Rational Intransitive Preferences

Peter Baumann

Abstract

According to a widely held view, rationality demands that the preferences of a person be transitive. The transitivity assumption is an axiom in standard theories of rational choice. It is also *prima facie* very plausible. I argue here that transitivity is not a necessary condition of rationality; it is a constraint only in some cases. The argument presented here is based on the non-linearity of differential utility functions. This paper has four parts. First, I present an argument against the transitivity assumption (I, II). Second, I discuss some objections; this will also bring out some features of the view defended here, like the essentially comparative nature of choice (III). Then, I discuss certain proposals that promise to avoid my conclusion (IV). Finally, I go into some implications concerning the nature of rational decision making (V). To accept some intransitivities as rational does not make our life easier; it can even lead to hard choices. However, it does not lead to skepticism about rational choice.

Suppose Mary now prefers a banana to an apple and an apple to a peach. Suppose further Mary is rational. Does it follow that she prefers a banana to a peach? How could she, under these circumstances, prefer a peach to a banana without being irrational? One of the most important standards for the rationality of values and preferences is the axiom of transitivity[1]:

---

1   - with "x", "y", "z" ranging over options (see also below). "xPy" is short for "the person prefers x to y" (or similar expressions).

$$(T) \qquad \forall x \, \forall y \, \forall z \, ((xPy \, \& \, yPz) \rightarrow xPz)$$

This principle is of crucial importance for the theory of rationality. It is, of course, not an empirical assumption about the structure of people's preferences but a normative principle about what that structure should look like. Fulfilling the transitivity condition might not be sufficient for rationality but it seems to be necessary.

Many people – psychologists, philosophers, social scientists and especially economists - think that this is an obvious and fundamental truth. In von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (T) is among the axioms of the theory[2] and the authors take it to be "a plausible and generally accepted property".[3] Many other authors agree.[4] (T) might seem so plausible as to appear trivial and beyond reasonable doubt. However, I will argue that (T) leads to serious problems. Not only can one doubt it, one should do so. My conclusion will be that transitivity is not a necessary condition for the rationality of preferences. This has important implications for our views on rationality. This paper has four parts. First, I will argue for the rejection of the transitivity assumption (I, II); the ideas of non-linearity as well as of utility differences rather than utilities play an important role here. Second, I will discuss some objections and thereby also explain the view proposed here more (III), for instance its essentially comparativist nature. Then, I will talk about some proposals to avoid my conclusion (IV). Finally, I will say a bit about implications (V).

---

2   See von Neumann/ Morgenstern 1953, 26-27.
3   von Neumann/ Morgenstern 1953, 27.

I. The Main Argument against (T): Preference

(a) Preliminaries

First, some short remarks on the concept of preference.[5] I am dealing here with the relation of "strict preference" ("P"); "preference" and related terms shall refer to strict preference here. To prefer (in this sense) something to something else implies that one is not indifferent ("I") between the options. If a person is indifferent between two options, then neither option is strictly preferred to the other. A "weak preference" ("W") between two options x and y allows for both strict preference of x over y and indifference between the two. One could, of course, also use the notion of weak preference as the basic notion here: A person is indifferent between two options if and only if both options are weakly preferred to the other one; a first option is strictly preferred to a second one if and only if the second one is not weakly preferred to the first one but the first one is weakly preferred to the second one.

It does not matter here whether one takes options, objects, propositions, states of affairs, or sentences being true, etc., as the relata of the relation. I will go with options here.[6] We can then say that preference is a binary relation on a set of options.

---

4   See Ramsey 1990, 78, 75; Luce/ Raiffa 1957, 23, 25, 28; Savage 1972, 18; von Wright 1963, 21ff.; Davidson/ McKinsey/ Suppes 1955, 142, 145-146, 154-155; Jeffrey 1974, 391; Nozick 1993, 140; Putnam 1986, 3ff.; Maher 1993, ch.2.

5   See, e.g., Suppes 1958, 63, 169, 224, 213ff.

6   In the following, I will restrict the domain of quantification to the members of the set of options.

One might have second thoughts about whether the preference relation also fulfills the following completeness or connectedness condition:

(CO) $\forall x \, \forall y \, (xPy \lor yPx \lor xIy)$

(or, in other words: $\forall x \, \forall y \, (xWy \lor yWx)$. Referring to fundamentally incomparable (or "incommensurable") options, some authors argue that (CO) is not among the principles of decision theory and is no rationality constraint for preferences.[7] I cannot go into this here and will not assume that there is a problem with the completeness assumption. Let us instead focus upon the transitivity condition (T).

(b) Non-Linearity and Differential Utilities

Consider the following example. Fred is doing his work and is getting hungry. He decides to buy a slice of pizza for $2 at a take-away. Before he arrives there, he remembers that for only $2 more he can get a whole plate of spaghetti at an inexpensive Italian restaurant. Since $2 do not really count for Fred and since he is much better off with a whole plate of spaghetti he changes his mind and starts walking to the restaurant. Then he has a new idea: for only $6 he can get a saltimbocca at a trattoria nearby. Since $2 still do not count very much for him and since he likes saltimboccas much more than spaghetti he decides to go to the trattoria. But then he remembers his first plan: he just wanted to go out and get a slice

---

7   See, e.g., Putnam 1986, 3ff. but also Luce/ Raiffa 1957, 23, 25. For more detailed discussion of incomparability see: Raz 1986, ch.13; Chang 1997; Chang 2002. See also Baumann 2018.

of pizza for $2. For Fred, however, a price difference of $4 makes a remarkable difference whereas $2 does not; since he is not extremely fond of saltimboccas he decides to get back to his original plan (never mind the somewhat unrealistic nature of the prices!).

What has happened here? Let:

"c" stand for "getting pizza for $2",

"b" stand for "getting spaghetti for $4" and

"a" stand for "getting saltimbocca for $6".

Assume Fred's preferences have not simply changed during his deliberations (we might have independent evidence for this assumption). Then, Fred's preference ranking is the following one:

(P$_1$)   bPc & aPb & cPa.

In other words, Fred's preferences violate (T).[8] There is no best option for Fred; whichever option he considers, there is always one he likes better.[9]

---

8   I am focusing here on cycles of the form aPbPcPa. This is, of course not the only way of violating (T): aPbPcIa would be another form (or incomparability of a and c still another one, if one wants to assume this possibility). For the sake of simplicity I focus on cycles here and use the term "intransitivity" to refer to this case. See also fn.19 below for terminology.

9   Here is another example. I intend to buy a cheap radio that I could use in the bathroom; there is one for 9.99. I am ready to buy it when the salesperson in the store tells me that for a bit more money (for 14.99) I can have a radio with a tape deck. Since the price difference is negligible for me and I like the addition, I prefer the second option. The sales-person then tells me that for 19.99 I can have a radio with both tape deck and CD player. Again, the price difference is negligible and I prefer the last option to the second one. It is easy to imagine how this could be continued until I am ready to get a whole stereo and tv system for 199.99. However, then I might think of the first option, again, and

Does this mean Fred is irrational? Is he bound to make sub-optimal choices? What advice should we give Fred? Or is there something wrong with (T)? Let us suppose, for a change, that a, b and c do not differ in price or in any other relevant aspect except Fred's gusto. We can call preferences between options that differ only in one relevant aspect "one-criterion preferences". If you prefer a banana to an apple and an apple to a peach and nothing but the taste of the fruit counts for you and if you are rational then you prefer a banana to a peach. In such cases violating (T) would be irrational; there seems to be no reason to suspect otherwise (see also below).

However, our original lunch-example was more complex. There are two independent relevant aspects for the decision: price and quality. That two criteria or dimensions are independent roughly means that the utility of a good can vary in one dimension without varying in the other dimension. The quality of food can vary with constant prize and the prize can vary given constant quality. Fred's preferences are "two-criteria preferences".

Moreover, a difference of $2 does not really count for Fred whereas a difference of $4 does. The money-criterion is "non-linear" in the sense that doubling the expense of money does not mean "doubling" the negative subjective utility. However, talk about "doubling" utility can be very misleading because utility is not additive in this sense. Since the notion of non-linearity is crucial here, we need to take a closer look at it.

One might want to propose something like the following. We can use a curve to illustrate the utility function for some good G (that is, the function that maps quantities of Gs into numbers; the differences between those numbers then represent differences

---

compare it with the last option. The huge price difference between 9.99 and 199.99 outweighs everything else and I would rather accept the first than the last offer.

between utilities). However, it turns out that it is better here to imagine a slightly more complex kind of utility function: one that maps differences between amounts of Gs into differences of utility (but see below for an important caveat here). Our example above concerning the money-criterion would offer an illustration: Monetary differences map into utility differences.[10] Now, in case the first derivative of the differential utility function is not constant, that is, in case the gradient or slope of the curve varies, the (differential) utility function – or the corresponding criterion – is non-linear. One can, for instance, imagine a function that is quadratic in a certain interval of x-values (values of differences in amounts of Gs); the y-values (values of utility differences) give the curve its characteristic upward slope.

One might be tempted to object to this proposal that that it involves the possibility of a numerical representation of utilities which in turn presupposes the principle (T) I am criticizing here. If the preference relation can be represented by the greater-than-relation for numbers, then the former inherits the transitivity of the latter. However, the argument presented here does work by assigning numbers not to utilities but rather to utility differences. Hence, the objection has no force.[11] - The hints given here should, however, be

---

10 Since most goods show a decrease of marginal utility we would, of course, have to be aware of the relativity of this function to the amount of goods already available to the person. - One must also clearly distinguish between non-linearity and decreasing marginal utility. Non-linearity is explained here via a function which maps differences onto differences while the idea of decreasing marginal utility uses functions which map absolute quantities of goods to utilities, not to utility differences. See below for problems with this standard way of constructing utility functions. – Non-linearity can also not be reduced to risk aversion: The phenomena discussed here can appear even for persons who show no particular attitude towards risk.

11 Here is another attempt to explain non-linearity. If a criterion is non-linear, then the person would not be indifferent between any p-chance on n units of G and a p/x-chance on xn units of G. For instance, I might be desperately in need of £15 right now in order to buy a train ticket to London. Under such circumstances, I would prefer a 90%-chance on £15 to a 45%-chance on £30. I am then certainly also not indifferent between, say, one gamble which gives me a 90%-chance on £30 and a 10% change on £10 and a gamble which gives me a 90%-chance on £15 and a 10% chance on £16. I, like many in

sufficient for some understanding of what I mean by "non-linearity". It seems quite plausible to assume that the phenomenon exists, and that assuming this neither presupposes the transitivity principle nor its denial.[12]

The kind of non-linearity relevant here is not due to a mere threshold effect. It is, for instance, not due to "perception" thresholds such that very small "objective" utility differences do not correspond to differences in subjective utility. Rather, it is presupposed here that the person notices the relevant features of the situation. However, there are, of course, threshold effects that result from humans' limited sensitivity to small differences between two options along one dimension. One can consider this to be an extreme case of the non-linearity above: one where the smallest utility difference equals 0 (more on this below).

Here is an illustration of how non-linearity in one of two criteria can lead to intransitivity. Suppose we are comparing some options x and y with respect to criteria A and B. Let "$\Delta_A(x,y)$" refer to the utility advantage of y over x in dimension A, and let "$\Delta_B(x,y)$" refer to the utility advantage of y over x in dimension B. In our case above, we can put taste (T) for A and money (M) for B. $\Delta_T(x,y)$ then expresses how much utility difference there is between the taste of x and of y; if y has better (worse, equal) taste than x, then $\Delta_T(x,y)$ is positive (negative, equals 0). Similarly, $\Delta_M(x,y)$ expresses how much

such a situation, would prefer the second gamble. One problem with this explanation of non-linearity is that this kind of lack of indifference could also be due to (rational) risk averseness rather than to non-linearity of the utility function. Or it could be due to decreasing marginal utility of the relevant good. How could we possibly tell? – Along similar lines one could imagine a social case. Suppose a person is asked to minimize total disutility for a group of two persons. She can either take 2 Gs (Dollars) from each person or 4 Gs (Dollars) from one person. If the person is not indifferent between these two options, then this might, again, be due to other factors (e.g., considerations of distributive justice) than to non-linearity with respect to the utility of Gs (money) for our person.

12   See also Wodak 2019 here.

utility difference there is between paying for x and paying for y; if x is cheaper (more expensive, equally expensive) than y, then $\Delta_M(x,y)$ is negative (positive, equals 0). $\Delta_T(x,y)$ and $\Delta_M(x,y)$ are skew-symmetric functions.[13] We can then say that y is preferred to x just in case:

$$\Delta_T(x,y) + \Delta_M(x,y) > 0.[14]$$

Suppose we have for our food example above the following numbers:

$\Delta_T(c,b)=.3$, $\Delta_M(c,b)=-.001$ and thus $\Delta_T(c,b)+\Delta_M(c,b)=.299>0$,

$\Delta_T(b,a)=.3$, $\Delta_M(b,a)=-.001$ and thus $\Delta_T(b,a)+\Delta_M(b,a)=.299>0$,

$\Delta_T(a,c)=-.6$, $\Delta_M(a,c)=+.799$ and thus $\Delta_T(a,c)+\Delta_M(a,c)=.199>0$.

Hence,

(P$_1$)   bPc & aPb & cPa.[15]

---

13   "A function is called skew-symmetric if the value of the function is multiplied by -1 when any two variables are interchanged. An example of such a function is H(x,y) = z-y." (Helson 1994, 98) We are not dealing with cases here where the utility differences between x and y are zero. Hence, we can also say that our Δ-function is skew-symmetric if $\Delta(x,y) > 0$ just in case $\Delta(y,x) < 0$.

14   And: x is preferred to y just in case: $\Delta_T(x,y) + \Delta_M(x,y) < 0$. Finally: There is indifference between x and y just in case: $\Delta_T(x,y) + \Delta_M(x,y) = 0$. - This is just one possible (skew-symmetric) way of aggregating the two dimensions; there are others but here it doesn't matter which one is the best or correct one (if there is one) since it should only be illustrated here how rational intransitivity is possible.

15   Since non-linear, multi-dimensional preferences do not always result in intransitivity, one can wonder under what general conditions they do. I think that in this generality nothing more informative can be said than the following. Intransitivity results just in case the following condition is met: For each ordered option pair – (c,b), (b,a), (a,c) - the sum of the Δ-values is positive. Thanks to a referee here.

It should be mentioned as an aside here that one can have intransitivities even if one does not start with differential utility functions but rather attributes utilities to options on the different dimensions.[16] Suppose that the gustatory utilities for a, b, c above are 2.5, 2, and 1 while the utilities of the money spent for each option a, b, c are 2.25, 1.75 and 1 respectively (if you prefer values between 0 and 1 you can, for instance, just divide by 10). Suppose further that we are only considering Δ_T and Δ_M for the ordered pairs (c,b), (b,a) and (c,a). Finally, let us assume that y is preferred to x just in case $\Delta_T(x,y) - (\Delta_M(x,y))^2 > 0$, and that x is preferred to y just in case $\Delta_T(x,y) - (\Delta_M(x,y))^2 < 0$. As one can easily see by computing the numbers, the same intransitivity results. However, for reasons explained in more detail below, I prefer the approach via functions of utility differences above (the particular choice of ordered pairs above also seems somewhat arbitrary). - Finally, Peter Fishburn uses a skew-symmetric bilinear function (which is linear in each of the variables separately but not in both together) to illustrate the possibility of rational intransitivities for the case of prospects or gambles.[17]

(c) Implications

Given all the above explanations it seems obvious that Fred is not irrational.[18] There is no way for him to be more rational. To say that he should not have these preferences because they are not transitive would beg the question in favor of (T). Life can be more complex

---

16   Helene Shapiro proposed this approach to me.

than textbooks might suggest. As long as no specific reason is given for charging Fred with irrationality (see also below), he should enjoy the presumption of innocence or rationality. It is not "his fault" that his preference ranking has the form (P$_1$) or that this is a case of multi-criteria and non-linear preferences; rather, the form of his preference ranking is due to the givens of the choice situation. If Fred cannot possibly do better, then it is not true that he ought to do better; here, too, *ought* implies *can* (and *cannot* implies the absence of the corresponding *ought*). Fred violates (T) without being irrational. And there is no reason not to apply the concept of preference in Fred's case. Hence, (T) cannot be a universal rationality constraint for preferences:

(N) $\neg[\forall x\ \forall y\ \forall z\ ((xPy\ \&\ yPz) \rightarrow xPz)]$.[19]

Given (N) we can reject (T) but keep a restricted version:

(T') Transitivity is a rationality constraint for

(a) one-criterion preferences and

(b) linear multi-criteria preferences.

One might ask whether we even have reason to accept (T'). Why should rationality require transitivity even in cases (T'a) or (T'b)? Since we can, in some cases, easily give up the

---

17   See Fishburn 1984, esp. 133.
18   See also Davidson 1985, 350-351 who draws a very different conclusion from a similar example.
19   For simplicity's sake, I stick with the general term "intransitivity" here when I have (N) in mind, leaving other uses of the term aside (as most authors seem to do).

transitivity condition one might wonder why we shouldn't "go all the way", take further steps and doubt or deny that transitivity has anything to do with rationality in the first place. Consider Kurt: He prefers a banana to an apple and an apple to a peach but also prefers a peach to a banana. And nothing but the taste counts for him. Should we say he is irrational? At least, he is a very strange guy. It is hard to understand how he can, given the above conditions, prefer a peach to a banana. Notice the contrast with the case of Fred: Here we can explain how the intransitivity arises; it results from the combination of two unproblematic factors and aspects of human choice behavior, namely non-linearity and multi-dimensionality. For this reason, the charge of irrationality does not apply to Fred even though it applies to Kurt. It is also hard to understand how we could have reason to doubt that the concept of preference is applicable at all in Kurt's case. Transitivity seems to be built into the concept of a one-criterion preference. Hence, if we agree that the concept of a preference is applicable in Kurt's case, then the best explanation for his preferring a peach to a banana is some kind of irrationality on his side.

Given all the above considerations it is plausible to propose the following general thesis:

(T*) If

    (a)   the person's preferences are two-or-more-criteria preferences and if

    (b)   at least one of these criteria is non-linear to a certain degree

           then

    (c)   the person's preference ranking may (but need not) violate (T) without the person being irrational.[20]

(T*a,b) formulate sufficient conditions for the possibility of rational intransitivity of preferences; it is not my aim here to present necessary and sufficient conditions for rational intransitivity. By the way, when I say that there is rational intransitivity I am using the word "rational" in the rather weak sense of "not irrational" but not in a stronger sense which would imply that rationality demands intransitivity in such cases (the latter is left open here).

It seems very plausible to assume that most of our preferences are more-criteria preferences and most of the criteria are non-linear. This becomes even more plausible as we move from oversimplified textbook examples to the typically much more complex examples from real life. Especially the truly important choices in life tend to involve many criteria and non-linearity. If all this is true (and I think it is), then it is also very plausible to assume that in many important cases (T) is no rationality constraint for preferences. Many rational preference rankings are intransitive or "cyclical". Only in the two special cases of one-criterion preferences and of linear multi-criteria preferences does rationality imply transitivity.[21]

---

20   See also Tversky 1969, 31ff., 43, and 47-48 for a proof; Bar Hillel/ Margalit 1988, 124; Philips 1989, 478. See also Broome 1993, 53ff. but also Broome 1991b, 80ff.

21   There are philosophers, psychologists, decision theorists and economists who agree that transitivity is not necessary for the rationality of preferences: see Michalos 1967; Tversky 1969, 31ff.; Schumm 1987, 435ff.; Bar-Hillel/ Margalit 1988, 120ff.; Fishburn 1991, 113ff.; Anand 1993, 55ff.; Gendin 1996, 482ff.; Horowitz 2006. See also the discussion between Hughes 1980; Lee 1984; Philips 1989 and Rawling 1990. Temkin 1996, 180ff. offers a very different argument against (T) and would not accept what has been said here (see 202ff.); but also see Temkin 2012, esp. chs. 2, 5 and 452-455. For a similar argument see Rachels 1998 who somewhat surprisingly holds that intransitivity can only be

Recently, some philosophers have presented spectrum arguments against the transitivity assumption.[22] One variant is this: There is a series of n states such that each state i is a tiny bit worse in one dimension (slightly more intense pain) but clearly better in another dimension (less duration of pain) than state i+1 so that the i+1th state is worse than the ith state. However, it is also very plausible that the nth state (hangnail for a very long time) is better than the 1st state (severe torture for 3 days). Intransitivity results. There are some analogies but also some differences between such spectrum arguments and the argument presented above. The latter already works with only 3 options. To be sure, as the example above (fn.9) suggests, the argument presented here also works for more than 3 options; however, this doesn't mean that spectrum arguments could also work with 3 options. While this might not amount to a deep difference, there are more important ones. First, spectrum arguments depend on difficult and controversial assumptions about how the two dimensions relate to each other[23]: How much intensity of pain can be traded for what duration? There are good reasons to doubt that the first is comparable with the second in any degree of precision that would be required here; and this is so even if we should be able to construct scales for the badness of intensity of pain as well as for the badness of duration of pain. The argument above does not depend on any such controversial

---

the exception (see 83). Doody 2019 argues that incomparability leads to violations of transitivity. Mandler 2001 argues that one has to choose between completeness (for one notion of preference) and transitivity (for another notion of preference), and cannot have both at the same time. For the case of social choice see Arrow 1963. Morrison 1962 is a pioneering study which deals with intransitive comparisons more generally, not just intransitive preferences.

22 This goes back to Parfit 1984, chs. 17-19, and also Rachels 1998, 2004 as well as Temkin 1996, 2000 and especially Temkin 2012. See also, for instance, Carlson 1996, Binmore and Voorhoeve 2003, Carlson 2003, Broome 2004, 55-59, Carlson 2005, Qizilbash 2005, Knapp 2007, Voorhoeve 2008, Benbaji 2009, Huemer 2013, Coons 2014, Handfield 2014, Kagan 2015, Handfield 2016, Klocksiem 2016, Handfield and Rabinowicz 2018, Nebel 2018, Pummer 2018, Herlitz 2020a, 2020b.

23 Temkin 2012 has a lot of discussion about this.

assumptions. Furthermore, Larry Temkin, for instance, thinks that the criteria as well as their significance or relevance might vary with pairs of options. I am not making any such assumptions here. Change of criteria or of their relevance across different choice situations makes any argument against the transitivity assumption vulnerable to triviality charges ("No wonder you don't get transitivity if you change the rules of the game between cases!"). Finally, the argument above is about "preferred", not about "better than"; more controversial assumptions about value relations are avoided here (see below for more on this difference).

An argument similar to the one above has been made some time ago by Amos Tversky.[24] However, Tversky uses the assumption of a lexicographic ordering of the different criteria: It is only when the differences in one of the dimensions are big enough that the relevant criterion kicks in - but then it outweighs the other criteria.[25] No assumption of such orderings of criteria is made here. Some years before Tversky, Howard Raiffa presented cases of intransitive preferences which do not seem to violate standards of rationality.[26] However, he did not capture how big utility differences between options are; his example only represents utility differences on ordinal scales, not on interval scales. This is a restriction of the argument which is also not present here.[27]

---

24   See Tversky 1969.
25   See similarly Horowitz 2006, 108-109. See also Temkin 2012, 219-221 and chs. 11 and 12 more generally. Mackie 1985 discusses intransitivity resulting from certain combinations of different "principles of preference". However, he rather has ordering of options in mind here than subjective preferences of persons. Apart from that, he is simply assuming that intransitivity is inconsistent.
26   See Raiffa 1968, 77-78.
27   Apart from all that, one can bring in probabilities in at least two different ways. First, one could discuss stochastic principles of transitivity: Instead of talking about outright preference, one could put everything in terms of the probability that the person would prefer or choose one option over another one. This, more behavior-related, way of discussing things would make the argument presented here slightly more complicated but the main points would remain the same. Therefore, I disregard it here. Second, one could discuss everything in terms of expected utilities rather than utilities *simpliciter*.

II. A Related Argument against (T): Indifference

Very close to the argument presented above is another argument against (T): one which exploits the intransitivity of indifference.[28] Consider this example. Typically, persons are indifferent between any two cups of coffee which only differ (relevantly) insofar as one cup contains one more grain of sugar than the other cup. One grain of sugar more or less in the coffee, we may assume, lies below the thresholds of our discriminatory abilities. The person is thus indifferent between any two such cups of coffee: between a cup with 0 grains and a cup with 1 grain, between a cup with 1 grain and a cup with 2 grains, between a cup with 2 grains and a cup with 3 grains, …. and between a cup with 9999 grains and a cup with 10000 grains. However, the person is not indifferent between a cup of coffee with no sugar in it and a cup with 10000 grains of sugar in it; the latter is much too sweet for her and she clearly (strictly) prefers the one without any sugar. This is an illustration of the intransitivity of indifference. [29]

It is not hard to see how this leads to intransitivity for two criteria when at least one of them is not linear. Suppose that our person likes coffee with some sugar in it. She is offered a series of choices between pairs of cups of coffee such that one cup contains one grain of sugar more than the other and she is paid 1 extra cent for each extra grain of sugar;

Again, this would just make things a bit more complicated but not touch on the main points here. Therefore, I disregard this option here.

28   See, e.g., Armstrong 1939, Lehrer and Wagner 1985, or Gauthier 1986, 40-41 See also Raiffa 1968, 79-80.

29   So, suppose aIb, bIc but cPa. Doesn't this show that b should be placed in intermediate position in the preference order between a and c? But then we would have to deny that aIb and that bIc – which

the series starts with a cup with 1 grain of sugar and a payment of 1 cent. Given the discriminatory threshold described above and given typical attitudes with respect to money the person will for any number n clearly prefer the cup with n+1 grains of sugar (which will earn her n+1 cents) to the cup with n grains of sugar (which will earn her n cents). However, we can easily imagine that the person finds the idea of drinking a cup of coffee with 10000 grains of sugar in it so disgusting that even $100 cannot persuade her; she prefers the coffee with just 1 piece of sugar in it (and earnings of just 1 cent). In other words, her preferences are intransitive.[30]

Again, there are at least two independent criteria of evaluation; in addition, one of them shows an intransitivity of indifference based on discriminatory thresholds. All this results in the possibility of intransitive preferences. Again, it is hard to see why this should count as a case of irrationality. We could add this further case to the conditions in (T*) above but since this preference intransitivity is based on an indifference intransitivity one can consider it as a special, extreme, case of the one above: one where the utility difference between two adjacent options equals 0 in the relevant dimension. Interestingly, the intransitivity of indifference is much more accepted in the literature than the intransitivity of preferences – even though the above remarks suggest that if the first is not problematic, then the latter also isn't.[31]

---

contradicts our assumptions and also seems incorrect. Apart from that, this would presuppose a transitivity principle and thus beg the question (see also Maher 1993, 57-60).

30   See also Quinn 1990 here, and Voorhoeve / Binmore 2006 who follow up on Quinn but disagree with the argument presented above.

31   See Schumm 1987, Hansson 1993, 481-482, Aldred 2007, 398-400.

III. Objections

Two of the 8 objections I am going to discuss concern the specific argument against (T) presented here (sections (a) on sorites and (b) on essentially comparative choices). The other 6 objections concern the denial of (T) more generally. I start with what is by far the most common objection against (T): the money pump argument (c). This is followed by a concern about limited measurement of preferences (d), - a topic at the center of classical rational choice theories. This is followed by two related concerns of a more axiological nature (e, f). Finally, I comment briefly on implications for dominance principles (g) as well as on the idea that change of preference rather than intransitivity explains the phenomena under discussion here (h). In addition to all that, one could argue that the content of preferences needs to be described in a more fine-grained way than presented here, and in a way that removes the appearance of intransitivity. I will discuss this in the next section (IV).

(a) Sorites?

Does the argument for rational intransitivity presented above, especially for the case of indifference, constitute a bad sorites argument? I don't think so. There is no vague predicate involved here nor a (vague) set of borderline cases which separates the clear cases of the applications and the non-applications of the predicate. Even if "as good as" should be vague, the argument does not exploit any such vagueness and would work in the

same way for a perfectly sharp notion of being "as good as". What would create a paradox of transitivity here is not the iteration of some substantial premise (e.g., *if n is F, then n+1 is F*) but rather the assumption of transitivity of indifference; the paradox can thus be avoided by giving up transitivity. A substantial sorites-producing premise would have the form of "If coffee-option n is good, then coffee-option n+1 (with 1 more piece of sugar in it) is good" (in the case of indifference) or of "If food option n is good, then food option n+1 (with 2 more Dollars spent) is good" (in the case of preference). In contrast, the argument presented here works with "better", not with "good" (and "A is better than B" does not imply that A or B are good). Furthermore, the argument for intransitive indifference can easily be shortened to one involving only 3 options and two steps; sorites typically, at least, take many more steps. Finally, in contrast to a soritical piece of reasoning, the argument for intransitivity does not end up in a clearly false conclusion (even if the denial of transitivity is false it is not clearly false).[32]

(b) Basically Non-Comparative Choices?

How about the following objection? It is wrong, one might say, to do what I did: to evaluate single options only in relation to other options. Rather, one should evaluate options one by one and assign cardinal values to them according to their "absolute" utility or desirability for the person. More precisely, one should not compare the options in one dimension of evaluation after the other, determine the differences between the options in

---

32   See also Aldred 2007, Temkin 1996, 197-201 or Temkin 2012, 277-296, 531-534, especially 278-284 and also Nebel 2018, sec.4.2.

each dimension and then aggregate the different dimensions; rather, one should evaluate one option across the different dimensions first, then aggregate everything and compare it with the results for the other options. In other words, we should go for separate rather than joint evaluation of the options[33], for an "additive model" rather than an "additive difference model".[34] If we do this for all the options we can construct a ranking according to the cardinal numerical values we assigned to the options. Since the relation "bigger than" is transitive for cardinal numbers the ranking of the options must be transitive, too. In other words, what I said would be wrong.

This objection fails because we cannot assign cardinal values or cardinal utility to single options. There simply is no way to measure absolute utility because there is no absolute utility. The desirability of an option can only be evaluated in comparison to other options. Talk about the "strength of a desire for an option" only makes sense if interpreted as a statement about the relative desirability of one option versus other options. That I desire that p implies that I prefer that p to non-p. In this sense, utility is "relative". The two-place concept of preference is prior to the one-place concept of a desire. If the corresponding expression "better than" should turn out to be very closely related to "preferred" (see below), then "better than" is more fundamental than the expression "good (to this or that degree)" and "good" has to be explained in terms of "better". This point is of fundamental importance for the argument presented here. In some cases we cannot even assign ordinal values to options. In these cases transitivity is no rationality constraint for

---

33   See Hsee et al 1999 who give an overview over empirical research on preference reversals between separate and joint evaluations of options. What they say does, however, not imply that there is a substantial conceptual difference between two kinds of evaluation both available to us.
34   See Morrison 1962, 14-15, 55-56, and Tversky 1969, 41.

preferences.[35] A further argument concerns the intransitivity of indifference (see above). If one cannot assign numbers to options expressing their "absolute" utility, then the comparativist approach to preference and utility appears to be the best theoretical option here.[36]

If we cannot look at one option in isolation – can we perhaps look at three at a time? If we could look at three options x, y, z at the same time and determine which one is the best, then intransitivity could be avoided. However, one cannot just assume that this is possible (otherwise one would beg the question against the argument proposed here). How could one determine, when faced with three options, which one is the best? It is hard if not impossible to see how that could work. We really have to look at pairs of options. Preference is essentially a binary relation, not a ternary or n-ary (for n>2) relation. This also explains why a traditional form of critique of intransitive preferences does not work. Gordon Tullock, for instance, argued that intransitivity leads to a form of logical inconsistency: "If the individual is alleged to prefer A to B, B to C, and C to A, we can inquire which he would prefer from the collection of A, B, and C. Ex-hypothesi he must prefer one, say he prefers A to B or C. This, however, contradicts the statement that he prefers C to A, and hence the alleged intransitivity must be false."[37] If it is true that a person has a ternary preference, then she cannot have intransitive preferences because to say so would entail a contradiction. The answer to that is that we just do not and cannot

---

35   See also Temkin 2012, esp. chs. 11 and 12 here, as well as Handfield 2016.
36   The argument against the transitivity principle presented here supports this view of choice as essentially comparative (and the latter also supports the former). It would go beyond the limits of this paper to discuss this in more detail here. Some (see, e.g., Barrett 2019) prefer to talk about the monadic relation of desire and its strength rather than about the binary relation of preference. The problem, however, is then to show how talk about strengths of desires do not in the end reduce to talk about preferences.

have ternary preferences which are not based on binary preferences. Apart from that, Tullock is implicitly assuming the possibility of transitivity.

Similar things would have to be said about a related idea: What if one presents a person with more complex alternatives (compound lotteries)? Couldn't the person answer the question whether

$$x \, P \, (y \lor z)?$$

If she could, then transitivity would hold. But again, all this presupposes that we do not have to go back to pairwise evaluations of options. It is hard to see how we could possibly do without that. To recapitulate all this: One option is not enough and three options are too many.

(c) Money Pumps?

There is a very common objection to my conclusion: the "Money Pump"-argument.[38] The objection runs as follows. Suppose Fred has a preference ranking of the form (P_1). He agrees to give one penny to Mary if she offers him an option that seems to him to be clearly better than the one he already has. Fred does have this general preference as well as another preference of owning more money over owning less money. Suppose Fred starts with the option c. Then, he is ready to pay one penny to Mary if she offers him the

---

37   Tullock 1964, 403.

preferred option b. But then again, he will pay one penny to get the preferred option a. Since he prefers c to a he will pay a third penny to get option c, and so forth. Fred will run around in his circle of intransitivity until there is no money left to spend on the money pump. The preferences that get him into the money pump are inconsistent with his preference to own more money rather than less money.[39] Many people think that the possibility of a money pump proves that intransitive preferences are irrational.

But this objection is not convincing. First, a rational person would simply refuse to make the money pump-deal as soon as she realizes (as she would) what is going on. Why should the intransitive preferences be the culprit here rather than the pumping-agreement? If the pumping cannot be predicted, then we can say that bad things sometimes happen even to agents who are not irrational or blameworthy in some relevant way.[40] Second, and more importantly, the money pump-argument presupposes that one can, in principle, always avoid intransitive preferences. But this is not true if what was said above is true. The money pump-argument thus seems to beg the question under discussion here insofar as it presupposes the truth of (T). Third, intransitivity of indifference is much more accepted than intransitivity of preference; however, money pump arguments can as easily be made for the former than for the latter; so, if that argument creates problems for the intransitivity of preferences, then it should equally create problems for the intransitivity of

---

38   See Ramsey 1990, 78-79 and Raiffa 1968, 78. The argument is also known as a "dutch book-argument".

39   One has to be careful here. First, the inconsistency is rather indirect. The general agreement to make this kind of deal leads to a series of deals which results in being left with less rather than more money. So, the result of some preferences (the pump preferences) is incompatible with the fulfillment of another preference, the preference not to lose money. Second, that each individual exchange of an option for a penny is rational does not mean that the series of such exchanges is also rational (see Huemer 2013, 333-334).

40   Thanks to a referee for pressing me on this point.

indifference (which it doesn't seem to do for many).[41] Finally, even if the money pump argument works (is valid or even sound), this would only show that a person with intransitive preferences is irrational in some sense. However, if such intransitivity is acceptable for independent reasons, then this particular irrationality of pumping is blameless (given permissible starting preferences) and inevitable, given the nature of things and the situation at hand and given the preferences of the person. Hence, this would at best show that blameless and inevitable irrationality can happen (given that intransitivity is not prohibited by independent reasons). One can reject this reply only if one already assumes transitivity from the start and thus is begging the question here. Overall then, the money pump-argument is not a good argument and does not succeed in showing that intransitivity entails irrationality.[42]

(d) No Numbers then?

There is a more serious problem which arises most clearly with the pattern aIb, bIc and cPa. If we were to assign real numbers to options, a core aim of classic decision theory, then we would have to assign the same number to a, b, and c insofar as aIb and bIc but we would also have to assign different numbers to a and c insofar as cPa. Given the latter, we would have a reason to assign b the number of a as well as the different number of c. Which one is the correct one? Or should we rather claim that there is a basic relativity of measurement to the choice situation, a lack of context-invariance of the "utility function"?

---

41   See, e.g., Hansson 1993, 478-479 and Aldred 2007, 398-400.

This shows again how much in the classical theory depends on the transitivity assumption, and how much has to change if one gives up the transitivity of indifference (and preference; though this particular problem appears in more drastic form for the case of indifference). This problem also arises for the idea of a utility difference function (see above). This shows that one has to take talk about this kind of function with a grain (or two) of salt: It is a theoretical fiction meant to help make a certain point. The real lesson to be taken from this is that there is no choice- or comparison-invariant assignment of values if one denies transitivity (see above). (All this does, of course, leave intact the possibility of having some clear preference relations).

(e) Bad Better than Good?

Here is another very interesting problem. First, some explanations. It seems that some things are non-instrumentally good for a person, in the light of her preferences (e.g., being very happy), some things are non-instrumentally bad for that person (e.g., being in severe pain), and some things are neither the one nor the other for her, or "neutral" (e.g., having a prime number of hairs on one's head – which is utterly irrelevant to most of us).[43] What can one mean by "good", "bad" and "neutral"? Here is a good way of explaining this. An option is neutral (N) just in case it is as good as an option x the presence of which is as good as its absence (xInot-x). An option is good (G) just in case it is better than a neutral

---

43   In the following I will drop the qualifier "non-instrumentally" for the sake of ease of exposition. For a few paragraphs I will also drop the qualifier "for a person" but I will get back to that below.

42   See also Schick 1986, 116ff.; Anand 1993, 61ff.; Fishburn 1988, 44, Maher 1993, 36-38, Temkin 2012, 185-188, and Baumann Ms. I won't go more into this very widely discussed topic here.

option. An option is bad (B) just in case it is worse than a neutral option. And a good option is better than a bad option. Hence, we have transitivity here: The good is better than the neutral, the neutral is better than the bad, and the good is better than the bad. Or, somewhat schematically:

(GNB) G>N, N>B, G>B.

All this[44] seems extremely plausible and hard to give up. But can there then be intransitive cycles involving 3 options not all of which are good and not all of which are bad? Without having to go into all the possible combinations, we can consider the case of a good option a, a neutral option b, and a bad option c. If it were the case that

(P_1) bPc & aPb & cPa.

then we would have to accept that a bad option (c) is better (preferred in the light of the person's preferences) than a good option (a). However, this seems clearly unacceptable.[45] How could the above proposal of acceptable intransitive prevent this abominable conclusion?

---

44   See Chisholm and Sosa 1966 for this kind of proposal, and especially 245-248 for reasons to distinguish between neutrality and indifference. I will leave details and possibly necessary amendments aside here.

45   See for this problem also Danielsson 1996, Nebel 2018, 877, and, with respect to satisficing: Klocksiem 2016, 1321-1322.

The way out is surprisingly straightforward: We have to restrict possible cycles of intransitivity to sets of options that are either all good or all bad.[46] But isn't this an ad hoc solution, just motivated by the attempt to solve our problem? No, (GNB) is independently plausible and even has the flair of an axiom hardly anyone would want to deny. And (GNB) does not entail that there cannot be intransitive cycles amongst options that are all good or all bad. And this latter assumption has good reasons in its support. So, both (GNB) and the argument against the transitivity assumption presented here are independently plausible. So, there is a good reason to accommodate them both by making a restriction of the sort proposed here (in contrast, it would be very much ad hoc, for instance, to just rule out all non-linear and multi-dimensional preferences). All this also explains why (GNB) does not support a defense of (T) all the way. Thus, this "hybrid" solution is well supported and pretty plausible insofar as it assumes that there are limits to possible intransitivity.[47]

(f) Preferring Worse?

There is a related problem we should consider here. So far, I have used the term "better than" as well as related terms ("worse than", "as good as", "good", "bad", "neutral") in a somewhat "subjective" sense: "better than" roughly, at least, in the sense of "preferred by

---

46   There is no possible rational intransitivity in case all options are neutral (though there might be if they are mutually indifferent).

47   Interestingly, Aristotle seems to disagree with all this in the following very brief passage in the *Topics* (III.2, 117a20-22): "Also it is quite possible for what are not good things to be more desirable than a number of good things, e.g., the combination of happiness and something else which is not good may be more desirable than the combination of justice and courage." I have to put questions concerning the

the person". However, we can also understand terms like "better than" in a more "objective" sense according to which the following holds: x being better than y does not entail that x is preferred by the person over y, and x being preferred by the person over y does not entail that x is better than y.

Given this notion of betterness, there seems to be a problem. John Broome has argued that betterness is a comparative relation (like x having more F than y) and thus necessarily transitive.[48] But how can the preference-relation then allow for intransitive cycles? This would mean that one could prefer x to y even if y is better than x. The acyclicity of "better than" would contrast with the potential cyclicity of "preferred". This, however, seems problematic.

One response could consist in arguing that "better than" also allows for intransitive cycles.[49] The betterness relation could be conceived not as one where one relatum has more of goodness than the other. The semantics of "better than" would have to be quite different from that of "more Fy than". For instance, "better than" could be conceived of as "contextual" in the sense that x being better than y does not entail anything (within the restrictions of (GNB) above) about the betterness relation between other pairs of options.

I have some sympathies for this way out of our problem but there are other, somewhat better (sic!) options. If "betterness" is understood in a more objective (see above) sense, then it is not that astonishing that someone could rationally prefer x to y, not knowing that y is better than x. This seems as uncontroversial as that one can prefer the

---

interpretation of Aristotle aside here, like the one whether Aristotle wants to cut off any close conceptual link between what is good and what is desirable (but see below on this topic).
48   See Broome 2004, 50-63, esp. 50-51.
49   See Herlitz 2020a, 2020b and also Vong 2018.

bad over the good (as the classic opponent of the idea of weakness of the will, Plato, already knew).[50] If, on the other hand, we understand "betterness" in a more subjective sense, then the problem does not get started in the first place because there is no difference between being better and being better in the light of the person's preferences or being preferred by the person.[51]

(g) No Dominance?

Another problem[52] which might be seen as an objection has to do with the principle of dominance. Suppose aPb, bPc, and cPa. Suppose further that there are 2 feasible acts A1 and A2 and 3 equally probable circumstances C1, C2, C3. Finally, assume the following outcome matrix:

	C1	C2	C3
A1	a	b	c
A2	b	c	a

Then A1 dominates A2 even though they have the same outcomes with the same probabilities, just under different circumstances. This is certainly implausible and suggest a modification of the principle of dominance for intransitive relations. It should be no

---

50 See Plato, Meno 77b-78b. - See also Andreou 2019 who defends the compatibility of acyclic betterness and potentially cyclic preference, interpreting "better" as more "choice-worthy".
51 See also Goldman 2020 who argues that the intransitivity of betterness or preference suggests a view according to which value is subjective or dependent on subjective attitudes.

surprise that giving up transitivity will have repercussions elsewhere in the theory. In this case, there should be an acceptable modification of the principle of dominance. Whatever the details, simple (and initially plausible) principles of dominance won't hold in cases of lack of transitivity. This should not be too surprising, given that other simple (and initially plausible) principles of classical theories of rational choice, like principles of utility maximization, will also not hold without at least some restrictions (see below on that latter point). Principles of dominance can become complex anyway when one considers particular cases.

(h) Briefly: Changed Preferences?

A final short objection: Apparent intransitivity might turn out to merely indicate a change of preferences during deliberation. This can, of course, happen. But one needs evidence for this: Saying that our preferences must have changed because otherwise they would be irrational is not convincing (and question-begging). Moreover, the possibility of a change of preferences in no way implies that our preferences cannot be intransitive without being irrational.[53]

---

52   See Fishburn 1991, 116, 118-119. This issue has not been raised much at all in the discussion about transitivity.

53   One might also be tempted to make the following point. If for options a, b, and c it is true that aPb, bPc, and cPa, then it is also true, e.g., that aPa and that bPa. However, this is necessarily false because strict preference is neither symmetric nor reflexive. Taken as an argument against rational intransitivity, this move is question-begging because it presupposes the transitivity-principle. But it would seem to give the defender of (T) an additional explanation of how, in their view, intransitive preferences are irrational.

IV. Alternative Descriptions: An Alternative?

So far I have argued for and defended the conclusion that transitivity is no general rationality constraint. It might seem that this conclusion can be avoided by changing the method of description of preferences. I want to discuss two such proposals now. The basic idea underlying them is that the description must be much more fine-grained.

According to the first proposal we should replace expressions of the form "mPn" by expressions of the form

"(m, when compared to n) P (n, when compared to m)".

It is obvious that this avoids intransitivity. Again, let

"c" stand for "getting pizza for $2",

"b" stand for "getting spaghetti for $4" and

"a" stand for "getting saltimbocca for $6".

Take our example about Fred. According to the present proposal it would have to be analyzed in the following way:

(b, when compared to c) P (c, when compared to b),

(a, when compared to b) P (b, when compared to a),

(c, when compared to a) P (a, when compared to c).

Clearly, this does not show any intransitivity. The pattern is this[54]:

(P$_2$)   xPy & zPu & vPw.

However, transitivity, too, would be hard to get. If a is preferred to b and b is preferred to c we cannot conclude that a is preferred to c because

(a, when compared to b) P (b, when compared to a)

and

(b, when compared to c) P (c, when compared to b)

do not (just relying on the transitivity assumption) imply that the following is true (for a rational person):

(a, when compared to c) P (c, when compared to a).

If we describe the situation in this way we do not get this preference structure

---

54   I am using the letters "x', "y", "z", "u", "v", and "w" as referring to the 6 relata of the preference relation just mentioned.

(P₃)   aPb & bPc & aPc

but this one:

(P₄)   zPu & xPy & wPv.

Obviously, neither intransitivity nor transitivity is conceivable anymore if we describe the situation in this way. Furthermore, this kind of re-description violates the intuitive idea that we should not describe the things we compare in terms of the comparison itself.[55] Hence, this way to re-describe the situation won't work.

There is a second and more interesting way to describe the situation in a more fine-grained way. According to it, it is not the options but rather the differences along the relevant evaluative dimensions that have to enter into the description of the preferences. Here is what I have in mind. Let us re-describe our first example. Instead of:

"(getting spaghetti for $4) P (getting pizza for $2)"

we have

"(getting spaghetti instead of pizza) P (spending $2 rather than $4)";

instead of

"(getting saltimbocca for $6) P (getting spaghetti for $4)"

we have

"(getting saltimbocca instead of spaghetti) P (spending $4 rather than $6)";

and instead of

"(getting pizza for $2) P (getting saltimbocca for $6)"

we have

"(spending $2 rather than $6) P (getting saltimbocca instead of pizza)".

Thus described, Fred's preference ranking is unproblematic. It has the form[56]:

(P$_5$) mPn & oPp & qPr.

---

55 See, for a similar proposal, Anand 1990, 94-96 and Lee 1984, 130ff., and, against it, Philips 1989, 481. See also Broome 1991a, 100ff. Anand 1990, 95-96 shows in addition how transitive preferences can be redescribed using the same method so as to turn them into intransitive preferences.

This is not at all paradoxical. The intransitivity vanishes if we give a fine-grained analysis of the preference ranking. We do not have to give up (T). Transitivity could still be a general rationality constraint.

There is nevertheless a price to pay for this "way out": transitivity, again, loses its place, too. Let us modify our original example. Assume Fred also prefers a \$6 saltimbocca to a \$2 pizza. According to the rough analysis, his preferences are transitive and have the form:

(P$_3$)   aPb & bPc & aPc

However, if we switch to the fine-grained analysis, his preference ranking has the following form:

(P$_6$)   mPn & oPp & rPq.

With the fine-grained analysis, not only the possibility of intransitivity but also the possibility of transitivity disappears. Preferences lose their internal connections. For our preference-triples the question of transitivity would not even arise.

Thus, we seem to face the following situation. We can either stick to a rough description of preferences and accept that transitivity is no general rationality constraint for

---

56   I am using the letters "m', "n", "o", "p", "q", and "r" as referring to the 6 relata of the preference relation just mentioned.

preferences (or: a rationality constraint only for some cases) or we can switch to a fine-grained description of preferences and keep the principle of transitivity at the price of devaluing or trivializing it.[57] In both cases, transitivity does not play the important role it prima facie seems to play. In other words: whether our preferences conform to or violate the principle of transitivity depends on how we describe them.[58] Under fine-grained descriptions the question of transitivity and intransitivity does not even arise. Under rough descriptions, transitivity is possible but only in some cases. We should make a choice between the two. I think there are good reasons to stick with the rough description.[59]

V. Implications

The main conclusion that transitivity is no necessary condition for the rationality of preferences has far-reaching implications and raises serious questions.

(a) Simple Cases vs. Extreme Cases. First of all, there is a methodological lesson to be learnt from all this. Why does (T) seem so convincing as a universal rationality constraint in the first place? It seems to work for the most basic and simple cases: if you prefer a banana to an apple and an apple to a peach then you should also rationally prefer a banana to a peach. One might be tempted to argue that if transitivity is a rationality

---

57   One might also want to say that the concepts of transitivity or intransitivity have no meaningful application in such cases. See Temkin 2012, 17 and Kagan 2015, 474-475 on this.
58   See Anand 1993, 103ff.
59   See Broome 1991b, 80-81 and Temkin 2012, 457-465. – According to one claim, choice is essentially comparative (see above); according to another claim, the above idea of fine-grained description, the comparison is part of the content of the options. The first claim (favored here) does not entail the second. The rejection of the second proposal is compatible with the defense of the utility-difference model defended here. Both use comparisons of differences (in taste, in money) but while these

constraint for such simple cases then it is also valid for the more complex cases. The Cartesian methodological rule to start the analysis with simple cases is, of course, a good one. However, we have seen that the simple banana-apple-peach case is not a basic but a limiting case. It is a typical textbook case but it surely does not represent many real-life cases. Thus, it can be misleading to focus on simple cases. As Wittgenstein once remarked: "A main cause of philosophical disease -- a one-sided diet: one nourishes one's thinking with only one kind of example".[60]

(b) The Good, the Beautiful, etc. "Better than" is just one among many other evaluative relations, like "more beautiful than" or "morally better" than. One could suspect that one can construct analogous anti-transitivity-arguments for morality or aesthetics or whatever. Should "x is a better work of art than y" really be a transitive relation? It seems much easier to doubt this than the claim that rational preferences are transitive. But why? The reason cannot be that aesthetic judgments are not subject to rationality constraints; they are. Similar questions can be raised about moral betterness.[61] My aim here is only to point to this kind of question.

(c) No Rational Choice? How can there still be a rational choice in the case of intransitive or cyclical preferences? Isn't the person forced to run around in circles of intransitivity and end up like Buridan's ass? Or is the person forced to make an irrational decision? No, it is rather that her choice of one of the options is beyond rationality or

---

comparisons are part of the content of the options according to the second proposal rejected here, they are rather part of the method of evaluation of options defended in this paper.

60   Wittgenstein 1958, § 593.

61   On the acyclicity of "morally better than" see Andreou 2019, sec.5. For intransitivity in theory choice, see Baumann 2005; for intransitivity with respect to "is a more meaningful life", see Baumann 2015. For interesting cases of "legal intransitivity" see Naeh and Segal 2009.

irrationality; rationality theory remains silent about which option to choose. This does not mean that there are no demands of rationality at all here. Certainly, the person should make some decision (Buridan's ass died because it neglected this important point). That we do have a reason to do something does not imply that there is something we have a reason to do. The person therefore does not need to run around in circles forever. She can and should just pick some option.[62] Thus, giving up (T) as a general rationality constraint does not lead to the disaster of irrationality or to a paralysis of decision-making.[63]

(d) Maximization and the Best. Many theories of individual rationality as well as many moral theories are based on the idea of maximization. Classical decision theory is just one proponent of the idea that the rational person maximizes her own well-being or expected utility. Utilitarianism tells us that we should maximize the happiness of the greatest number of people. All these different ideas of maximization are facing problems, given the possibility of rational intransitivity. A "local" problem arises with respect to a particular rational cycle of preferences: Within the relevant set of options there is then no best option for the person. If the cycle is such that each of the options involved is preferred to any other option not involved, then the problem is "global": Overall, there is no best option for the person then. The possibility of rational intransitivity at least loosens the connection between the better and the best and raises some doubts about the possibility of maximization.[64] I argued above that we have to explain "good" in terms of "better"; at the

---

62   See for the important difference between picking and choosing: Ullmann-Margalit/ Morgenbesser 1977, 757ff.

63   See, e.g., Fishburn 1991, 117-118 or Herlitz 2020a and 2020b.

64   For a very different argument for a very similar conclusion see Slote 1989. – One can imagine an ordered series of infinitely many options which has neither a beginning nor an end and where each option is better than its predecessor. Such a series would have a betterness ordering without a best or a worst option. The notion of a best or a worst would have no application in this (admittedly far-

same time, the conceptual connection between "best" on the one hand and "better" or "good" on the other hand becomes looser. Similar things can be said about the idea of the "worst".

(e) How to Live without (T)? Without transitivity, it is not clear how and to what degree one can still assign numerical values to options and measure utilities and strengths of preferences; even a restricted version of the "representation theorem" is on uncertain grounds. More generally, it is not clear which mathematical techniques are still available to the decision theorist. Many would not like such loss of mathematical power. But this doesn't mean that one cannot use formal methods any more[65], - though the question remains what a not so orthodox, non-transitive decision theory could look like.[66] Could there be a different, non-classical theory of rational choice which still uses formal methods and is still able to measure preferences in some way? Or should we be more radical and give up the idea of a "theory of rationality" in the sense of something that gives us a set of rules of rationality? Should we rather go back to old ideas of implicit knowledge, practical wisdom and phronesis? Should we go even further and forget about rationality and become nihilists or skeptics or relativists about rationality? Orthodoxy in despair about losing (T) might expect this. I don't think there is reason for any such extreme reaction. There doesn't seem to be any good reason to give up the project of finding another theory of rationality, one that is certainly more modest in its ambitions than orthodox theories of rational choice.

---

fetched) case. For other comments on cases involving infinity see Kagan 2015, 471-472 or Andreou 2019, 1629.

65   But see, for instance, the work of Fishburn (e.g., 1991).

66   But see, for instance, Fishburn 1988.

References

Aldred J (2007) Intransitivity and Vague Preferences. *The Journal of Ethics* 11: 377-403.

Anand P (1990) Interpreting Axiomatic (Decision) Theory. *Annals of Operations Research* 23: 91-101.

Anand P (1993) *Foundations of Rational Choice under Risk*. Oxford: Clarendon.

Andreou C (2019) Better than. *Philosophical Studies* 176: 1621-1638.

Aristotle (1991) *Topics*. Transl. Pickard-Cambridge W A. *The Complete Works of Aristotle: The Revised Oxford Translation* (ed.: Jonathan Barnes). Princeton: Princeton University Press, vol.1.

Armstrong W E (1939) The Determinateness of the Utility Function. *The Economic Journal* 49: 453-467.

Arrow K J (1963) *Social Choice and Individual Values* (2.ed.). New York: Wiley.

Bar-Hillel M and Margalit A (1988) How Vicious are Cycles of Intransitive Choice? *Theory and Decis*ion 24: 119-145.

Barrett J (2019) Interpersonal Comparisons with Preferences and Desires. *Politics, Philosophy & Economics* 18: 219-241.

Baumann P (2005) Theory Choice and the Intransitivity of *Is A Better Theory Than*. *Philosophy of Science* 72: 231-240.

Baumann P (2015) Meaningful and More Meaningful. A Modest Measure. *The Journal of Philosophy of Life* 5.3: 33-49.

Baumann P (2018) What Will Be Best for Me? Big Decisions and the Problem of Inter-World Comparisons. *Dialectica* 72: 253-273.

Baumann P Life Pumps. Unpublished manuscript.

Benbaji Y (2009) Parity, Intransitivity and a Context-Sensitive Degree Analysis of Gradability. *Australasian Journal of Philosophy* 87: 313-335.

Binmore K and Voorhoeve A (2003) Defending Transitivity against Zeno's Paradox. *Philosophy and Public Affairs* 31: 272-279.

Broome J (1991a) *Weighing Goods*. Oxford: Blackwell.

Broome J (1991b) Rationality and the Sure-Thing Principle. In: Meeks G (ed) *Thoughtful Economic Man*. Cambridge: Cambridge University Press, pp. 74-102.

Broome J (1993) Can a Humean be Moderate? In: Frey R G and Morris C W (eds) *Value, Welfare, and Morality*. Cambridge: Cambridge University Press, pp. 51-73.

Broome J (2004) *Weighing Lives*. Oxford: Oxford University Press.

Carlson E (1996) Cyclical Preferences and Rational Choice. *Theoria* 62: 144-160.

Carlson E (2003) On some Recent Examples of Intransitive Betterness. In: Segerberg K and Sliwinski R (eds) *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åquist*. Uppsala: Department of Philosophy, Uppsala Philosophical Studies 51 / Uppsala University, pp. 181-195.

Carlson E (2005) Intransitivity without Zeno's Paradox. In: Ronnow-Rasmussen T and Zimmerman M (eds) *Recent Work in Intrinsic Value*. Dordrecht: Springer, pp. 273-277.

Chang R (ed.) (1997) *Incommensurability, Incomparability and Practical Reason*. Cambridge/MA: Harvard University Press.

Chang R (2002) *Making Comparisons Count*. New York & London: Routledge.

Chisholm R M and Sosa E (1966) On the Logic of "Intrinsically Better". *American Philosophical Quarterly* 3: 244-249.

Coons C (2014) Hope for Fools: Four Proposals for Meeting Temkin's Challenge. *Analysis* 74: 292-306.

Danielsson S (1996) The Refutation of Cyclic Evaluations. *Theoria* 62: 161-168.

Davidson D (1985) Incoherence and Irrationality. *Dialectica* 39: 345-354.

Davidson D, McKinsey J C C, and Suppes P (1955) Outlines of a Formal Theory of Value I. *Philosophy of Science* 22: 140-160.

Doody R (2019) Opaque Sweetening and Transitivity. *Australasian Journal of Philosophy* 97: 579-591.

Fishburn P C (1984) Dominance in SSB Utility Theory. *Journal of Economic Theory* 34: 130-148.

Fishburn P C (1988) *Nonlinear Preference and Utility Theory*. Baltimore and London: The Johns Hopkins University Press.

Fishburn P C (1991) Non-Transitive Preferences in Decision Theory. *Journal of Risk and Uncertainty* 4: 113-134.

Gauthier D (1986) *Morals by Agreement*. Oxford: Clarendon.

Gendin S (1996) Why Preference Is not Transitive. *The Philosophical Quarterly* 46: 482-488.

Goldman A H（2020）Intransitivity, Essential Comparativeness, and Objective Value. *Metaphilosophy* 51: 14-31.

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.

Handfield T and Rabinowicz W（2018）Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness. *Philosophical Studies* 175: 2372-2387.

Hansson S O (1993) Money Pumps, Self-Torturers and the Demons of Real Life. *Australasian Journal of Philosophy* 71: 476-485.

Helson H（1994）*Linear Algebra* (2.ed.). Delhi: Hindustan Book Agency.

Herlitz A（2020a）Non-Transitive Better than Relations and Rational Choice. *Philosophia* 48: 179-189.

Herlitz A（2020b）Correction to: Non-Transitive Better than Relations and Rational Choice. *Philosophia* 48: 431.

Horowitz T (2006) Making Rational Choices when Preferences Cycle. In: Horowitz T, *The Epistemology of A Priori Knowledge* (ed.: Joseph L. Camp, jr.). Oxford: Oxford University Press, pp. 102-122.

Hsee C K, Blount S, Loewenstein, G F, Bazerman M H (1999) Preference Reversals between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis. *Psychological Bulletin* 125: 576-590.

Huemer M (2013) Review of Temkin, Rethinking the Good. *Ethics* 123: 318-45.

Hughes R I G (1980) Rationality and Intransitive Preferences. *Analysis* 40: 132-134.

Jeffrey R C (1974) Preference among Preferences. *The Journal of Philosophy* 71: 377-391.

Kagan S (2015) The Costs of Transitivity: Thoughts on Larry Temkin's *Rethinking the Good*. *Journal of Moral Philosophy* 12: 462-478.

Klocksiem J (2016) How to Accept the Transitivity of *better than*. *Philosophical Studies* 173: 1309-1334.

Knapp C (2007) Trading Quality for Quantity. *Journal of Philosophical Research* 32: 211-233.

Lee R (1984) Preference and Transitivity. *Analysis* 44: 129-134.

Lehrer K and Wagner C (1985) Intransitive Indifference: The Semi-Order Problem. *Synthese* 65: 249-256.

Luce R. D and Raiffa H (1957) *Games and Decisions. Introduction and Critical Survey*. New York etc.: Wiley.

Mackie J L (1985) The Combination of Partially-Ordered Preferences. In: Mackie J L, *Persons and Values. Selected Papers*, Volume II (eds.: Mackie J and Mackie P). Oxford: Oxford University Press, pp. 249-253.

Maher P (1993) *Betting on Theories*. Cambridge: Cambridge University Press.

Mandler M (2001) A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but not Both. In: Millgram E (ed) *Varieties of Practical Reasoning*. Cambridge, MA: MIT Press, pp. 373-402.

Michalos A C (1967) Postulates of Rational Preference. *Philosophy of Science* 34: 18-22.

Morrison Jr. H W (1962) *Intransitivity of Paired Comparison Choices*. University of Michigan, Ann Arbor (Ph.D. Dissertation).

Naeh S and Segal U (2009) The Talmud on Transitivity. *Boston College Working Papers in Economics* 687: 1-12.

Nebel J M (2018) The Good, the Bad, and the Transitivity of *Better than*. *Noûs* 52: 874-800.

Nozick R (1993) *The Nature of Rationality*. Princeton: Princeton University Press.

Parfit D (1984) *Reasons and Persons*. Oxford: Clarendon.

Philips M (1989) Must Rational Preferences be Transitive? *The Philosophical Quarterly* 39: 477-483.

Plato (1997) *Meno*. Transl. Grube G M A. In: *Complete Works* (ed.: John M. Cooper). Indianapolis and Cambridge: Hackett.

Pummer T (2018) Spectrum Arguments and Hypersensitivity. *Philosophical Studies* 175: 1729-1744.

Putnam H (1986) Rationality in Decision Theory and in Ethics. *Critica* 18: 3-16.

Qizilbash M (2005) Transitivity and Vagueness. *Economics and Philosophy* 21: 109-131.

Quinn W S (1990) The Puzzle of the Self-Torturer. *Philosophical Studies* 59: 79-90.

Rachels S (1998) Counterexamples to the Transitivity of *Better Than*. *Australasian Journal of Philosophy* 76: 71-83.

Rachels S (2004) Repugnance or Intransitivity. In: Ryberg J and Tannsjo T (eds) *The Repugnant Conclusion*. Dordrecht: Kluwer, pp. 163-186.

Raiffa H (1968) *Decision Analysis. Introductory Lectures on Choices under Uncertainty*. Reading/MA etc.: Addison-Wesley.

Ramsey F P (1990) Truth and Probability. In: Ramsey F P, *Philosophical Papers* (ed.: David Hugh Mellor). Cambridge: Cambridge University Press, pp. 52-94.

Rawling P (1990) The Ranking of Preference. *The Philosophical Quarterly* 40: 495-501.

Raz J (1986) *The Morality of Freedom*. Oxford: Clarendon.

Savage L J (1972) *The Foundations of Statistics* (2. ed.). New York: Dover Publications.

Schick F (1986) Dutch Bookies and Money Pumps. *The Journal of Philosophy* 83: 112-119.

Schumm G F (1987) Transitivity, Preference and Indifference. *Philosophical Studies* 52: 435-437.

Slote M (1989) *Beyond Optimizing: A Study of Rational Choice*. Cambridge/MA: Harvard University Press.

Suppes P (1958) *Introduction to Logic*. Princeton: Nostrand.

Temkin L S (1996) A Continuum Argument for Intransitivity. *Philosophy and Public Affairs* 25: 175-210.

Temkin L (2000) An Abortion Argument and the Threat of Intransitivity. In: Crisp R and Hooker B (eds) *Well-Being and Morality. Essays in Honor of James Griffin*. Oxford: Clarendon, pp. 263-279.

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

Tullock G (1964) The Irrationality of Intransitivity. *Oxford Economic Papers* 16: 401-406.

Tversky A (1969) Intransitivity of Preferences. *Psychological Review* 76: 31-48.

Ullmann-Margalit and Morgenbesser S (1977) Picking and Choosing. *Social Research* 44: 757-785.

von Neumann J and Morgenstern O (1953) *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

von Wright G H (1963) *The Logic of Preferences*. Edinburgh: Edinburgh University Press 1963.

Vong G (2018) Making Non-Transitive Betterness Behave. *Journal of Moral Philosophy* 15: 495-515.

Voorhoeve A (2008) Heuristics and Biases in a Purported Counter-example to the Acyclicity of 'Better Than'. *Politics, Philosophy, and Economics* 7: 285-299.

Voorhoeve A and Binmore K (2006) Transitivity, The Sorites Paradox, and Similarity-Based Decision-Making. *Erkenntnis* 64: 101-114.

Wittgenstein L (1958) *Philosophical Investigations* (2. ed.). Oxford: Blackwell.

Wodak D (2019) What if Well-Being Measurements are Non-Linear? *Australasian Journal of Philosophy* 97: 29-45.

Short Author Biography

Peter Baumann is Full Professor of Philosophy at Swarthmore College. His research interests are mainly in Theoretical Philosophy and in theories of rationality as well as in 18th Century Philosophy.