

Intransitivity

According to *Transitivity*, if A is better than B, and B is better than C, then A is better than C. We may understand “better than” as short for any of the following: “intrinsically better than,” “all things considered better than,” “hedonically better than,” and “better for a person than.” The same puzzle arises on each interpretation.

Transitivity seems entrenched in our conceptual scheme, if not analytically true; its failure implies, implausibly, that some possibilities cannot be ranked in terms of value; scads of three-member sets inductively confirm it; and hypothetical money-pumpers coax its detractors out of house and home by proffering a series of trades—C for A, B for C and A for B (and then repeating the cycle)—for an endlessly accumulating price. Nevertheless, two highly plausible principles entail that Transitivity is false:

Duration: Pleasure for a given duration is better than slightly more intense pleasure which lasts not more than 1% as long.

Lexicality: No duration of very mild pleasure (at each moment barely preferable to unconsciousness) is better than a century of ecstasy.

Giant leaps in length swamp small losses of intensity; hence, Duration; a century of ecstasy is not worse than the drab eternity of muzak and potatoes—in fact, it’s better (Parfit, 1986; Griffin, 1986); hence, Lexicality. The proof that Transitivity, Duration and Lexicality form an inconsistent triad goes like this. According to Duration, 100 years of ecstasy (=A) is worse than 10,000 years of pleasure slightly less intense (=B); B is worse than 1,000,000 years of pleasure slightly less intense (=C); C is worse than 100,000,000 years of pleasure slightly less intense (=D); and so on to Z, which is 10^{52} years of very mild pleasure. Given these premises, Transitivity entails that A is worse than Z. However, Lexicality entails that A is not worse than Z. So, some member of the triad is false. But which one? This is the puzzle.

Consider Duration. Might a 100-fold increase in a pleasure’s length fail to outweigh a mild loss in intensity? According to a romantic view, B is not better than A; the greatest pleasures need not, and perhaps should not, be abandoned for much more of what is almost as good. But even if this is true, wouldn’t B-Z still violate Transitivity, even though B is not the greatest pleasure? Anyway, wouldn’t one be foolish not to favor the 10,000 years of

near-ecstasy over the century of ecstasy, where these are the only options? Many people feel that Duration might be false where the pleasures involved are very mild. So, perhaps X (10^{48} years of mild pleasure) is not better than W (10^{46} years of slightly more intense pleasure). But even if this is true, wouldn't A-W (or even B-T) still violate Transitivity? Anyway, X should be better than W, given that the pleasures in X are better than unconsciousness at every moment and last 100 times as long as W. Duration, I think, cannot be plausibly denied.

Next, consider Lexicality. Perhaps, in judging A to be better than Z, we don't grasp how much goodness accumulates over 10^{52} years; maybe A—a century of ecstasy—is actually worse than Z— 10^{52} years of very mild pleasure. Is this true? Normally one compares periods of pleasure by deciding which one would prefer, or by interviewing others who are competent to judge. In this instance, competence requires knowing very mild as well as very intense pleasures. I have asked about 50 people (ignorant of this trilemma) to compare A and Z, and *none* have preferred Z—none have denied Lexicality. About 85% prefer A, while the rest are unsure. While some of my interlocutors might not have tasted ecstasy, this would make them *less* likely to prefer A to Z.

For all this, Lexicality might still be false. This would be an extraordinary result. If Lexicality is false, then a long enough duration of mild pleasure is better than any given duration of intense pleasure. And if so, then what Derek Parfit calls “The Repugnant Conclusion” would likely be true: a population of ten billion people, all with a very high quality of life, would be *worse* than a sufficiently large population in which each person's life is barely worth living.

Now consider Transitivity. Must it be retained? Conceptual arguments for Transitivity ring hollow; a conceptual scheme that dismisses Duration and Lexicality *a priori* might itself need dismissing. Transitivity, like any other well-supported inductive generalization, is vulnerable to disconfirmation (on my view). Moreover, savvy rejectors of Transitivity can't be money-pumped. Suppose you think that A is better than B, B is better than C, and C is better than A. You have A. Would you swap it for C, then swap C for B, then swap B for A, pitching in a dime at each turn? Obviously not—you don't want to pay for what you already have. Instead, you'll deny that it's always wise to trade for something better, and so you won't be forced around the circle.

Rejecting Transitivity would have few immediate practical implications; we could still assume Transitivity in everyday decision-making, since exceptions to it are rare. But

such a denial would have broad theoretical implications. Many philosophers think of value quantitatively or linearly; they conceive value as something like a line-segment on which points representing better states of affairs are to the right of points representing worse states of affairs. If “better than” or “more value than” isn’t transitive, then this picture fails, for “to the right of” is transitive. Transitivity seems analytically true insofar as this picture seems built into the concept of value. Need we think of value in this way? Thomas Kuhn says that “. . . all revolutions involve . . . the abandonment of generalizations the force of which had previously been in some part that of tautologies.” If so, perhaps we should abandon Transitivity despite its tautological force. Many philosophers, however, are understandably reluctant to do so.

Bibliography

Griffin, James. *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Oxford University Press, 1986. The citation in the text is to p. 86.

Kuhn, Thomas. *The Structure of Scientific Revolutions, second edition*. Chicago: The University of Chicago Press, 1970. The quotation in the text comes from pp. 183-184.

Lemos, Noah. *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge University Press, 1994. See chapter 4 for a nice discussion of lexical differences.

Norcross, Alastair. “Comparing Harms: Headaches and Human Lives.” *Philosophy and Public Affairs* 26 (1997). Norcross discusses Temkin’s (1996) version of the puzzle and rejects what I call Lexicality (see pp. 144-158).

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984. Parfit defines the Repugnant Conclusion on p. 388.

—“Overpopulation and the Quality of Life,” in *Applied Ethics*, ed. Peter Singer. Oxford: Oxford University Press, 1986. Here Parfit reluctantly endorses a variant of the view I called “romantic.” (pp. 161-164) He discusses the Century of Ecstasy vs. the Drab Eternity on pp. 160-161.

Rachels, Stuart. "A Theory of Beneficence." Oxford University undergraduate thesis (unpublished), 1993. Rachels rejects Transitivity in Part B, where the puzzle is first presented.

—"Counterexamples to the Transitivity of *Better Than*," *Australasian Journal of Philosophy* (March, 1998). Further develops the puzzle.

Schwartz, Thomas. "Rationality and the Myth of the Maximum." *Nous* 6 (1972).

Temkin, Larry S. "Intransitivity and the Mere Addition Paradox." *Philosophy and Public Affairs* 16 (1987). A brilliant and complicated exploration of possible counterexamples to Transitivity (though not the one presented here).

—"A Continuum Argument for Intransitivity," *Philosophy and Public Affairs* 25 (1996). Temkin's version of the puzzle presented here. He rejects Transitivity.

Tversky, Amos. "Intransitivity of Preferences," *Psychological Review*, Vol. 76 No. 1 (1969).

Stuart Rachels