Hello World

Brian Weatherson

2025-10-22

Abstract

There is a natural view of the relationship between preference and choice: an option is choiceworthy if and only if no alternative is strictly preferred to it. I argue against this view on two grounds. First, it makes false predictions about which options are choiceworthy in games and in multi-dimensional choice settings. Second, it conflates two distinct attitudes: choiceworthiness, which is assessed ex ante, and preference, which is assessed ex post. I explore the consequences of rejecting this natural view, including how it simplifies the relationship between game theory and decision theory, and how it complicates debates about what Ruth Chang calls 'parity' between options.

Subjective decision theory concerns norms about three concepts: values of options, preferences between pairs of options, and choices out of sets of options. A common assumption, at least implicitly, is that norms about preferences are prior to norms on values and choices. One way to put this assumption, following Amartya Sen ([1970] 2017), is that choice functions are *binary*; they are grounded in binary relations of preference and indifference.

I'm going to argue against this for two reasons. First, preferences, being binary comparisons, don't provide a rich enough base to ground all the norms. Sometimes decision theorists need to take as primitive comparisons the chooser (hereafter, Chooser) makes between larger sets of options. Second, preference is an ex post notion, in a sense to be made clearer starting in **?@sec-multieq**, while choiceworthiness is an ex ante notion. And ex ante norms are not grounded in ex post ones.

Choice and Choiceworthiness

Choice gets less attention in philosophical decision theory than one might expect. The focus is usually on either value (e.g., this has value 17 and that has value 12) or preference (e.g., this is preferable to that). Norms on choice are almost an afterthought in standard presentations. After a long discussion of values, preferences, or both, the typical theorist breezily says that the norm is to choose the most valuable or most preferred option.

There is a long tradition in economics, going back to Paul Samuelson (1938) and Herman Chernoff (1954), of taking choice to be primary. Some of this literature rested on largely behaviourist or positivist assumptions. It was better to theorise with and about choice because it was observable, unlike preferences or values. The picture was not dissimilar to this recently expressed view:

> Standard economics does not address mental processes and, as a result, economic abstractions are typically not appropriate for describing them. (Gul and Pesendorfer 2008, 24)

That's not going to be my approach here. I'm going to start not with observable choice dispositions, like the economists, or with choice frequencies, as psychologists like R. Dunacn Luce (1959) do, but with judgments about choiceworthiness. In familiar terminology[1], I'm taking a mentalist approach not a behaviourist approach. Much of the formal work on choice theory has been done by theorists from the behaviourist side, and I'll be inevitably drawing from their work. But the most important source I'll be using is someone much more sympathetic to mentalism: Amartya Sen. In particular, I'll draw heavily on his "Collective Choice and Social Welfare" (Sen [1970] 2017), and also on the literature that grew out of that book.

I'm not going to take a stance on the metaphysics of choiceworthiness judgments. I'll sometimes talk as if they are beliefs. However, if someone wanted to maintain a sharp belief-desire split and hold that choiceworthiness involves an interplay of the two (like preference), I wouldn't object. The real assumption is that the mental state being ascribed in choiceworthiness ascriptions is the same kind of state as that being ascribed in preference ascriptions. The biggest difference between choice and preference is that choiceworthiness is a relation to an arbitrarily sized set of options, while preference is a relation to a pair of options.

Values First?

Both to clarify what kind of question I'm asking, and to set aside one kind of answer, I'm going to start looking not at preferences or choices, but at numerical values. At first glance, it might seem that many decision theorists take values to ground the other norms. One should prefer the more valuable. So we get theorists discussing Newcomb's Problem largely by offering theories about the value of uncertain outcomes (like taking both boxes) in terms of the values and probabilities of those outcomes.

On second glance, though, there are at least four reasons it is implausible that values are really what ground the other attitudes.

First, it's surprising to have a numerical measure like this not have a unit. We'll sometimes say things like this outcome has value 17 *utils*, but this is a placeholder, not a real unit like kilograms or volts. This is related to the second reason.

[1]See, for instance, Hansson and Grüne-Yanoff (2024), where I learned about the Gul and Pesendorfer quote.

The orthodox view is that these values are only defined up to a positive affine transformation. If it's appropriate to represent Chooser with value function v, it's appropriate to represent them with any value function f where $f(o) = av(o) + b$, for positive a. Why is this transformation allowed? Because all the values are doing is reflecting Chooser's preferences over outcomes and lotteries, and this transformation doesn't change those preferences. In other words, the transformation is allowed, and the definedness claim is true, because values are grounded in comparatives.

Third, it's not at all obvious why values should be anything like numbers. Indeed, the thought that they should be numbers starts causing problems when we get to various puzzles about infinite goods.[2] Why should values have the topology of the reals rather than any number of other possible topologies? Why aren't they, for example, quintuples of rational numbers ordered lexicographically? If one takes preferences to be primary, and generates utility functions via representation theorems, as in Ramsey (1926) or von Neumann and Morgenstern (1944), there is a reason for why values should be numbers. But if values are primitive, it seems like an unanswered and I'd say unanswerable question.

Finally, there is something very strange about the idea of values that are not in any way comparative. How valuable something is just seems like it should be a notion that reflects the value of alternatives to it.

The argument of this section will not be, I suspect, particularly contentious. It's a widespread view, if often implicit, that that values, and norms on values, are ultimately grounded in comparatives. What is going to be contentious is the claim that preferences can't do the job, and judgments of, and norms about, choiceworthiness ultimately ground both values and preferences.

Coherence

There is one other striking thing about the picture we get in von Neumann and Morgenstern (1944), and which is still I think broadly endorsed by contemporary decision theorists. The aim is to put norms on preference, and hence on values and choices. But these norms are almost always defined in terms of other preferences. For example, if one strictly prefers x to y, and y to z, one should prefer x to z. It is strange to talk about grounding the normative facts about preference when other preferences play such a crucial role. It looks like the grounding relation will be at least cyclic, and possibly intransitive.

If there's a puzzle about this apparent circularity, there are two (related) ways out. One is to take the view, which perhaps Hume held, that the only constraints on preference are coherence constraints. Another is to say that while there might be some non-coherence constraints on preference, e.g., it is in fact wrong to prefer

[2] See, e.g., Nover and Hàjek (2004), or Goodman and Lederman (2024).

the world's destruction to one's finger getting scratched, these are part of a separate subject to decision theory. The result is the same: decision theory largely is about what it takes for various preferences to cohere with one another.

I don't particularly agree with this picture, but I'm going to accept half of it for the purposes of this paper. That is, I'll assume it's not the role of decision theory to criticise the person who prefers the destruction of the world to the scratching of their finger. That person either violates no norms, or violates a different kind of norm to the norms of decision theory. Decision theory, on the latter view, takes Chooser's preferences over ends as given and judges Chooser on how well their instrumental preferences serve these preferences over ends.

I'll argue that even if decision theory is about coherence, it should be about coherence between choiceworthiness judgments. So the big question is not whether preferences should satisfy transitivity or independence, but whether their choiceworthiness judgments should satisfy conditions like those described in Section . As we'll see, the questions about which conditions are genuine coherence constraints on choiceworthiness are tied up with the metaphysical question about the priority of preferences and choices.

Sen on Preference

The main binary relation Sen uses, which he denotes R, is such that xRy means that Chooser either prefers x to y, or is indifferent between x and y. The first disjunct obtains if $\neg yRx$, the second disjunct obtains if yRx. (As he sometimes puts it, these are the symmetric and asymmetric parts of the relation.) We'll write these two disjuncts as xPy and xIy. Formally, that is, they can be defined in terms of R in (1) and (2). (Throughout, I'm leaving off wide scope universal quantifiers over free variables.)

(1) $xPy \square (xRy \wedge \neg yRx)$

(2) $xIy \square (xRy \wedge yRx)$

It is important that indifference not be understood as equality. It is not assumed that I is transitive; indeed Sen makes great use of models where it is not. If we take transitivity to be part of the definition of equality, it is misleading to gloss xRy as that x is greater than **or equal** to y (for Chooser). For this reason I won't use $\square$ to denote it, since that is most naturally read as that x is better than or equal to y.[3]

A more common way of doing things in contemporary philosophy is to start with P and a fourth relation E, where xEy means that x and y are equally good. On this picture, both (1) and (2) are true, but the explanatory direction in both cases is right-to-left. So xRy just is $\neg yPx$, and then xIy is still defined via (2). On the version Sen uses, it's a little trickier to define E, but (3) looks like a plausible conjecture.

[3]Indeed, I'll use $x \square y$ to explicitly mean that x is better than or equal to y, where equality is understood in the sense of the next paragraph.

(3) $xEy \leftrightarrow [(xRz \leftrightarrow yRz) \wedge (zRx \leftrightarrow zRy)]

That is, two options are equally good iff they are substitutable in other preference relations. Given all these results, we can show that the following claims are all tightly connected.

(4) $xPy \vee xEy \vee yPx$

(5) $(xPy \wedge yIz) \rightarrow xPz$

(6) $(xIy \wedge yIz) \rightarrow xIz$

(4) is what Ruth Chang (2017) calls the trichotomy thesis. (5) is what Sen calls PI-transitivity, and (6) is what he calls II-transitivity.

Sen makes very few assumptions about R, but it will simplify our discussion to start introducing some assumptions here.[4] We'll assume that R is reflexive, everything is at least as good as itself, and that P is transitive. Sen ([1970] 2017, 66) notes that if P is transitive and R is 'complete' in the sense that $xRy \vee yRx$ holds for arbitrary x and y, then (5) and (6) are equivalent. It's also easy to show that given (3) plus these assumptions, (4) and (6) are equivalent.[5]

What should we call the principle (4)? The terminology around here gets potentially confusing. If we define $x \succeq y$ to just mean the disjunction $xPy \vee xEy$, and assume that E is symmetric, then (4) is equivalent to $x \succeq y \vee y \succeq x$. That's what Gustafsson (Forthcoming) calls *completeness*, and I feel that's often what philosophers understand by 'completeness'. In his economic work, Sen ([1970] 2017) uses the term 'completeness' for a slightly different property of preference relations, namely $xRy \vee yRx$. In both cases the claim is that some preference relation is guaranteed to hold in one direction or other. The issue is whether that relation is the disjunction of P and E, or the disjunction of P and I.

Both of these notions are useful to have. There has been a huge amount of literature on (4), i.e., $x \succeq y \vee y \succeq x$, less so on $xRy \vee yRx$. But the latter is useful because various interesting possibilities open up both in social choice theory, and

[4]He makes few assumptions because he was interested in exploring what assumptions about preference are crucial to the impossibility theorem that Arrow (1951) derives. He initially noticed that without (6), Arrow's theorem didn't go through. This turned out to be less significant than it seemed, because Allan Gibbard (2014) proved that a very similar theorem can be proven even without (6). See Sen (1969) for the original optimism that this might lead to an interesting way out of the Arrovian results, and Sen ([1970] 2017) for a more pessimistic assessment in light of Gibbard's result. Sen reports that Gibbard originally proved his result in a term paper for a seminar at Harvard in 1969 that he co-taught with Arrow and Rawls. Much of what I'm saying in this paper can be connected in various ways to the literature on Arrow's impossibility theorem, but I won't draw out those connections here.

[5]Proof: Assume (4) is false. So the right hand side of (3) is false. Without much loss of generality, assume that $xRz \wedge \neg yRz$; the other cases all go much the same way. So all the disjuncts are false. From $\neg xPy$ and $\neg yPx$ we get $yRx \wedge xRy$, i.e., xIy. And xRz implies zIx. So we have a counterexample to II-transitivity, since zIx and xIy, but since $\neg yRz$, yIz is false. So if (4) is false, (6) is false. In the other direction, assume we have a counterexample to (6), i.e., xIy and yIz but not xIz. From xIy we immediately get that the two outer disjuncts of (4) are false. From yIz we get yRz and zRy. So if xEy, (3) implies that xRz and zRx, i.e., xIz. But we assumed that $\neg xIz$. So all three disjuntcs of (4) are false. That is, if (6) fails, so does (4), completing the proof that they are equivalent.

in the relationship between preference and choice, if it is dropped.[6]

While (4) is widely discussed in philosophy, it isn't often discussed as such in the economics literature. That literature does contain several works discussing (6), starting with important works by Wallace E. Armstrong (1939, 1948, 1950). In most of those works it is assumed that P is transitive and $xRy \lor yRx$, so (4) and (6) are equivalent. But the different focus leads to more terminological confusion.

More generally, I find using 'completeness' to denote either $x \succeq y \lor y \succeq x$ or $xRy \lor yRx$ potentially confusing, since in either case it is easy to mistakenly believe the author is talking about the other property. So I'll follow Chang (2017) and say that relations satisfying (4) are *trichotomous*, and use *definedness* for $xRy \lor yRx$.

Properties of Choice Functions

In philosophy we're sufficiently familiar with possible properties of preference relations, e.g., that they are transitive, reflexive, acyclic, etc, that these terms don't need to be defined. We're mostly less familiar with properties of choice functions. A choice function takes a set of options as input, and returns a non-empty subset of that set as output. The elements of the output are the choiceworthy members of the original set. So $C(\{a, b, c, d\}) = \{a, c\}$ means that if Chooser has to pick from a, b, c and d, then a and c are choiceworthy, and b and d are not.

I'll start by laying six properties that choice functions may have which will be important in what follows. The first four are discussed in some detail by Sen ([1970] 2017), and I'll use his terminology for them. The fifth is due to Aizerman and Malishevski (1981), and is usually named after Aizerman. The sixth is discussed by Blair et al. (1976).

Property α $(x \in C(S) \land x \in T \land T \subseteq S) \rightarrow x \in C(T)$

That is, if x is choiceworthy in a larger set, then it is choiceworthy in any smaller set it is a member of. This is sometimes called the *Chernoff condition*, after Herman Chernoff (1954), and sometimes called *contraction consistency*.

Property β $(x \in C(T) \land y \in C(T) \land T \subseteq S) \rightarrow (x \in C(S) \leftrightarrow y \in C(S))$

That is, if x and y are both choiceworthy in a smaller set, then in any larger set they are either both choiceworthy or neither is. Intuitively, if x and y are ever both choiceworthy, then anything better than x is also better than y.

Property γ $(x \in C(S) \land x \in C(T)) \rightarrow (x \in C(S \cup T))$

That is, if x is choiceworthy in two sets, it is choiceworthy in their union. This is sometimes called *expansion*, e.g., by Hervé Moulin (1985).

Property δ $(x \in C(T) \land y \in C(T) \land T \subseteq S) \rightarrow (\{y\} \neq C(S))$

This is a weakening of β. It says that if x and y are both choiceworthy in the smaller set, then after options are added, it can't be that only one of them is the only choiceworthy option remaining. If x is not choiceworthy in the larger set,

[6] On the latter, see Bradley (2015).

that's because some other option, not y, is chosen in place of it.

Property Aiz $(C(S) \square\ T \wedge T \square\ S) \square\ C(T) \square\ C(S)$

That is, if the smaller set contains all of the choiceworthy members of the larger set, then no option is choiceworthy in the smaller set but not the larger set. If x is an unchoiceworthy member of S, then it can only become choiceworthy by deleting choiceworthy members of S, not other unchoiceworthy ones..

Path Independence $C(S \cup T) = C(C(S) \cup C(T))$

The same options are choiceworthy from a union of two sets as are choiceworthy from the union of the choiceworthy members of those sets. This is a sort of independence of irrelevant alternatives principle; the availability or otherwise of unchoiceworthy members of S and T doesn't affect what should be chosen from $S \cup T$.

I'll describe the effects of these properties in more detail in subsequent sections.

Property α

This is the most commonly used constraint on choice functions, and it does seem intuitive. If x is choiceworthy from a larger set, deleting unchosen options shouldn't make it unchoiceworthy. Sen ([1970] 2017, 323–26) discusses two possible counterexamples.

One is where the presence of options on the menu gives Chooser relevant information. If the only two options are tea with a particular friend or staying home, Chooser will take tea. But if the option of cocaine with that friend is added, Chooser will stay home. The natural thing to say here is that when one gets new information, C changes, so there isn't really a single C here which violates α.[7]

The more interesting case is where the value Chooser puts on options is dependent on what options are available. So imagine Chooser prefers more cake to less, but does not want to take the last slice. If the available options are zero slices or one slice of cake, Chooser will take zero. But if two slices of cake is an option, Chooser will take one, again violating α.

This is a trickier case, and the natural thing to say is that Chooser doesn't really have the same options in the two cases. Taking the last slice of cake isn't the same thing as taking one slice when two are available. But this move has costs. In particular, it makes it hard to say that C should be defined for any set of options. It doesn't clearly make sense to ask Chooser to pick between *taking one slice, which is the last*, and *taking three slices when five are available*.

Still, I'm going to set those issues aside and assume, like most theorists do, that α is a constraint on coherent choice functions, and that choice functions are defined over arbitrary sets of options.

[7] For a quick argument for that, if Chooser learns the only options are tea and staying home because the friend has just run out of cocaine, they might still stay home.

Aizerman, M., and A. Malishevski. 1981. "General Theory of Best Variants Choice: Some Aspects." *IEEE Transactions on Automatic Control* 26 (5): 1030–40. https://doi.org/10.1109/TAC.1981.1102777.

Armstrong, W. E. 1939. "The Determinateness of the Utility Function." *The Economic Journal* 49 (195): 453–67. https://doi.org/10.2307/2224802.

———. 1948. "Uncertainty and the Utility Function." *The Economic Journal* 58 (229): 1–10. https://doi.org/10.2307/2226342.

———. 1950. "A Note on the Theory of Consumer's Behaviour." *Oxford Economic Papers* 2 (1): 119–22. https://doi.org/10.1093/oxfordjournals.oep.a041384.

Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. New York: John Wiley & Sons.

Blair, Douglas H, George Bordes, Jerry S Kelly, and Kotaro Suzumura. 1976. "Impossibility Theorems Without Collective Rationality." *Journal of Economic Theory* 11 (3): 361–79. https://doi.org/10.1016/0022-0531(76)90047-8.

Bradley, Richard. 2015. "A Note on Incompleteness, Transitivity and Suzumura Consistency." In *Individual and Collective Choice and Social Welfare: Essays in Honor of Nick Baigent*, edited by Constanze Binder, Giulio Codognato, Miriam Teschl, and Yongsheng Xu, 31–47. Berlin: Springer.

Chang, Ruth. 2017. "Hard Choices." *Journal of the American Philosophical Association* 3 (1): 1–21. https://doi.org/10.1017/apa.2017.7.

Chernoff, Herman. 1954. "Rational Selection of Decision Functions." *Econometrica* 22 (4): 422–43. https://doi.org/10.2307/1907435.

Gibbard, Allan F. 2014. "Social Choice and the Arrow Conditions." *Economics and Philosophy* 30 (3): 269–84. https://doi.org/10.1017/S026626711400025X.

Goodman, Jeremy, and Harvey Lederman. 2024. "Maximal Social Welfare Relations on Infinite Populations Satisfying Permutation Invariance." https://arxiv.org/abs/arXiv:2408.05851.

Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." In *Foundations of Positive and Normative Economics*, edited by Andrew Caplin and Andrew Schotter, 2–40. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328318.003.0001.

Gustafsson, Johan E. Forthcoming. "A Behavioural Money-Pump Argument for Completeness." *Theory and Decision*, Forthcoming. https://doi.org/10.1007/s11238-025-10025-3.

Hansson, Sven Ove, and Till Grüne-Yanoff. 2024. "Preferences." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2024. https://plato.stanford.edu/archives/win2024/entries/preferences/; Metaphysics Research Lab, Stanford University.

Luce, R. Duncan. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Moulin, Hervé. 1985. "Choice Functions over a Finite Set: A Summary." *Social Choice and Welfare* 2 (2): 147–60. https://doi.org/10.1007/BF00437315.

Neumann, John von, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Nover, Harris, and Alan Hàjek. 2004. "Vexing Expectations." *Mind* 113 (450): 237–49. https://doi.org/10.1093/mind/113.450.237.

Ramsey, Frank. 1926. "Truth and Probability." In *Philosophical Papers*, edited by D. H. Mellor, 52–94. Cambridge: Cambridge University Press.

Samuelson, Paul A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Econometrica* 5 (17): 61–71. https://doi.org/10.2307/2548836.

Sen, Amartya. 1969. "Quasi-Transitivity, Rational Choice and Collective Decisions." *The Review of Economic Studies* 36 (3): 381–93. https://doi.org/10.2307/2296434.

———. (1970) 2017. *Collective Choice and Social Welfare:* An expanded edition. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674974616.