KEN BINMORE AND
ALEX VOORHOEVE

Defending Transitivity against Zeno's Paradox

Recently, Larry Temkin and Stuart Rachels have offered a new argument against the transitivity of the relationship 'all things considered better than'.[1] This argument, in its various guises, invokes our intuitions about our preferences over different bundles of pleasurable or painful experiences with different durations, which, it is argued, will typically be intransitive. This article defends against this argument the orthodox view that the relation 'all things considered better than' should be regarded as transitive by showing that Temkin and Rachels are mistaken in supposing that a preference relation satisfying their assumptions must be intransitive. It makes clear where the argument goes wrong by showing that it is a version of Zeno's paradox of Achilles and the tortoise.

Their argument centers around two kinds of counterexamples to the transitivity of 'all things considered better than', one involving painful and the other involving pleasurable experiences of differing duration. Since Temkin and Rachels both offer an example of painful experiences, we begin by discussing this case in Section I. Section II explains why the argument fails. Section III explains why the case of pleasure is of the

same form, and can be solved in the same way. Section IV contains some general remarks on counterexamples to the transitivity of 'all things considered better than'.

I. THE ARGUMENT

Temkin and Rachels's pain counterexample rests on the following three claims, which they hold to be true:[2]

> *Claim 1:* For any unpleasant or 'negative' experience, no matter what the intensity and duration of that experience, it would be better to have that experience than one that was only a little less intense but that lasted much longer.[3]
>
> *Claim 2:* There is a finely distinguishable range of unpleasant or 'negative' experiences ranging in intensity from mild discomfort to extreme agony.
>
> *Claim 3:* No matter how long it must be endured, mild discomfort is preferable to extreme agony for a significant amount of time.

Temkin and Rachels believe that these claims taken together contradict the transitivity of 'all things considered better than'. They invite us to imagine first a lengthy life that contains an amount of excruciating torture of significant but relatively short duration. Let us call this combination of torture and the time it must be endured T_0. It follows from the first claim that one would prefer a life containing T_0 to an otherwise identical life that contains T_1, where T_1 represents slightly less intense torture for a much longer period of time. Claim 2 then allows us to apply the first claim repeatedly with a slightly lower intensity of torture and a longer time period to generate a chain of preferences, which Temkin and Rachels argue runs from T_0 (excruciating torture for a short period of time) at one end to T_{MILD} (a mild pain for a very long time) at the other. Temkin envisages the mild pain as a hangnail, Rachels as a slight headache. Each member of this chain is preferred to its successor. By transitivity, T_0 is preferred to T_{MILD}. But by claim 3, T_{MILD} is preferred to T_0, so that claims 1, 2, and 3 generate an intransitive preference.

---

2. Rachels, "Counterexamples," pp. 72–75 and pp. 78–79 and Temkin, "Continuum," p. 179.
    3. Rachels, "Counterexamples," p. 73, mentions 100 times as long, while Temkin mentions twice as long. Nothing depends on this number.

Temkin and Rachels deduce far-reaching conclusions from their purported demonstration that it may be reasonable for people to have intransitive preferences. Rachels advocates abandoning what he calls "maximizing theories" for evaluating the goodness of states of affairs in favor of his own "quasi-maximizing theory" that embraces intransitivity.[4] Temkin quotes Derek Parfit to the effect that the argument leads to a "general skepticism about practical reasoning," and concludes that "whatever we say, in the end, about my arguments, they may require us to seriously rethink our understanding of the good, moral ideals and the nature of practical reasoning."[5] We believe no such drastic measures are required, since, as we show in the next section, claims 1, 2, and 3 are consistent with transitivity.

II. Why the Argument Fails

One case of a person of whom claims 1, 2, and 3 are true, but who holds transitive preferences, suffices to show that the argument is wrong. A person who maximizes a utility function has transitive preferences. But a person who maximizes the utility function

$$u(p, t) = \frac{-pt}{(1+t)}$$

where u is utility, $p \geq 0$ is the intensity of pain, and $t \geq 0$ the length of time it must be endured, satisfies the three claims. Such Cobb-Douglas utility functions are standard in economics when representing trade-offs between different commodities. The only differences are the negative sign (because the commodities are bads rather than goods) and the denominator, which represents the fact that for this person, an extra hour of pain is less troublesome after many days of pain than after a few hours of pain.

A person with utility function u satisfies claim 1, since for every level of pain, if we slightly decrease the intensity of pain and significantly increase the duration he must bear it, he will prefer the bundle with the slightly greater amount of pain. For example, suppose the level of pain

4. Rachels, "A Set of Solutions to Parfit's Problems."
5. Temkin, "Continuum," pp. 209–10.

for excruciating torture equals 10 and the duration the torture must be endured is 2. For a person with utility function u the disutility of this (pain, duration) bundle is 6.7. Unpleasant as it is for him, this person will still prefer this bundle to a (pain, duration) bundle of (9,4), the disutility of which is 7.2. Utility function u is also compatible with claim 2, since it allows for a continuum of levels of pain. Finally, for a person with a utility function like u, claim 3 holds as well: a mild pain, no matter how long endured, will simply never generate the disutility of the pain of a significant amount of extreme torture. For suppose that the level of pain of a hangnail equals 1, and consider again subjecting our protagonist to excruciating torture of level 10 for duration 2. As we saw, the disutility of this amount of torture for someone with u is 6.7. Now, no matter how large we make t, the disutility of a hangnail can never exceed 1.[6] For this person, as for the "competent judges" invoked by Stuart Rachels and the audiences polled by Larry Temkin, there is a fundamental gap between the badness of being tortured for a significant amount of time and the badness of enduring a hangnail. As Temkin writes:

> When I imagine having a hangnail for a very long time, it is not as if I imagine my situation getting closer and closer to being as bad as getting tortured for two years, but at such a slow rate that my imagination gives out long before I ever reach such a point. Rather, I imagine that there is a fundamental gap between the pain of being tortured for two years and the pain of the hangnail, and that gap is no closer to being bridged after 1000 years than it was after 100, or 50, or 10.[7]

Rachels and Temkin go wrong when they suppose that the chain of preferences $T_0 > T_1 > \cdots > T_n$ over bundles of intensity of pain and its duration that is generated by the repeated application of claims 1 and 2 will necessarily reach a bundle T_{MILD}, where the level of pain is equal to that of a hangnail or a slight headache. For a person with a

---

6. Note how our proposed utility function matches with Temkin, "Continuum," p. 192: "My model for this is something like the following. Torture's badness might range from 0 to 10, depending on its duration, with two years of torture being, say, a 7. A hangnail's badness might range from 0 to 1. Prolonging a hangnail increases the value of the decimal places representing its 'badness score', but the fundamental gap between 1 and 7 is never affected."

7. Temkin, "Continuum," pp. 191–92.

utility function like u, the levels of pain in the chain $T_0 > T_1 > \cdots > T_n$ will converge to some limit larger than the mild pain of a hangnail or a slight headache.

It may be illuminating to note that this argument is a version of Zeno's paradox of Achilles and the tortoise. To recall, Zeno argued that if given a head start at the time t_0 when the race starts, the tortoise can never be caught by Achilles, even though Achilles is faster, since at the moment t_1 when Achilles is where the tortoise was when Achilles started, the tortoise will have moved on. When Achilles reaches the tortoise's new position at time t_2, the tortoise will have moved on again. Continuing this way, we construct an infinite sequence $t_1 - t_0$, $t_2 - t_1$, . . . of the successive time intervals it takes Achilles to get to where the tortoise was when Achilles started each leg of his attempt to catch the tortoise. Zeno argued that the terms of this infinite sequence must sum to infinity, and hence the tortoise is never overtaken, but they actually sum to a finite number.

Figure 1 explains why the argument criticized in this article is a version of Zeno's paradox. The points T_0 and T_{MILD} are, respectively "two years of extreme pain" and "a very long time with a hangnail or a slight headache." With the utility function u, a person will be indifferent between all combinations of pain and duration that lie on a curve such as that drawn through T_0. The arrows show the direction of his preferences: he will prefer all points to the left of an indifference curve to any point on the curve, and he will prefer any point on the curve to all points that lie to the right of this curve. Note that each indifference curve has a vertical asymptote. The bold indifference curve drawn through T_0 has a vertical asymptote at the pain level a. As we decrease the pain intensity while remaining on the indifference curve (thereby remaining at the same utility level) the duration of the pain increases to infinity, but the pain intensity never falls below a.

The points T_0, T_1, T_2, and T_3 show the first elements of the chain Temkin and Rachels must construct for their argument. Like the sum of the successive time intervals in the story of Achilles and the tortoise in Zeno's paradox, the sequence p_0, p_1, p_2, . . . converges to some limit ℓ. Since $\ell \geq a > p_{\mathrm{MILD}}$, p_n can never reach p_{MILD}. The attempt to construct a chain of preferences starting with T_0 and ending with T_{MILD} therefore fails, no matter how much time we associate with T_{MILD}.
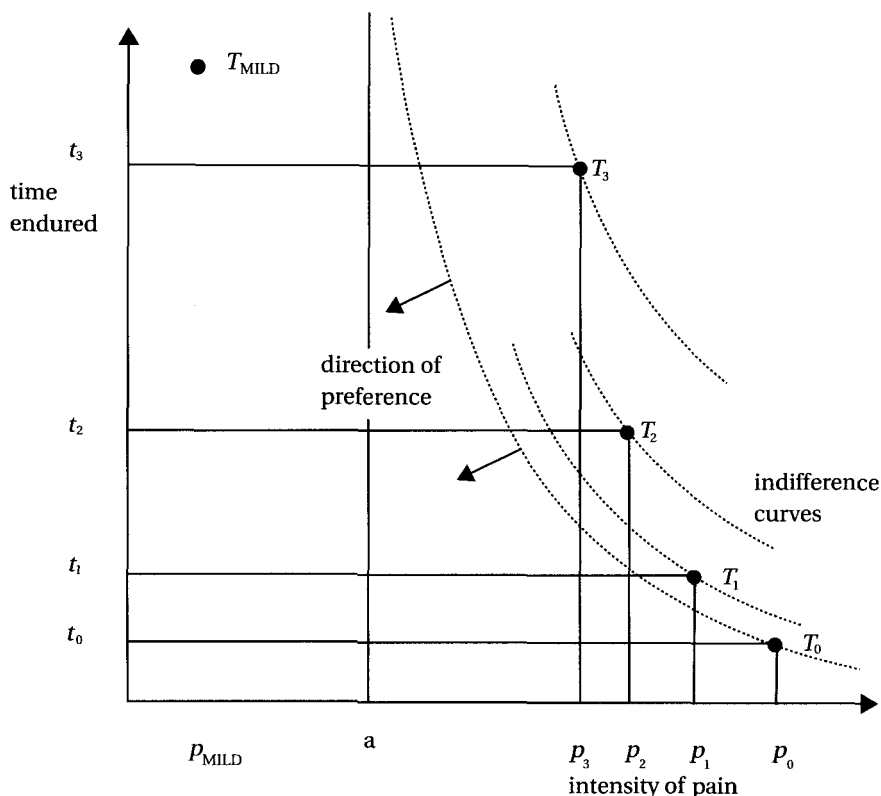
FIGURE 1
Indifference Curves in the Space of (Pain, Duration) Bundles

III. THE CASE OF PLEASURE

In several articles, Rachels offers an analogous counterexample to transitivity based on an analysis of a series of pleasurable experiences of differing intensity and duration.[8] He posits the following claims:

> *Claim 1*: For any experience of pleasure, no matter what the intensity and duration of that pleasure, it is better to experience a slightly less intense pleasure that lasts 100 times as long.[9]

8. Rachels, "Counterexamples," pp. 75–78, "A Set of Solutions," pp. 215–16, and "Intransitivity."
9. In "A Set of Solutions," Rachels calls this the "principle of duration."

Claim 2*: There is a finely distinguishable range of pleasurable or 'positive' experiences ranging in intensity from ecstasy to mild pleasure. *Claim 3**: No finite duration, no matter how long, of mild pleasure is as good as any significant duration of ecstasy, no matter how short.[10]

Rachels then invites us to imagine first an experience of ecstasy of significant duration. Let us call this bundle of ecstasy and time it is experienced E_0. It follows from claim 1* that one would prefer a bundle E_1, where E_1 represents slightly less pleasure lasting 100 times as long, to E_0. Claim 2* then allows us to apply the first claim repeatedly with a slightly lower intensity of pleasure and a longer time period to generate a chain of preferences, which Rachels argues runs from E_0 (ecstasy for a short, but significant period of time) at one end to E_{MILD} (mild pleasure for an extremely long time) at the other. Each subsequent member of this chain is preferred to its predecessor. By transitivity, E_{MILD} is preferred to E_0. But by claim 3*, E_0 is preferred to E_{MILD}, so that claims 1*, 2*, and 3* generate an intransitive preference.

Whether or not these claims are plausible, this counterexample fails once again because we can construct a utility function for a person of whom claims 1*, 2*, and 3* are true. A person who maximizes the utility function

$$u(e, t) = \frac{et}{(1+t)}$$

where u is utility, $e \geq 0$ is the intensity of pleasure, and $t \geq 0$ the length of time it is enjoyed, satisfies the three claims. This is the same utility function as before but without the negative sign, because the commodities are now goods. As before, the denominator entails that an extra hour of bliss is less wonderful after many days of bliss than after a few minutes. For a person with this utility function, a night of ecstasy is worth more than a lifetime of bourgeois comfort. Indeed, it might be said to represent well the feelings of Dostoevsky, whom Rachels approvingly quotes.

"In certain moments, I experience a joy that is unthinkable under ordinary circumstances, and of which most people have no comprehension. Then I feel that I am in complete harmony with myself and the whole world, and this feeling is so bright and strong that you could give up

10. In "A Set of Solutions," Rachels calls this the "principle of lexicality."

ten years for a few seconds of that ecstasy—yes, even your whole life."[11]

We may conclude, contra Rachels, that should Dostoevsky also have held claim 1* to be true, we would nevertheless have no reason to presume his preferences were intransitive.

IV. A REMARK ON COUNTEREXAMPLES TO TRANSITIVITY

Temkin's and Rachels's arguments fail. But suppose that with a different example we could all be persuaded that our choices in a pairwise comparison of three alternatives A, B, and C express a genuinely intransitive preference relation: $A > B > C > A$. An economist would then say that this determines how we would choose from each of the feasible sets $\{A, B\}$, $\{B, C\}$, and $\{C, A\}$. He would also say, however, that we have failed to consider everything that matters because we do not specify how we would choose from the full set $\{A, B, C\}$. Whatever we plan to choose from this set, say A, we are in trouble if someone now removes C from $\{A, B, C\}$ on the grounds that its presence is irrelevant because it is not going to be chosen and because our valuation of the options A and B is independent of its presence or absence. When C is gone, we are left with $\{A, B\}$, from which we are committed to choose B. We believe that a notion of rational choice that allows such violations of the Independence of Irrelevant Alternatives would lack any useful content.[12]

---

11. Rachels, "Counterexamples," p. 77.
12. The Independence of Irrelevant Alternatives was introduced by John Nash in "The Bargaining Problem," *Econometrica* 18 (1950): 155–62. It differs from Kenneth Arrow's well-known condition of the same name that relates individual and social preferences. Amartya Sen, in "Internal Consistency of Choice," *Econometrica* 61 (1993): 495–521, expresses regret at the fact that Nash's and Arrow's conditions are often confused and renames Nash's condition "basic contraction consistency." If a choice function C specifies for any admissible nonempty set S of alternatives a nonempty subset $C(S)$ called the choice set of S, then Sen formulates the condition as

$$(x \in C(S) \text{ and } x \in Q \subset S) \text{ implies } x \in C(Q).$$

For the condition to make sense, the alternatives must be formulated in a manner that makes it possible to eliminate alternatives from a feasible set without altering our relative valuation of the alternatives that remain. Sen describes cases where this property appears to be violated by a rational chooser, but all his examples involve cases where the removal of one of the unchosen options changes the (expected) worth of the remaining alternatives. This means that the removed options are not irrelevant to the remaining alternatives.