIVITY, as Huemer (2013: 325) points out.

<sup>16</sup>It is not just Temkin (2012: 229, 386); others agree (Friedman 2009: 292; Coons 2014: 293, fn. 14).

<sup>17</sup>Temkin's main definition of the Internal Aspects View doesn't mention 1D SCORING (Temkin 2012: 370; see also Temkin 1999: 777), nor do others' definitions (Coons 2014: 292; Roberts 2014: 309; Handfield 2016: 4; Pummer 2018: 1740; an exception is Cusbert 2017: 74).

<sup>18</sup>Should 1D SCORING be part of the Essentially Comparative View? I don't think so. (And while Temkin does include it at 2012: 229, he doesn't at 2012: 371.)

<sup>19</sup>I might not need to apologize; arguably, this is Temkin's official definition! He often says 1D SCORING is not part of the Internal Aspects View proper, but rather part of a natural interpretation (or 'representation') of the view, and it is only the view *so interpreted* that conflicts with TRANSITIVITY (Temkin 1987: 159; 2012: 385; 2014b: 272; 2020: 96). But, admittedly, he does not always distinguish the view from the interpretation (see e.g. Temkin 2014b: 272; 2015b: 371).

INTERNAL ASPECTS VIEW does not require 1D SCORING, and can be paired with all sorts of other rules – even MAJORITY RULE. Now let’s see how this particular pairing leads to cycles in Condorcet’s Paradox.<sup>20</sup>

#### 4. Nontransitive betterness from internal aspects

It is finally time to ask what value scores *are*. 1D SCORING suggests that they are like numbers on a line. But when TRANSITIVITY and ACYCLICITY fail, values can’t have that structure; the number line never cycles back. So what, then, is a value?

We can start from the idea that values have *multiple dimensions*. In the snack example, there are three relevant dimensions: goodness for Alan, for Betty and for Chico. For simplicity, let’s just stick with these three, and let’s assume we can measure them with cardinal numbers – units of utility. We give each outcome a 3D score (a, b, c), representing how good it is for Alan, Betty and Chico, respectively. This is a single score: the outcome’s ‘real’ value. It depends only on ‘internal features’ – namely, utilities – and it measures how good an outcome is, all things considered. It is meant to be the ‘key feature’ in determining which outcomes are better.

We thus have a conception of scores that obeys SCORES and INTERNAL SCORING, though not necessarily 1D SCORING. The first key idea is:

3D SCORES

$$V \supseteq \{(x, y, z): x, y, z \in \mathbf{R}\}.$$

This says that values can be ordered triples of real numbers. In our case, the first number measures goodness for Alan, the second for Betty and the third for Chico. (More generally, the idea is to have  $n$ -tuple scores given  $n$  dimensions of value.)

This does not by itself, however, get us beyond 1D SCORING and TRANSITIVITY. For suppose we used the utilitarian’s additive scoring rule:

ADDITION

$$(a_1, \dots, a_n) \gg (b_1, \dots, b_n) \text{ if } a_1 + \dots + a_n > b_1 + \dots + b_n.$$

This squishes our  $n$  dimensions into one number: the sum, which then settles which things are better.  $A > B$  if  $A$ ’s sum is higher. This is barely a step beyond 1D SCORING.

Instead of adding the numbers, we could use a (long-foreshadowed) spin on MAJORITY RULE. The rule says:

MAJORITY RULE (SCORING)

$$(a_1, \dots, a_n) \gg (b_1, \dots, b_n) \text{ if for most } i, a_i > b_i.$$

In other words: one value is higher than another, all things considered, if it is higher *along the majority of dimensions*.

<sup>20</sup>Sometimes Temkin says that, on the Internal Aspects View, a situation’s value must be appraised abstractly, without any consideration of ‘who its members are’ (1987: 159). I don’t think we need this restriction. What’s more internal to a situation than the facts about who is *in* that situation? (Here, I follow Dancy 2005: 6–7.)

Now suppose we add 3D SCORES and MAJORITY RULE (SCORING) to the INTERNAL ASPECTS VIEW. The resulting 3D INTERNAL ASPECTS VIEW leads to nontransitivities in Condorcet's case. We can use the numbers from before:

	Apples	Bananas	Cherries
Alan	10	5	0
Betty	0	2	1
Chico	1	0	2

Where Alan's preferences are most intense. The 3D INTERNAL ASPECTS VIEW yields a cycle:  $A > B > C > A$ . For A is better than B in most respects; B is better than C in most respects; and C is better than A in most respects. (This is possible because, thanks to 3D SCORES, we have more dimensions to play with.) By MAJORITY RULE (SCORING), we will get a betterness cycle, just as MAJORITY RULE leads to cycling preferences. This cycle persists even if we modulate the intensity of individual preferences, so long as everyone's ranking remains the same.

The same kind of cycle can arise in other cases, too, not just clashes of utilities. Suppose we are evaluating the work of three philosophers – Dina, Ellie and Fumi – along the dimensions of depth, elegance and funniness. The 3D values might be:

	Dina	Ellie	Fumi
Depth	10	5	0
Elegance	0	2	1
Funniness	1	0	2

ADDITION gives a transitive ranking:  $Dina > Ellie > Fumi$ . Because the differences in depth are so big, they are decisive. But MAJORITY RULE (SCORING) would turn this into a cycle by ranking Fumi over Dina overall, given that his work is better than hers in most ways.

The 3D INTERNAL ASPECTS VIEW thus entails cycles in a variety of cases. But I want to emphasize something: it is still clearly an INTERNAL ASPECTS VIEW. The view obeys SCORES and INTERNAL SCORING, though of course the scores are not just single numbers anymore; the value of giving everyone apples, for instance, is (10, 0, 1). We might not be used to thinking of values as internal 3D scores.<sup>21</sup> But it should be immediately clear that these scores, whatever they are, are *not* essentially comparative. They are not even comparative in the broad sense of having a dimension that counts for more in one comparison than in another. Each vote counts equally in MAJORITY RULE.

And so, the 3D INTERNAL ASPECTS VIEW retains the core of the INTERNAL ASPECTS VIEW, without any whiff of comparativity, while still allowing cycles.

<sup>21</sup>Some may find it odd to say that a value score – a thing's goodness, all things considered – is a triple. If the question is 'how good is this?' don't we need a single answer? Yes, we do. But (0, 1, 2) is a single answer. It just has multiple parts, like a review of a restaurant that gives 4/5 stars for price and 5/5 stars for taste.



We have reached a surprising conclusion. It turns out that there are *two* distinct sources of nontransitivity: one is the Essentially Comparative View that (potentially) accepts 1D SCORING; the other is an INTERNAL ASPECTS VIEW that rejects 1D SCORING. TRANSITIVITY is assured if values are *both* reducible to one dimension *and* invariant across comparisons. But relaxing either assumption can lead to violations.

This result should fundamentally change the way we see the debate over TRANSITIVITY, which does not just turn on the INTERNAL ASPECTS VIEW. We can only get TRANSITIVITY by *also* assuming something like 1D SCORING or ADDITION.<sup>22</sup> This assumption deserves more direct discussion, and we should be much more careful to distinguish the question of whether values are internal from our choice of preferred scoring rule.

What scoring rule *should* we prefer? If we want to resist TRANSITIVITY, we have to forgo ADDITION. We might instead opt for MAJORITY RULE. But this is not very plausible when it comes to dimensions of value. If A is better than B in two ways, but vastly worse in one, it seems ridiculous to insist that A must be better overall.

This raises a question for future work. Is there an aggregation rule far enough away from ADDITION as to allow for nontransitivity, but not so close to MAJORITY RULE as to be ridiculous? If we find such a rule, we can establish a new kind of nontransitivity, consistent with the INTERNAL ASPECTS VIEW, that is not just possible but plausible. If we can find a problem inherent in such rules – something more direct than ‘they yield cycles’ – we will have found what Temkin wanted all along: a source for TRANSITIVITY.

## 5. Warm-up: Condorcet in context

I have argued that the 3D INTERNAL ASPECTS VIEW violates TRANSITIVITY in the case of Condorcet’s Paradox. As a warm-up for part two of the paper, let’s now put Condorcet’s case in context, by comparing it with some nearby nonmoral examples in the literature.

Interestingly, at the end of Temkin’s first paper on transitivity (1987: 187, fn. 51), he gives a case that is almost analogous: the *nontransitive dice*.

Suppose Die A’s six faces are 6, 6, 6, 2, 2, 2; Die B’s are 5, 5, 4, 3, 3, 2; and Die C’s are 6, 4, 3, 3, 3, 3. It is easily shown that rolling two die [*sic*] at a time, on any given roll, the probability of A beating B will be 6/5, the probability of B beating C will be 14/13, and the probability of C beating A will be 6/5. Hence, in the long run, one would do better to bet on A against B, B against C, and C against A.<sup>23</sup>

<sup>22</sup>There are ways of relaxing 1D SCORING and ADDITION that would recapture the transitivity of ‘>’ but not of ‘≥’ (‘is at least as good as’); one example is the use of intervals as value scores, and the rule that an interval is ‘higher’ than another just if all its points are greater (so [2, 4] is not higher than [1, 3], but it is higher than [0, 1]). This rule is proposed by Gert (2004). For critique, see Chang (2005) and Rabinowicz (2008).

<sup>23</sup>Temkin means odds, not probabilities.

This delightful case does not seem essentially comparative. Each die has a 6D internal score – the numbers – which then determine the chances of victory. Die X will probably beat Die Y if X's highest four numbers are greater than the highest four numbers of Y. With this rule, the 'will probably beat' relation is cyclic:  $A > B > C > A$ . This is, roughly, a non-moral spin on Condorcet's Paradox.<sup>24</sup> With MAJORITY RULE (SCORING), we can get nontransitive betterness by thinking of each die as an outcome and the numbers (ordered highest to lowest) as six people's utilities. This case shows that the principles behind Condorcet's case have been right under our noses.

Temkin's helpful 'sports analogy', by contrast, is only superficially Condorcet-like.<sup>25</sup> Suppose we rank baseball teams based on their results:

#### MORE WINS

$A > B$  if A has more wins than B does against teams in the group.

A team's 'score' is its number of wins. So understood, a score isn't intrinsic to the team; it depends on the whole group.<sup>26</sup> But the score is still 'internal' in that it doesn't vary across comparisons; it is also 1D, securing TRANSITIVITY. The case may seem Condorcet-like if teams beat each other in cycles. But since wins are summed, MORE WINS is really more like the 1D INTERNAL ASPECTS VIEW. I mention this case because it nicely illustrates that, *pace* Temkin, it does not really matter for TRANSITIVITY whether values arise from intrinsic properties; what matters is that they not vary across comparisons (see fn. 15, above). Values can be 'internal' even if they depend on extrinsic relations.

Finally, notice that Condorcet's case only features a cycle because there are more than two 'voters'. Does that show that, whenever there are cycles and only *two* dimensions, we need an essentially comparative story? No! Consider Handfield's 'ertnog' cycle, which emerges just from mass and height. First a definition:

Suppose that  $x$  is *more ertnog* than  $y$  if and only if: (i)  $x$  and  $y$  differ in height by 2 or more inches and  $x$  is taller than  $y$ , or (ii)  $x$  and  $y$  do not differ by 2 or more inches in height and  $x$  has a greater mass than  $y$ . (2016: 6)

<sup>24</sup>There is a difference: the rule for comparing dice is *insensitive to permutation of dimensions*. It does not matter which three of Die A's faces say '6' and which three say '2'. All that matters is that there are three sixes and three twos. (The score could just as well be a multiset as a sextuple.) By contrast, order *does* matter to MAJORITY RULE (SCORING); it matters which individuals are which (see fn. 20, above), or more generally, which dimensions are which. Notice, for example, that (2, 1, 0) loses along most dimensions to (0, 2, 1), even though the only difference is the order of the numbers.

<sup>25</sup>As Temkin would agree (see Temkin 1987: 180; 2012: 468).

<sup>26</sup>As a result, and as Temkin is well aware (1987: 181), MORE WINS violates the Independence of Irrelevant Alternatives: whether  $A > B$  can depend on other teams. (Suppose A wins against B, but A loses to C and D whereas B wins against them. Here A is better than B *only thanks to C and D*.) In this way, MORE WINS is like the Borda Rule:  $A > B$  if voters rank A higher on average (see Katz 2011: 4, and fn. 8, above).

Next, measurements:

	A	B	C
<b>Height (in.)</b>	11	12	13
<b>Mass (lbs)</b>	5	4	3

Where ‘>’ means ‘more ertnog’, it follows that  $A > B > C > A$ , even though ‘>’ depends only on height and mass, which are intrinsic and invariant. The result is another cycle arising not from comparativity, but multidimensionality, this time with 2D ‘scores’. We can see here that MAJORITY RULE is not the only conceivable rule that can squeeze nontransitive rankings from internal aspects; it is just a striking example because it is not comparative in any way: each voter counts equally in each pairwise comparison. Ertnog, by contrast, is comparative in that mass only counts when height is close. (We will analyse this kind of comparativity in §9, below.)

Handfield’s case is also, obviously, not normative. There is no ethics of ertnog. But I submit that there is a simple normative analogue: Temkin’s ‘fireplace’ example.<sup>27</sup> Temkin (2012: 4) reports having had three housewarming options:

	Nothing	Fireplace1	Fireplace2
<b>Cost (\$)</b>	0	800	1,100
<b>Comfort</b>	None	Medium	High

The options may appear to form a cycle. Fireplace2 > Fireplace1 because a small comfort boost is worth over \$300; Fireplace1 > Nothing because a medium comfort boost is worth over \$800; and yet Nothing > Fireplace2 because \$1,100 is too much to spend on any amount of fireside comfort. This is like the ertnog structure: a big difference in cost is decisive, but when costs are small, comfort is decisive, or at least can be. (With ertnog, big height differences are decisive, and otherwise mass is decisive.) Contra Temkin, then, we seem to have a moral case better suited to a non-additive, multidimensional treatment than an essentially comparative one.

Once we distinguish the two sources of nontransitivity, we have double the resources with which to interpret Temkin’s example and others like it. Rather than exhibiting comparativity, some paradoxes might instead be exceptions to ADDITION.

## 6. Internalizing

Let’s recap.

We began with Temkin’s idea that TRANSITIVITY is linked to our views of value. The Internal Aspects View entails TRANSITIVITY; the ESSENTIALLY COMPARATIVE VIEW does not. But I have argued that this link only holds given 1D SCORING, or something like it (such as additive 3D scores). If we allow for complex scores, aggregated in the right ways, there can be nontransitivity even on the INTERNAL ASPECTS VIEW – defined as SCORES plus INTERNAL SCORING. The upshot is that

<sup>27</sup>This case is a warmup for my treatment of ‘Spectrum Arguments’ in §§7, 9 below.

INTERNAL SCORING is *not* the fundamental issue on which TRANSITIVITY depends. 1D SCORING, though less familiar, is just as crucial.

Now I want to propose a new question, which will be my topic for the second half of the paper. Once we give up 1D SCORING, is there any need for the ESSENTIALLY COMPARATIVE VIEW? Given enough dimensions of value, aggregated in sufficiently funny ways, it is not just Condorcet's case and Temkin's fireplace that seem to fit the 'internal aspects' treatment. For any case where an ESSENTIALLY COMPARATIVE VIEW delivers a nontransitivity, we may be able to *internalize* the case, in the following sense: we could cook up an INTERNAL ASPECTS VIEW with multiple dimensions and a non-additive aggregation rule that gives rise to the very same nontransitivity. There would then be no need for essential comparativity.

The question, then, is whether every (apparent) instance of essentially comparative value can, and perhaps should, be 'internalized', where this involves reinterpreting the case to involve internal scores aggregated in a non-additive way.<sup>28</sup> We will consider two examples: Spectrum Arguments and person-affecting principles. Temkin sees these as the home turf of the ESSENTIALLY COMPARATIVE VIEW. It would be surprising if even they could be internalized.<sup>29</sup>

## 7. A spectrum argument

Consider 100 populations. In the first, many people live fabulous lives. In the last, vastly more eke out lives barely worth living. The 98 populations in between fall on a spectrum; each is bigger, and less fabulous on average, than the last. Within each population, lives are equally good, and no member shows up in any other population.<sup>30</sup>

The picture is this, where 'How Many' tells us the size of the population, and 'How Good' tells us the quality of life for each member. (My presentation draws on Hare (2017).)

	P1	P2	P3	...	P98	P99	P100
<b>How Many</b>	10 <sup>11</sup>	10 <sup>12</sup>	10 <sup>13</sup>	...	10 <sup>108</sup>	10 <sup>109</sup>	10 <sup>110</sup>
<b>How Good</b>	100	99	98	...	3	2	1

<sup>28</sup>Compare this to the question of whether every deontological view can and should be consequentialized – i.e. reimagined as a consequentialist view that takes certain verboten acts, such as killings, to be intrinsically bad outcomes (or at least, to be outcomes that we have reason to disprefer). For a fuller discussion of consequentializing, see Louise (2004), Dreier (2011), Portmore (2011), Setiya (2018), Howard (2021) and Muñoz (2021).

<sup>29</sup>Less surprising is the converse 'Temkinizing Thesis': any moral ranking that can be modelled with an *n*-dimensional INTERNAL ASPECTS VIEW can also be modelled with a 1D ESSENTIALLY COMPARATIVE VIEW. Here is how. If an INTERNAL ASPECTS VIEW ranks  $x = y$ , the Temkinized view could say that  $x$  and  $y$  have the same value – say, 1 – when compared with each other. If an INTERNAL ASPECTS VIEW says that  $x$  is better or worse, the Temkinized view could set  $x$ 's value to 1 or –1, respectively, while setting  $y$ 's value to 0 – again, with both values relative to the context of that pair. And since all values are relative to context, the view can assign  $x$  or  $y$  a different value when the alternative is  $z$ . Upshot: the 1D ESSENTIALLY COMPARATIVE VIEW is extremely flexible. It can model nearly anything (though it does require '≥' to be reflexive and '>' to be asymmetric).

<sup>30</sup>The case is famous from Parfit's treatment of the Repugnant Conclusion (Parfit 1984: Ch. 17). Other cases, involving pains or pleasures rather than populations, are included under the general heading of 'spectrum arguments' for nontransitivity (Rachels 1993, 1998, 2001; Temkin 1994, 1996, 2012: Chs 5, 9; Binmore and Voorhoeve 2003; Huemer 2008; Handfield 2014; Nebel 2018).

Now the nontransitivity. When we compare populations 1 and 2, it seems clear that 2 is better, given that the population is so much bigger – ten times the people! – with only a tiny reduction in average quality of life. For the same reason, each population  $n > 1$  seems better than population  $n - 1$ . But when we compare populations 1 and 100, the first seems *worse*. Given that population 100 is vastly inferior in average quality of life, it seems that population 1 is positively better, all things considered, despite its puny size.

For Temkin, there is only one way to explain this apparent nontransitivity: his ESSENTIALLY COMPARATIVE VIEW. The ‘relevant factors’ (‘or the significance of those factors’) that determine betterness must change, depending on whether we are comparing P1 with P2 or P1 with P100.<sup>31</sup> What does this mean exactly? Temkin is not always clear.<sup>32</sup> But I think the basic idea is clear enough. *Higher quality* of life for each is decisive only when we are comparing populations with *big differences in quality* (see Temkin 1996: 193–196; 2012: 266). When quality is close – as with P1 vs. P2 – total quantity of wellbeing can be decisive. But when the gap in quality is a whopping 99 points – as with P1 to P100 – quantity is moot.

This is a breeze to model on the ESSENTIALLY COMPARATIVE VIEW with 1D SCORING. We could say that the value of a population P in a context is equal to the number of people times the average quality of life (its quantity) – *unless* P is being compared with another population P\* whose average quality is lower by at least 99, in which case P is given infinite value.<sup>33</sup> In this way, we get that each population  $n$  in the series is better than its predecessor  $n - 1$ , but population 1 is better than population 100.<sup>34</sup>

That is one option. But we can *also* treat the case using an INTERNAL ASPECTS VIEW, much like Handfield’s treatment of ‘ertnog’ and my take on Temkin’s fireplace (see §6).

Here’s the trick. We give each population P a two-dimensional score:  $\langle A_p, T_p \rangle$ , representing the average utility and total utility, respectively. (Where ‘utility’ measures quality of life, whatever that may be.)<sup>35</sup> On this 2D INTERNAL ASPECTS VIEW we then aggregate as follows:

<sup>31</sup>I am quoting from Temkin’s (1996: 194) treatment of a spectrum argument involving a series of pains from short and intense to long and dull.

<sup>32</sup>At one point, Temkin (1996: 191) suggests a 1D SCORING ‘model’ that is ‘something like’ the following (though he uses a different example; see fn. 31): adding more lives, all at a low level of pleasure, can only increase the overall value of the population up to 1; adding more lives, all at a high level of pleasure, can increase the overall value up to 10. We can suppose that the fairly hefty P1 has value 7, and that the gigantic P100 has a value near 1. This model can explain why repeatedly doubling the size of P1 will always lead to better populations, though never one that is better overall than P100 (cf. Binmore and Voorhoeve 2003). But the model can’t explain the *cycle*. Unless we *also* have a violation of INTERNAL SCORING, there is no way that P100 could be better than something that is better than something ... that is better than P1.

<sup>33</sup>We would have to gently loosen 1D SCORING to allow infinite values.

<sup>34</sup>For all  $n > 1$ ,  $V_{Pn-1}(Pn) \gg V_{Pn}(Pn-1)$ , but  $V_{P100}(P1) \gg V_{P1}(P100)$ .

<sup>35</sup>Of course, these 2D scores are simplified. We might eventually want to tack on more dimensions, or derive  $A_p$  and  $T_p$  from more detailed scores that represent each individual’s utility (like the scores in Condorcet’s case).

## QUANTITY

$P > P^*$  if  $|A_P - A_{P^*}| < 99$ , and  $T_P > T_{P^*}$ .

*Informally: when average utility is close, total utility is decisive.*

## QUALITY

$P > P^*$  if  $A_P - A_{P^*} > 99$ .

*Informally: when average utility isn't close, average utility is decisive.*

These rules, though simplistic, are vastly more plausible than MAJORITY RULE (SCORING), and they, too, can fit the INTERNAL ASPECTS VIEW.<sup>36</sup> The two dimensions really do measure how good a population is all things considered; the overall score is the key factor in determining betterness all things considered; and scores never change even as we swap out alternatives.<sup>37</sup> And yet we get a cycle:  $P1 > P2 > \dots > P100 > P1$ .

(Notice, by the way, that the nontransitivity here does not depend in any way on there being vagueness. Our simple model has a sharp cut-off for when total quantity starts trumping average quality, and the cycle does not need to include many tiny steps between  $P1$  and  $P100$ . There is a cycle between  $P1$ ,  $P100$  and *any* intermediate population – for example,  $P100 > P50 > P1$ . Now, I am not saying that there cannot be vagueness in this case. Maybe it's independently plausible that the cut-off should be vague. My point is just that vagueness is not essential to there being a nontransitivity here – or in Temkin's fireplace case and Handfield's ertnog case (§6). Multidimensionality can lead to cycles all on its own.)

We have just 'internalized' the Spectrum Argument, in the following sense. We have taken a putative nontransitivity that is supposed to be representable only on the ESSENTIALLY COMPARATIVE VIEW and shown how to model it on a multidimensional INTERNAL ASPECTS VIEW. Indeed, this doesn't just seem like *one* way to model the cycle; it seems like the *right* way to do it. Who needs essential comparativity?

## 8. A person-affecting principle

Let's internalize one more classic example.

Suppose that you are making a choice that will affect who exists. If you do *A*, the world will include a very happy Alan (utility: 10) and a content Chico (utility: 5). If you do *B*, the world will again feature Alan, merely content this time (utility: 5), alongside an ebullient Betty (utility: 20).

<sup>36</sup>QUALITY and QUANTITY are more plausible in cases without confounds such as inequality. (Compare them to Temkin's (2012: Ch. 2) two 'standard views'.)

<sup>37</sup>In the analogy with ertnog (Handfield 2016), quality is height, and quantity is mass.

	World A	World B
Alan	10	5
Betty	∅	20
Chico	5	∅

What should you do?

The utilitarian will say: make the world with the happiest people! Do B! But some say morality is not so concerned with making happy people; it is really about ‘making people happy’ (in Narveson’s (1967) slogan). This view is typified by principles like:

PERSON-AFFECTING PRINCIPLE

$A > B$  if  $A$  is better for someone and worse for no one.<sup>38</sup>

Where, crucially, we assume that doing  $A$  cannot be ‘better for’ or ‘worse for’ someone if the alternative for them is nonexistence. This principle is all about trying to do what is best for the people who will exist no matter what is done. It recommends, in a choice from  $A$  and  $B$ , that you do  $A$ .

But while the PERSON-AFFECTING PRINCIPLE may be appealing, it violates TRANSITIVITY and even ACYCLICITY. Compare doing  $B$  with doing  $C$ , where these give rise to the following worlds:

	World B	World C
Alan	5	∅
Betty	20	5
Chico	∅	20

Clearly, on the PERSON-AFFECTING PRINCIPLE,  $B > C$ , since  $B$  is better for Betty and worse for no one. But now compare:

	World C	World A
Alan	∅	10
Betty	5	∅
Chico	20	5

Here, the principle says  $C > A$ , since  $C$  is better for Chico and worse for no one. Put the pairs together, we have yet another cycle:  $A > B > C > A$ .

<sup>38</sup>This is a ‘narrow’ person-affecting view because it cares about how particular people fare, not how people are faring in general. See Ross (2015) for more on how to formulate and classify person-affecting principles.

What kind of view would allow this? Temkin's answer is, again, the ESSENTIALLY COMPARATIVE VIEW (2012: 422–423). The importance of Alan's welfare changes depending on whether we are comparing two options that both lead to his existence. That A is good for Alan is immaterial when the alternative C, but it greatly increases the value of A relative to B (since Alan will *also* exist, though less happily, if you do B).

But we can just internalize again, with 3D SCORES as in the snacks case. The only difference is that now we need to add gaps (' $\emptyset$ ') to represent the missing utilities of the nonexistent, and we need to stipulate that gaps disable dimensions: if  $a_i = \emptyset$  or  $b_j = \emptyset$ , then neither  $a_i > b_j$  nor  $b_j > a_i$ . We then aggregate using a rule known as (strong) Pareto:

PARETO (SCORING)

$(a_1, \dots, a_n) >> (b_1, \dots, b_n)$  if for some  $i$ ,  $a_i > b_i$  and for no  $j$ ,  $b_j > a_j$ .

Which says that a score is higher if it is higher along one dimension and lower along none. And this leads to the cycle above, all safely within the INTERNAL ASPECTS VIEW, without any need for essential comparativity.

## 9. Comparative effects from internal aspects

Now we are really pushing the INTERNAL ASPECTS VIEW to its limits.

In this paper's first half, I argued that the INTERNAL ASPECTS VIEW could allow *some* cases of nontransitive betterness. Now it seems that the view might underwrite *all* nontransitive betterness, or at least far more than expected. The PERSON-AFFECTING PRINCIPLE and Spectrum Arguments are the home turf of the ESSENTIALLY COMPARATIVE VIEW. But we were able to 'internalize' Temkin's take on them in a fairly natural way, reconstruing each to fit an INTERNAL ASPECTS VIEW with multiple dimensions.

What, then, do we make of the ESSENTIALLY COMPARATIVE VIEW? Should we give up on comparative values? One reason to say 'no' – the boring reason – is that we haven't looked at all of the examples. Perhaps we will eventually find some cases where the ESSENTIALLY COMPARATIVE VIEW resists internalizing. I'm open to this. It would be rash to chuck out the ESSENTIALLY COMPARATIVE VIEW too quickly.

There is, however, an interesting moral we can already draw from the two cases above. Some kinds of comparativity may be *compatible with*, or even *derivable from*, an INTERNAL ASPECTS VIEW. This is not 'essential comparativity', because scores aren't shifting. But it is comparativity in a broader sense: there is a value dimension that can count more or less depending on the options we are comparing.

Consider the 2D INTERNAL ASPECTS VIEW we used in the Spectrum Argument. The view has two principles, QUALITY and QUANTITY, which tell us how to combine our 2D scores. When quality is close, quantity is decisive; when quality is distant, quality is decisive. This is *lexicographic aggregation*. To determine whether  $P > P^*$ , we compare them one dimension of value at a time, in a specific order, and the better option overall



is the one that is the *first* to beat the other along a dimension.<sup>39</sup> In this case, quality comes first, and quantity comes second. P 'beats' P\* in quality if the average quality of life in P is 99 higher than life in P\*. P 'beats' P\* in quantity if the total quantity of good things in life is higher in P than in P\*.

This allows for a kind of comparative effect – *screening off*. Whenever P beats P\* in quality, *quantity doesn't matter* to whether  $P > P^*$ , even though it might be decisive in determining whether  $P^* > P^{**}$ . In our example above,  $P1 > P100$ , despite the massive quantity of good in P100, because quality screens off quantity. This is why P100's high total goodness is irrelevant *in this comparison* even though it *is* relevant, indeed decisive, when we compare P100 with P99, whose quality does *not* beat P100's.<sup>40</sup>

It is tempting to confuse effects such as screening-off, which are certainly comparative in some sense, with Temkin's ESSENTIALLY COMPARATIVE VIEW, which is basically the denial of the INTERNAL ASPECTS VIEW.<sup>41</sup> But there is a difference. Screening-off can fit the INTERNAL ASPECTS VIEW; indeed, it can be the *effect* of aggregating internal scores in a lexicographic way. Average quality and total quantity are internal to P1, but how quantity affects *betterness* depends on how close the alternative is in quality.

Now, what about the PERSON-AFFECTING PRINCIPLE? It's not lexicographic, but it also seems comparative, in some sense. We have three dimensions – one per person – and we are ignoring any dimension that has a gap in it. For example, when we compare A with B, we have the following scores: (10,  $\emptyset$ , 5) and (5, 20,  $\emptyset$ ). Only the first dimension, goodness for Alan, makes a difference; so,  $A > B$  even though B is great for Betty. But when we compare A with C, goodness for Alan is irrelevant, since Alan does not exist if C is chosen; its score is ( $\emptyset$ , 5, 20), with a gap in Alan's dimension.

This is not quite 'screening off'. It isn't as though Alan's wellbeing is being trumped by some inherently superior value.<sup>42</sup> Instead we have *gappy aggregation*.<sup>43</sup> If there is a gap in one of A's dimensions, that dimension does not affect whether A is better or worse than B. This seems to be compatible with the INTERNAL ASPECTS VIEW, since how good A is for Betty – and *whether* its goodness for Betty is defined at all – is fully intrinsic to A. The role of the gap in A's score, we might say, is to 'disable' the dimension, to make B's goodness for Betty irrelevant to whether  $A > B$ .

<sup>39</sup>For the formal definition, see Fishburn (1974: 1443). Why 'lexicographic'? Because the process is like the one that determines which of two words goes first in a dictionary. (Technically, the example in the text is a 'lexicographic semiorde', because A is not better than B *whenever* A is better in terms of quality; A is better whenever it is better than B in quality *by a certain margin*. Compare with the example described in Tversky (1969). Note also that the view focuses on the difference in quality between A and B, not whether A's quality passes some absolute threshold; see Klockseim (2016: 1319–1321) on 'threshold lexicality', which preserves TRANSITIVITY, and 'sufficient-difference lexicality', which does not.)

<sup>40</sup>Of course, P99's quality is *higher than* P100's. But the kind of 'beating' that matters in this method of aggregation is being *higher than by 99 points or more*.

<sup>41</sup>I say 'basically' because the ESSENTIALLY COMPARATIVE VIEW does include one positive commitment – SCORES. What kind of ethical view would reject this? I do not know. But I am reminded of a quip by Ernst Gombrich, who, when asked to name his favourite painter, replied, 'I don't give marks to artists'. See *The Charlie Rose Show*, Interview with Ernst Gombrich. PBS. Aired 22 September 1995.

<sup>42</sup>Chico isn't Lord of the Universe!

<sup>43</sup>For an example that may involve gappy aggregation, see Dancy (2005: 14). For a recent discussion of Pareto and incommensurable dimensions, see Steele (2021).

But there is a problem. We can get the same effect as gappy aggregation with the ESSENTIALLY COMPARATIVE VIEW. Recall our three options:

	World A	World B	World C
Alan	10	5	$\emptyset$
Betty	$\emptyset$	20	5
Chico	5	$\emptyset$	20

Now, when we compare A with B, we might think of their values simply as the two gappy triples: (10,  $\emptyset$ , 5) and (5, 20,  $\emptyset$ ). But notice that only the first dimension matters here, since it is the only one without a gap. So why not prune the other dimensions? We would then get svelte new scores:  $V_B(A) = (10)$  and  $V_A(B) = (5)$ . We can then get the same result as before with PARETO (SCORING):  $A > B$ . Ditto for the other pairs. Here, the betterness relations are exactly the same as before, but the triples are no longer values. They are instead ‘proto-values’ that determine an option’s value depending on context. For example, A’s proto-value is (10,  $\emptyset$ , 5). This means (i) A’s value is (10) when compared with something whose value has a gap in the third but not the first dimension; (ii) A’s value is (5) when compared with something whose value has a gap in the first but not the third dimension; and (iii) A’s value is (10, 5) when compared with something whose value has no gap in the first or third dimensions. The triple is not a value itself, but it tells us how to find the 2D or 1D value.<sup>44</sup>

Here we have two views that seem like notational variants. One says that values are internal, with gappy dimensions that can be ignored during aggregation; the other view says that proto-values are internal, with gappy dimensions that are pruned to arrive at an option’s true context-relative value. Both views generate cycles, and both involve a kind of multidimensionality. To decide between them, we would need some kind of a priori constraint on value theories to serve as a tiebreaker.

Is there any such tiebreaker? I think so. One constraint on a value theory is that there should be exactly one dimension of value for each genuine and distinct way in which things can be good or bad. (The ‘dimensionality constraint’, if you like.) For example, if A’s value is (10,  $\emptyset$ , 5) – where the dimensions reflect the utilities of Alan, Betty and Chico – it follows that A is good to degree 10 with respect to Alan’s utility, that A is good to degree 5 with respect to Chico’s utility, and that these are not the same way of being good: (10,  $\emptyset$ , 5) is not the same value as (5,  $\emptyset$ , 10). The two utilities are not fungible in ethical value as dollar bills are in their value as cash.<sup>45</sup>

Now, to my mind, one respect in which B is good is that it is good for Betty. I cannot say that B is *better* for Betty than A is, since she does not exist given A, nor can I say that A is good with respect to Betty’s happiness (there it is undefined). But I

<sup>44</sup>A fuller statement of this view would have to answer further questions. Could there be a zero-dimensional value? Should we distinguish different 1D values that are the result of pruning different dimensions? Note that we could also have multidimensional scores given this kind of ESSENTIALLY COMPARATIVE VIEW. To illustrate, imagine that there is a fourth person, Dina, whose utility is 1 at each of the three worlds. Then  $V_B(A) = (10, 1)$ .

<sup>45</sup>For more on nonfungible utilities (and the contrast with cash), see Chappell (2015).

can still say that B is good in virtue of Betty's happiness. That suggests that B's overall value score has to include a 'Betty' dimension with a score of 20. B's value cannot just be, say, (5). And yet, (5) is B's value when compared with A, on the ESSENTIALLY COMPARATIVE VIEW that prunes gappy dimensions. That's why I think we should prefer the view on which B has the internal value of (5, 20,  $\emptyset$ ).

I conclude that the PERSON-AFFECTING PRINCIPLE is probably better with gappy internal values than with gapless comparative values. For similar reasons, I think the Spectrum Argument is best understood with 2D internal values; average utility and total utility seem like genuine and distinct dimensions of value.<sup>46</sup>

But I might be wrong. In fact, I might be very wrong. On some views, there cannot *in principle* be any tiebreaker between two value theories so long as they agree about which value relations obtain between which things; any further difference is merely notational. I call this the:

#### VALUE EQUIVALENCE THESIS

Two theories of value are equivalent if they always agree, for any options  $x$  and  $y$ , on whether  $x$  is as good as  $y$ , all things considered, and whether  $y$  is as good as  $x$ , all things considered.<sup>47</sup>

If this is true, there is nothing more to a theory of value than what it says about overall value relations. If it's false, there must be more to values than the relations they underwrite. I myself suspect that the thesis is false. There should be many dimensions of value that do not change relative to the alternative, since there are many ways, independent of the alternatives, in which a thing can be good.

Let's sum up. The Spectrum Argument can be internalized as a case of lexicographic aggregation; this leads to a kind of comparativity – namely, screening – that isn't 'essentially' comparative. The PERSON-AFFECTING PRINCIPLE can be internalized as a case of gappy aggregation; this, too, leads to a kind of non-essential comparativity – namely, the disabling of dimensions. In both cases, I think the internalized view is an improvement, since there ends up being one value dimension for each genuine and distinct way in which things can be good. My

<sup>46</sup>We could get the same cycles on an essentially comparative view. The trick is to say that quantity is pruned when quality is decisive, and quality is pruned when quantity is decisive. On this proposal, when comparing P1 with P100, we prune the quantity dimension so that P1's true value is (100) and P100's is (1). But this implies that when we compare P1 with P100, P100 cannot be good with respect to quantity, which it clearly is. P100's high quantity of utility is, regardless of contest, a genuine way in which P100 is good. Rather than deny the goodness of high quantity, we should say that quantity is screened off from affecting overall betterness, whenever there is a big difference in average quality. (For relevant discussion, see Dancy (2005: 15), and see Parfit (1984: 412) on 'personal value' that 'does not make the outcome better' – it's still 'value!')

<sup>47</sup>This thesis is endorsed by Broome (1999: 163). Compare VALUE EQUIVALENCE THESIS to the 'deontic equivalence thesis', which says that theories of rightness are equivalent if they agree about which acts are right (see the 'consequentializers' in fn. 28, above). I should note two other reasons why one might doubt the VALUE EQUIVALENCE THESIS: (1) judgements of the form ' $x$  is  $n$  times as good as  $y$ ' may not be reducible to 'is as good as'; (2) also irreducible might be the difference between parity and 'incomparability', since both involve neither of two options being as good as the other (see Chang 2002). We could broaden the VALUE EQUIVALENCE THESIS so that equivalent theories have to agree about parity and ratio-judgements (and perhaps other value relations). My thanks to a referee for helpful comments about how to state the thesis.

assumption here is that values, and value dimensions, have some independent reality and can be assessed in themselves; they are not just ‘whatever determines betterness’. That said, I have left open the possibility that I am wrong about this, and I have not leapt to the conclusion that we should *always* internalize, though I think we probably should do it *sometimes*.

The moral is that we do not want a sharp divide between, on the one side, the INTERNAL ASPECTS VIEW, and on the other, all comparative phenomena. Complex scores make the relationship more nuanced and interlinked. To see a simple dichotomy here is restrict one’s view of some of the most dazzling ideas in axiology – lofty hypotheses about the nature of value – to the limited lens of 1D SCORING. Why look at the Milky Way through a magnifying glass?<sup>48</sup>

## 10. Conclusion

We began with a simple story: the fate of TRANSITIVITY is to be decided by a clash between two titans, the ESSENTIALLY COMPARATIVE VIEW and the INTERNAL ASPECTS VIEW. If value is internal, betterness must be transitive; if value is essentially comparative, betterness need not be transitive.

I have urged a rethinking of the TRANSITIVITY debate for two reasons. First, the INTERNAL ASPECTS VIEW cannot be the source of TRANSITIVITY, since there can be cycles even without comparativity; my main example was an ethical version of Condorcet’s Paradox. In this case, value scores do not change with context; instead, they are just complex, with multiple dimensions that aggregate in a non-additive way.<sup>49</sup>

Second, the INTERNAL ASPECTS VIEW can itself explain some cases that are supposed to be the exclusive purview of the ESSENTIALLY COMPARATIVE VIEW. By ‘internalizing’ these cases, we can also see how some kinds of comparativity, like screening-off, can flow from complex internal scores. This raises the question of whether internalizing changes the substance of a view; I suggested that it does, though I think it is too early to say for sure.

Let me finally return to a question raised in §4. Could there be an aggregation rule that is far enough away from ADDITION to violate TRANSITIVITY, but not so close to MAJORITY RULE that it is just ridiculous? We have seen a couple of candidates, but more importantly, now we know how to find more. Rather than conjuring principles from scratch, we can try internalizing more of the greatest hits of the ESSENTIALLY COMPARATIVE VIEW, bearing in mind that each dimension of value should embody a genuine and distinct way in which things can be good.

<sup>48</sup>I do not mean that 1D SCORING is false – only that it is unsuitable as a background assumption when trying to map out the space of possible moral views. There is a lot more space than ethicists have realized, and here I think it will be fruitful to mix and match ideas from philosophy and economics.

<sup>49</sup>There is nothing comparative about the 3D INTERNAL ASPECTS VIEW as it applies to our Condorcet case. That said, we could ‘Temkinize’ the view in a somewhat natural way. (See fn. 29 for a general but crude recipe for Temkinizing – i.e. turning a multidimensional view into a 1D ESSENTIALLY COMPARATIVE VIEW.) We can reconceive the 3D values – which tell us the utilities of Alan, Betty and Chico – as proto-values. We then say that  $V_B(A)$  is equal to the number of dimensions along which A’s proto-value is higher than B’s. The result would be the same cyclic judgements of overall betterness:  $A > B > C > A$ . It is not obvious which of these views is better; we would need to know more about whether, on these views, the individuals’ utilities each reflect a genuine and distinct way in which outcomes can be good. I myself would think they do. But if they do not, and goodness really is just a measure of the *number* of people who benefit more, we should prefer the Temkinized view.

Whatever method we use, and whatever answers we ultimately find, I hope you share my sense that there is much to explore, thanks in large part to pioneers such as Temkin. I have been critical of his views, but in a way we are on the same quest: we want to identify (and scrutinize) the source of our attraction to TRANSITIVITY. We want to understand how properties of *value relations* might arise from *what values are*. A clear map of the conceptual space here does not put a stop to anybody's questing; it invites us to go further, in all sorts of new dimensions.<sup>50</sup>

**Acknowledgements.** My thanks to the editors at *Economics and Philosophy* and to several anonymous referees, including three at this journal alone, for constructive comments. For discussion, I thank Alison Ross, Zach Barnett, Luc Bovens, Jonathan Dancy, Geoff Sayre-McCord, Tyler Doggett, Al Hájek, Nathaniel Baron-Schmitt, Daniel Cohen, Lloyd Humberstone, Kerah Gordon-Solmon, and audiences at Monash University, the Australian Catholic University, the Australian National University, and the University of North Carolina at Chapel Hill. Three friends deserve special thanks: Brian Hedden, who gave detailed feedback on an early draft, as well as Toby Handfield and Theron Pummer, who encouraged me to go beyond Condorcet's case to talk about internalizing.

## References

- Arrow K. 1963. *Social Choice and Individual Values* (2nd Edn). New York, NY: John Wiley & Sons.
- Arrow K. and H. Reynaud 1986. *Social Choice and Multicriterial Decision-Making*. Cambridge, MA: MIT Press.
- Baker K.M. 1975. *Condorcet: From Natural Philosophy to Social Mathematics*. Chicago, IL: University of Chicago Press.
- Binmore K. and A. Voorhoeve 2003. Defending transitivity against Zeno's Paradox. *Philosophy and Public Affairs* 31, 272–279.
- Broome J. 1999. *Ethics Out of Economics*. Cambridge: Cambridge University Press.
- Chang R. 2002. The possibility of parity. *Ethics* 112, 659–688.
- Chang R. 2005. Parity, interval value, and choice. *Ethics* 115, 331–350.
- Chappell R.Y. 2015. Value receptacles. *Noûs* 49, 322–332.
- Coons C. 2014. Hope for fools: four proposals for meeting Temkin's challenge. *Analysis* 74, 292–306.
- Cusbert J. 2017. Acting on essentially comparative goodness. *Thought* 6, 73–83.
- Dancy J. 2005. Essentially comparative concepts. *Journal of Ethics and Social Philosophy* 1, 1–15.
- Davidson D., J. McKinsey and P. Suppes 1955. Outline of a formal theory of value, I. *Philosophy of Science* 22, 140–146.
- Dreier J. 2011. In defense of consequentializing. In *Oxford Studies in Normative Ethics, Volume 1*, ed. M. Timmons, 97–119. Oxford: Oxford University Press.
- Fishburn P.C. 1974. Lexicographic orders, utilities and decision rules: a survey. *Management Science* 20, 1442–1471.
- Friedman A. 2009. Intransitive ethics. *Journal of Moral Philosophy* 6, 277–297.
- Gehrlein W.V. 2006. *Condorcet's Paradox*. Berlin: Springer.
- Gert J. 2004. Value and parity. *Ethics* 114, 492–510.
- Handfield T. 2014. Rational choice and the transitivity of betterness. *Philosophy and Phenomenological Research* 89, 584–604.
- Handfield T. 2016. Essentially comparative value does not threaten transitivity. *Thought* 5, 3–12.
- Hare, C. 2017. The great spectrum paradox. *Unpublished manuscript*. <http://web.mit.edu/~casparh/www/Papers/CJHareGreatSpectrumParadox.pdf>
- Howard C. 2021. Consequentialists must kill. *Ethics* 131, 727–753.
- Huemer M. 2008. In defence of repugnance. *Mind* 76, 62–72.
- Huemer M. 2013. Transitivity, comparative value, and the methods of ethics. *Ethics* 123, 318–345.
- Humberstone L. 2016. *Review of James W. Garson, Modal Logic for Philosophers*, Second Edition. Cambridge: Cambridge University Press.

<sup>50</sup>I myself have tried to 'go further' with multidimensionality in a paper about whether one may save the few instead of the many (Muñoz [Forthcoming](#)).

- Hurley S. 1985. Supervenience and the possibility of coherence. *Mind* **94**, 501–525.
- Katz L. 2011. *Why the Law is So Perverse*. Chicago, IL: University of Chicago Press.
- Katz L. 2015. On Larry Temkin's *Rethinking the Good*. *Journal of Moral Philosophy* **12**, 414–427.
- Klockseim J. 2016. How to accept the transitivity of better than. *Philosophical Studies* **173**, 1309–1334.
- Louise J. 2004. Relativity of value and the consequentialist umbrella. *Philosophical Quarterly* **54**, 518–536.
- May K.O. 1954. Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* **22**, 1–13.
- Muñoz D. 2021. The rejection of consequentializing. *Journal of Philosophy* **118**, 79–96.
- Muñoz D. Forthcoming. The many, the few, and the nature of value. *Ergo*.
- Narveson J. 1967. Utilitarianism and new generations. *Mind* **76**, 62–72.
- Nebel J.M. 2018. The good, the bad, and the transitivity of *better than*. *Noûs* **52**, 874–899.
- Parfit D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Persson I. 2014. Internal or external grounds for the nontransitivity of 'better/worse than'. *Law, Ethics, and Philosophy* **2**, 120–138.
- Portmore D.W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.
- Pummer T. 2018. Spectrum arguments and hypersensitivity. *Philosophical Studies* **175**, 1729–1744.
- Rabinowicz W. 2008. Value relations. *Theoria* **74**, 18–49.
- Rachels S. 1993. *A theory of beneficence*. Undergraduate thesis, University of Oxford.
- Rachels S. 1998. Counterexamples to the transitivity of *better than*. *Australasian Journal of Philosophy* **76**, 71–83.
- Rachels S. 2001. A set of solutions to Parfit's problems. *Noûs* **35**, 214–238.
- Roberts M.A. 2014. Temkin's essentially comparative view, wrongful life, and the mere addition paradox. *Analysis* **74**, 306–326.
- Ross J. 2015. Rethinking the person-affecting principle. *Journal of Moral Philosophy* **12**, 428–461.
- Sen A. 2017. *Collective Choice and Social Welfare: Expanded Edition*. Cambridge, MA: Harvard University Press.
- Setiya K. 2018. Must consequentialists kill? *Journal of Philosophy* **115**, 92–105.
- Steele K. 2021. Incommensurability that can(not) be ignored. In *Value Incommensurability*, eds. H. Andersson and A. Herlitz, 231–246. New York, NY: Routledge.
- Temkin L. 1987. Intransitivity and the mere addition paradox. *Philosophy and Public Affairs* **16**, 138–187.
- Temkin L. 1994. Weighing goods: some questions and comments. *Philosophy and Public Affairs* **23**, 350–380.
- Temkin L. 1996. A continuum argument for intransitivity. *Philosophy and Public Affairs* **25**, 175–210.
- Temkin L. 1999. Intransitivity and the person-affecting principle: a response. *Philosophy and Phenomenological Research* **59**, 777–784.
- Temkin L. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.
- Temkin L. 2014a. Reply to Persson: intransitivity and the internal aspects view. *Law, Ethics and Philosophy* **2**, 139–152.
- Temkin L. 2014b. Summary of *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. *Analysis* **74**, 265–291.
- Temkin L. 2015a. Rethinking *Rethinking the Good*. *Journal of Moral Philosophy* **12**, 479–538.
- Temkin L. 2015b. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning: A précis*. *Journal of Moral Philosophy* **12**, 363–392.
- Temkin L. 2020. To boldly go where no man, or woman, has gone before! In *Explorations in Ethics*, ed. D. Kaspar, 81–115. Cham: Palgrave Macmillan.
- Tversky A. 1969. Intransitivity of preferences. *Psychological Review* **75**, 31–48.

Daniel Muñoz is Assistant Professor of Philosophy at UNC Chapel Hill, where he is also Core Faculty in Philosophy, Politics, and Economics. He is writing a book with Sarah Stroud called *50 Paradoxes, Puzzles, and Thought Experiments in Ethical Theory* (Routledge, under contract). Other projects include *What We Owe to Ourselves* (proposal under review) and *How Values Combine: Multidimensionality in Ethics, Economics, and Beyond* (with Brian Hedden). <https://philosophy.unc.edu/people/daniel-munoz/>