Are Choices Binary?

BRIAN WEATHERSON

University of Michigan

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, often implicit, 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 Section 15, while choiceworthiness is an ex ante notion. And ex ante norms are not grounded in ex post ones.

1. 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). Standard presentations treat norms on choice as almost an afterthought. 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 theorize with and about choice because, unlike preferences or values, it was observable. 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. Duncan 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 inevitably draw 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 key assumption is that choiceworthiness ascriptions involve the same kind of mental state as preference ascriptions. The biggest difference is that choiceworthiness relates to sets of any size, while preference relates to pairs.

2. 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 sometimes say an outcome has value 17 *utils*, but 'util' is a placeholder, not a real unit like kilograms or volts. This is related to the next reason.

Second, 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. This transformation is allowed because values merely reflect Chooser's preferences over outcomes and lotteries, which the transformation preserves. 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, treating values as numbers creates problems in cases involving 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

---

1. See, for instance, Hansson and Grüne-Yanoff (2024), where I learned about the Gul and Pesendorfer quote.
2. See, e.g., Nover and Hàjek (2004), or Goodman and Lederman (2024).

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. Value appears to be inherently comparative, reflecting how options compare to their alternatives.

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 of grounding values and preferences. Instead, that task will have to be done by judgments of, and norms about, choiceworthiness.

3. Coherence

There is one other striking thing about the picture we get in von Neumann and Morgenstern ([1944](#)), which I think is still broadly endorsed by 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 seems circular to ground normative facts about preference in other preferences. It looks like the grounding relation will be at least cyclic, and possibly intransitive.

There are two (related) responses to this circularity. One is to take the view, which perhaps Hume held, that the only constraints on preference are coherence constraints. Another response allows for non-coherence constraints on preference—for example, it may be wrong to prefer the world's destruction to one's finger getting scratched—but treats these as part of a separate subject from 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 criticize 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 from those 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 central question shifts. Instead of asking whether preferences should satisfy transitivity or independence, we ask whether choiceworthiness judgments should satisfy conditions like those described in Section 5. 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.

4. 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 them. 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, they are defined in terms of R in (1) and (2). (Throughout, I'm leaving off wide scope universal quantifiers over free variables.)

(1) $xPy \leftrightarrow (xRy \wedge \neg yRx)$
(2) $xIy \leftrightarrow (xRy \wedge yRx)$

It is important not to conflate indifference with 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 $\succsim$ to denote it, since that is most naturally read as that x is better than or equal to y.[3]

Contemporary philosophy more commonly starts 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) is plausible.

(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

---

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

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.

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 $x \succsim y$ to just mean the disjunction $xPy \lor xEy$, and assume that E is symmetric, then (4) is equivalent to $x \succsim y \lor y \succsim x$. 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 $xRy \lor yRx$. 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., $x \succsim y \lor y \succsim x$, less so on $xRy \lor yRx$. But the latter is useful because various interesting possibilities open up both in social choice theory and in the relationship between preference and choice, if it is dropped.[6]

While philosophers widely discuss (4), the economics literature discusses it less often under that formulation. 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 that $xRy \lor yRx$, so (4) and (6) are equivalent. But the different focus leads to more terminological confusion.

More generally, using 'completeness' for either $x \succsim y \lor y \succsim x$ or $xRy \lor yRx$ is potentially confusing, since readers may not know which property the author intends. So I'll follow Chang (2017) and say that relations satisfying (4) are *trichotomous*, and use *definedness* for $xRy \lor yRx$.

5. Properties of Choice Functions

In philosophy, we're sufficiently familiar with properties of preference relations (transitivity, reflexivity, acyclicity, etc.) that these terms don't need defining. 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 present six important properties that choice functions may have. 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).

---

5. Proof: Assume (4) is false. So the right hand side of (3) is false. Without much loss of generality, assume that $xRz \land \neg yRz$; the other cases all go much the same way. So all the disjuncts are false. From $\neg xPy$ and $\neg yPx$ we get $yRx \land xRy$, 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 disjuncts of (4) are false. That is, if (6) fails, so does (4), completing the proof that they are equivalent.

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

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

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

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

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 both choiceworthy together, then anything better than x is also better than y.

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

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 $(x \in C(T) \land y \in C(T) \land T \subseteq S) \to (\{y\} \neq C(S))$

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 y alone is choiceworthy. If x is not choiceworthy in the larger set, that's because some other option, not y, is chosen in place of it.

Property Aiz $(C(S) \subseteq T \land T \subseteq S) \to C(T) \subseteq C(S)$

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 choiceworthy options from a union of two sets equal the choiceworthy options from the union of each set's choiceworthy members. This is a kind of independence of irrelevant alternatives principle; the presence or absence 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.

6. Property α

This is the most commonly used constraint on choice functions, and it seems 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 having tea with a particular friend or staying home, Chooser will choose tea. But if the option of taking 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 choose zero. But if two slices of cake is an option, Chooser will choose 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'll set these individuation issues aside and assume, following most theorists, that α constrains coherent choice functions and that choice functions are defined over arbitrary sets of options.

7. Assumptions

I've said a few times I'm assuming this or that, so it's a good time to put in one place the assumptions I'm making. These aren't intended to stack the deck in my favour; if any of these assumptions are false, their falsity makes the view that choices are not binary (a) more plausible, but (b) harder to state. Anyway, here's what has been assumed.

(i) P is transitive, i.e., $xPy \land yPz \to xPz$
(ii) R is 'defined', i.e., $xRy \lor yRx$.
(iii) R is reflexive, i.e., xRx.
(iv) C is non-empty, i.e., $C(S) \neq \emptyset$.
(v) C is defined everywhere, i.e., there is a universe of options U and all subsets of U are in the domain of C.
(vi) C satisfies α.
(vii) The universe U of options, that S is a subset of, and x is drawn from, is finite.

In Section 6 we saw one reason to reject (v), namely that we might want to individuate options in terms of what else is available. The cases we'll discuss in Section 16 provide another, but rather than explore that, I'll follow standard practice and assume (v) throughout this paper.

When R satisfies (i)-(iii), I'll follow Luce (1956) and call it a *semiorder*. When it also satisfies trichotomy, i.e., (4), I'll call it a *weak order*.

8. Defining Binariness

With these seven assumptions on board, it's easy to state what it is for a choice function to be binary.

First, we'll define an inversion function B (for binary) that maps preference relations to choice functions, and vice-versa. Both are sets of ordered pairs, which we'll define directly. I'll assume that

---

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.

there is a universe U of options and every option and set of options is drawn from it.

If the input to B is a preference relation R:

(7) $B(R) = \{\langle S, x \rangle \colon \forall y (y \in S \rightarrow xRy)\}$

That is, $B(R)$ is the choice function that, for any set S, selects what Sen calls 'maximal' members—those members to which nothing is strictly preferred.[8]

If the input to B is a choice function C:

(8) $B(C) = \{\langle x, y \rangle \colon x \in C(\{x, y\})\}$

That is, $B(C)$ is the preference relation stating that in pairwise choices, an element is preferred only if it could be chosen from the pair. Sen ([1970] 2017, 319) calls these functions 'basic binary', but the distinction marked by 'basic' doesn't matter given that R is a semiorder and α holds. Since we're assuming α, we'll use the simpler version.

A choice function C is **binary** iff (9) holds:[9]

(9) $C = B(B(C))$

Converting C into a preference relation and back yields the same function. This will fail if the part of C which concerns choice from menus with three or more members contains extra information than just the restriction of C to pair sets.

The core claim of this paper is that there are coherent choice functions which are not binary. A related claim is that a plausible pair of coherence constraints that you can state using B do not in fact hold. The constraints require that $C = B(R)$ and $R = B(C)$.

9. Property β

If we start with choice functions, the definition of E in (3) is too simple. A better definition is in (10).

(10) $xEy \leftrightarrow [\forall S (\{x, y\} \subseteq S \rightarrow (x \in C(S) \leftrightarrow y \in C(S)))]$

That is, x and y are equal iff one is never chosen when the other is not.[10] Given this notion of equality, there is an intuitive gloss on β: Two options are both choiceworthy iff they are equal.[11]

If choice functions are grounded in numerical values, then β follows naturally. Assume there is some function v from options to numbers, and o is choiceworthy iff $v(o)$ is maximal. Then if x and y are both choiceworthy, they must have the same value. In any set, they will either both be choiceworthy (if no alternative is more valuable) or neither will be (if some alternative is more valuable).

---

8. Hansson (2009) calls this the 'liberal maximisation' rule. He contrasts it with five other rules, which are distinct in general but equivalent given R is a semiorder.

9. Sen calls these functions 'basic binary', but the distinction he's drawing attention to by adding 'basic' doesn't make a difference given R is a semiorder and α.

10. Without α, this is too weak, since it doesn't entail that x and y are intersubstitutable in general. But we won't worry about that.

11. This gloss also assumes α.

More generally, given the assumptions from Section 7, C satisfies β iff $B(C)$ is trichotomous. This is equivalent to $B(C)$ satisfying II-transitivity. Unsurprisingly, the two historically significant cases of intuitive counterexamples to II-transitivity also generate intuitive counterexamples to β.

The first example involves distinct but indistinguishable options.[12] Assume that Chooser prefers more sugar in their coffee to less, but can only tell two options apart if they differ by 10 grains of sugar or more. Now consider these three options:

x = Coffee with 100 grains of sugar.
y = Coffee with 106 grains of sugar.
z = Coffee with 112 grains of sugar.

This is said to be a counterexample to II-transitivity because Chooser is indifferent between x and y, and between y and z, but strictly prefers z to x. It's also a counterexample to β. Chooser would choose either from x and y, but when z is added, y is still choiceworthy but x is not.

This example was historically important, but it's rarely discussed in the contemporary philosophical literature. It could be because philosophers were convinced by the argument by Delia Graff Fara (2001) that phenomenal indistinguishability is in fact transitive. But it was widely discussed in economics, especially after Luce (1956, 1959) used similar examples to argue that preferences form a semiorder

The cases that were more important in the philosophical literature are what Chang (1997) calls 'small improvement' cases. The earliest example of this form I know is from Luce and Raiffa (1957).[13] In their notation, $P(x,y)$ represents the probability that Chooser will select x when x and y are both available.

> Suppose that a and b are two alternatives of roughly comparable value to some person, e.g., trips from New York City to Paris and to Rome. Let c be alternative a plus \$20 and d be alternative b plus \$20. Clearly, in general $P(a, c) = 0$ and $P(b, d) = 0$. It also seems perfectly plausible that for some people $P(b, c) > 0$ and $P(a, d) > 0$, in which event a and b are not comparable, and so axiom 2 [i.e., (4)] is violated. (Luce and Raiffa 1957, 375)

An example with the same structure, involving a boy, a bicycle, and a bell, is discussed by Lehrer and Wagner (1985), and mistakenly attributed to Armstrong (1939).[14]

The usual way these cases are discussed, starting with Luce and Raiffa, is that they violate a certain kind of comparability. For example, Luce and Raiffa say there is a sense in which the two holidays are 'not comparable'. I want to resist this reading. The core intuition in small improvement cases is that β fails. Chooser would choose either option from $\{a, b\}$, but if a^+ is added as an option, a becomes unchoiceworthy. If we add the assumption that $R = B(C)$, then it does follow that trichotomy fails, and there is a sense in which they are incomparable. But without that assumption,

---

12. The idea that humans can't distinguish similar options is important in Fechner (1860), a work which is discussed in Beiser (2024). The earliest connection I've found between this and indifference being intransitive is in Armstrong (1939). Armstrong's example is rather confusing; the one I'll use here is based on Luce (1956).

13. I haven't found a case like this in Luce's sole-authored works, and indeed Debreu (1960) notes that a related case raises problems for one of the central assumptions of Luce (1959).

14. Many authors subsequently made the same attribution; if you want to see some examples, search for the word 'bicycle' among the citations of Armstrong's paper on Google Scholar.

it's consistent to say that these are counterexamples to β but not to trichotomy. We'll return to this point in Section 18.

10. Properties γ and δ

Assume R does not satisfy trichotomy, but is a semiorder, and $C = B(R)$. Then β will fail, but γ and δ will hold. Conversely, for any C where γ and δ hold, there is a semiorder R such that $C = B(R)$ (Sen [1970] 2017, 320). We're not going to be very interested in δ, but we will be very interested in γ.

The reason γ holds when R is a semiorder and $C = B(R)$ is instructive. If x is choiceworthy among S, then nothing in S is better than x. Similarly, if x is choiceworthy among T, then nothing in T is better than x. So nothing in $S \cup T$ is better than x. So x is choiceworthy among $S \cup T$.

Conversely, if there are cases where C should not satisfy γ, then we'll have an argument that C should not be grounded in some semiorder R. Showing that there are such cases will be one of the main tasks of the rest of this paper.

We had two kinds of counterexamples to β, but only one of them will be relevant here. I don't think there are any intuitive counterexamples to γ that start with Fechner-style reflections on the intransitivity of indifference. But there are going to be variations on the bicycle and bell example that generate intuitive counterexamples to γ. We'll come back to these in Section 19.

It is common to say that when $C = B(R)$ for some semiorder R, C is **rationalizable**, and when R is a partial order, C is **rationalizable by a partial order**. I find this terminology tendentious - why should semiorders be the only things that can make C rational? And as we'll see in Section 13, it conflicts with the notion of a choice being *rationalizable* in game theory. But it's a common enough terminology that I wanted to mention it here.

11. Aizerman's Property

The property Aiz is not particularly intuitive. Fortunately, it turns out to be equivalent, given our assumptions, to one that is: Path Independence.[15] Path Independence says that to find what's choiceworthy from a union of sets, you only have to consider which options are choiceworthy in the smaller sets.

Note that this isn't just saying that the options choiceworthy from the union are choiceworthy from one of the members. That is implied by α. What it says is that the presence or absence of unchosen options from S and T doesn't affect which options are choiceworthy from the union.

There is a very natural kind of model where α and Path Independence hold, but β and γ do not. Let O be a set of total orderings of U. (A total ordering is a relation R such that $xRy \lor yRx \lor x=y$.) Further, let $C(S)$ be the set of Pareto optimal members of S relative to those orderings. That is, it is the set of members of S such that no other member of S is better according to every member of the set of orderings. Every such model satisfies α and Path Independence, but some violate β and γ.

---

15. Unless stated otherwise, the results in this section, along with proofs, can be found in the helpful survey by Moulin (1985).

There is a further relation between Path Independence and these models. For any C that satisfies α and Path Independence, there is some set of total orderings over S such that $C(S)$ is the set of Pareto optimal members of S according to that set of orderings.

It might seem strange after all the talk of weak orderings and semiorders that we're now using total orders. Given that U is finite, this follows naturally from well-known properties of finite semiorders. Any semiorder with a finite domain (and hence any such weak order) can be represented by a set of total orders: x is strictly preferred to y in the semiorder iff it is preferred in all the total orders. (In fact we can put a sharp cap on how many such orders there must be.) So being Pareto optimal relative to some set of semiorders (or weak orders) is equivalent to being Pareto optimal relative to a larger set of total orders.

If C is determined by a set of orders in this way, it is said to be **pseudorationalizable**. These choice functions are not always binary. Consider one simple example, where U is $\{x, y, z\}$, $C(U) = \{x, z\}$, and for any other S, $C(S) = S$. That is the choice function determined by the pair of orderings $x \succ y \succ z$ and $z \succ y \succ x$. This satisfies α, δ and Aiz, but not β or γ. And it isn't binary. $B(C)$ is the universal relation R, since whenever S is a pair set, $C(S) = S$.

The approach discussed here is obviously similar to what is called the "multi-utility" approach to representing incomplete preferences (Evren and Ok (2011)). In this approach we find a set of utility functions and say xRy iff $u(x) \succsim u(y)$ for all u in the set.

I'm going to argue over subsequent sections that this is a coherent choice function, and hence not all coherent choice functions are binary.

12. Preference and Trade

Let's go back to why we might have wanted to focus on binary preference relations. One reason is that preference might naturally explain trade. If Chooser trades a cow for some magic beans, it's natural to explain that by saying Chooser preferred the magic beans to the cow.

In the monetary economies we live in, very little trade involves barter. Most trade involves money, which is primarily valuable instrumentally. If Chooser buys some shoes for $100, we could say that Chooser prefers the shoes to the money. But that doesn't seem like the full story, since there is a reason Chooser values the money as they do. The deeper point is that Chooser's money creates a budget constraint, and Chooser judges the shoes to be the best use of that $100 among available options. That is, it seems more informative to describe Chooser as choosing the shoes from the menu of things that cost $100 than to describe them as preferring the shoes to the money. In general, choiceworthiness seems more relevant to explaining market behaviour in monetary economies than preference, unless choiceworthiness can be defined in terms of preference. So let's turn to some reasons to think that it cannot.

13. Degenerate Games

Say a two-player game is **degenerate** iff the payoffs to one of the players are constant in all outcomes of the game. For convenience, assume Column is the player with constant payoffs. So Table 1,

Table 2 and Table 3 are degenerate games.

Table 1: A degenerate two-option game

	Left	Right
Up	10,0	0,0
Middle	1,0	1,0

Table 2: Another degenerate two-option game

	Left	Right
Middle	1,0	1,0
Down	0,0	10,0

Table 3: A degenerate three-option game

	Left	Right
Up	10,0	0,0
Middle	1,0	1,0
Down	0,0	10,0

Start with the following two assumptions, which seem fairly plausible for games like these.

- If a move is part of a Nash equilibrium, it is choiceworthy.
- A move is choiceworthy only if there is some probability distribution over the other player's moves such that the move maximizes expected utility given that distribution.[16]

In degenerate games, these necessary and sufficient conditions for choiceworthiness coincide, but in general they are rather different.[17]

In Table 1 and Table 2, both options are choiceworthy by this standard. The Nash equilibria of Table 1 are Up-Left and Middle-Right, while the Nash equilibria of Table 2 are Middle-Left and Down-Right. So each option is choiceworthy for each player in each game. But in Table 3, the only choiceworthy options for Row are Up and Down. Whatever probability Row assigns to Left or Right, Middle will not maximize expected utility. So this is a counterexample to γ. Middle is choiceworthy from {Up, Middle} and from {Middle, Down}, but not from their union.

—————————

16. The notion of a rationalizable choice, in the sense of B. Douglas Bernheim (1984) and David Pearce (1984), slightly strengthens this constraint. A choice is rationalizable iff it maximizes expected utility relative to a probability assignment that only gives positive probability to the other player, or players, making rationalizable choices. That's circular as stated, but one can remove the circularity at the cost of making the definition somewhat less intuitive.

17. Since rationalizability is between these notions, it also coincides with them for degenerate games.

It follows immediately from Lemma 3 in Pearce (1984) that in degenerate games, a choice satisfies those conditions for being choiceworthy iff it is not strictly dominated, where this includes being dominated by mixed strategies. (In Table 3, Middle is strictly dominated by the 50/50 mixture of Up and Down.) This means that the choices will satisfy Path Independence. An option is not dominated by the options in $S \cup T$ iff it is not dominated by the undominated options in $S \cup T$, i.e., by the options in $C(S) \cup C(T)$. So removing options that are not choiceworthy in S and T from the union doesn't change what is undominated, i.e., choiceworthy.

This last paragraph is the start of a pattern in the examples that follow. Although I'll be arguing against γ, I won't be arguing against Aiz/Path Independence. I'm not going to offer anything like a conclusive argument for Aiz, but the pattern suggests adding Aiz to α yields the fundamental constraints on coherent choiceworthiness.

14. Choice Under Uncertainty

Luce and Raiffa (1957) discuss what they call choices under 'uncertainty', by which they mean choices where Chooser cannot assign probabilities to the states. Martin Peterson (2017) calls these choices under 'ignorance'. None of the proposed decisive rules for choice under uncertainty/ignorance are particularly compelling; all lead to very strange outcomes.

The best approach, in my opinion, is to treat these choices like degenerate games. Indeed, degenerate games are really a paradigm of choice under ignorance; Row has no reason to assign any particular probability to Column's choice. Further, what the game theory textbooks say about degenerate games seems fairly plausible; any undominated option is choiceworthy. The same goes for choices under ignorance; any undominated option is choiceworthy.

If that's right, then the three examples from Section 13 can be repurposed as examples of choice under ignorance, replacing Left and Right with p and $\neg p$, and the same analysis will hold. Again γ will fail, because Middle is choiceworthy when there is one other option, but not when there are two. So there's no binary comparison of Middle with the other two options that explains the facts about what is choiceworthy in the three cases.

15. Multiple Equilibria

This is a decision theory paper, so we need to introduce a demon who can reliably predict Chooser's choices. We'll start with a version of what Skyrms (1982) calls 'Nice Demon'; we'll just call them Demon. In Table 4, Chooser selects Up or Down, and Demon either predicts Up (PU), or predicts Down (PD). Whatever Chooser does, Demon is very likely to have predicted correctly.

Table 4: First version of Nice Demon

	PU	PD
Up	6	0
Down	0	4

Jack Spencer (2023) argues against views, like the one defended by Dmitri Gallow (2020), which say only Up is choiceworthy in Table 4.[18] Spencer's argument relies on a simple principle. If Chooser plans to play Down, then Chooser knows Down will have the best return, and it's not irrational to make a choice if one knows, when one makes it, that the choice will have the best return. Gallow (2024) replies that in some cases, especially those involving high stakes, it isn't always rational to do what one knows will produce the best return. Since I think it's a fundamental feature of knowledge that it does rationalize action in the way Spencer suggests (Weatherson (2024)), I think Spencer has the better of this exchange. So I'll follow him and assume that both options are choiceworthy in games like Table 4 where multiple options are self-verifying.[19]

Now add a third option, Exit, which has a guaranteed return of 1. So the game table looks like Table 5.

Table 5: Second version of Nice Demon

	PU	PD
Up	6	0
Exit	1	1
Down	0	4

In Table 5, Exit is not choiceworthy. Whatever credences Chooser has about what Demon has done, it is better in expectation to choose one of Up or Down. But if either Up or Down were unavailable, Exit would be a permissible choice. This follows from the earlier assumption that any self-verifying choice in cases like Table 4 is choiceworthy. So like in Section 13 and Section 14, we have a counterexample to γ.

There is an important general lesson from this case. What makes an option choiceworthy in cases like this is that it is utility maximizing *once it is chosen*. We'll turn next to a more dramatic illustration of this point.

16. Mixed Strategies

Demon has stopped being nice, and now wants to play Rock-Paper-Scissors with Chooser. Given Demon's powers, this could go badly for Chooser. Happily, Chooser can choose to randomize[20], and all Demon can predict is the probability that Chooser's random process will come up with any

---

18. Evidential decision theory also says that only Up is choiceworthy in Table 4. What's distinctive about Gallow's position is that it says only Up is choiceworthy here, although only two-boxing is choiceworthy in the original Newcomb Problem.

19. The version of causal decision theory set out by David Lewis (1981) has a more complicated verdict on Table 4. It says the uniquely rational choice is whatever choice Chooser believes they will make. This isn't the view I'm assuming, since I take both options to be permissible. But it's very close to the picture I present, without endorsing, in Section 18. On that view, once Chooser commits to a choice, it becomes the only permissible one, so the view isn't far from Lewis's.

20. I'm not going to argue for this here, but I think it is part of being ideally practically rational that one is able to randomize, just like it is part of being ideally practically rational that one can make calculations costlessly.

option. So the best thing for Chooser to do is to pick one of the three options at random. That alone will maximize expected utility conditional on being chosen.

Not coincidentally, the only Nash equilibrium of Rock-Paper-Scissors is for each player to randomize. Nash equilibria are arguably the only sensible strategies if one assumes that every other player has Demon-like abilities to detect what one is doing. But it's a long-running puzzle in game theory how it can be uniquely rational to randomize. Why can choosing randomly be better than choosing one of the things one is randomizing between? To turn this rhetorical question into an argument, note that the following three principles are inconsistent.

(11) randomizing is the only choiceworthy strategy in Rock-Paper-Scissors.

(12) If only one choice is choiceworthy, it is rationally preferred to all other choices.

(13) It is irrational to prefer a random mixture of some choices to every one of the choices.

Since (11) is a standard view in game theory, (12) is a standard view in choice theory, and (13) is a standard view in preference-based decision theory, it is a little disconcerting to see they are inconsistent.

The example in Section 15 shows how to steer through this trilemma. Choiceworthiness is fundamentally an ex ante notion, and preference is fundamentally an ex post notion. The reason Spencer's view about Table 4 is right is not that the rational chooser is indifferent between Up and Down. Rather, choosers don't have preferences between these choices until they have chosen, and once they choose, they prefer their choice.

Similarly in Rock-Paper-Scissors (especially against a Demon), what's true is that prior to deciding, the only rational choice is to randomize. Before the choice, there is no such thing as Chooser's preferences over the options; what I earlier called definedness fails. Once one has chosen, one should be strictly indifferent between the options; they each have the same expected utility. In particular, one should not prefer to re-randomize rather than put into effect the result of the random process.

So we should reject (12) in its most natural interpretation. randomizing in Rock-Paper-Scissors is the only choiceworthy option, but until a choice is made, Chooser simply shouldn't have preferences over these options.

Most of the arguments in this paper against the binariness of choice turn on counterexamples to γ, but this is a distinct argument. Sometimes, as in Rock-Paper-Scissors, there are grounds for rational choice, but no grounds for rational preference. The only preferences that would ground the choice would violate (13). So rational choiceworthiness cannot be grounded in rational preference.

This is the deepest reason why $C = B(R)$ must be wrong; C and R are fundamentally about different kinds of attitudes. C is about what is rational prior to making a choice, R is a constraint that must be satisfied once a choice is made. Outside of Newcomb-like cases, this distinction won't often matter, but it is another reason that the equation fails.

17. Multiple Attributes and Decisiveness

Sartre (1946/2007) has a famous example of a young man (we'll call him Pierre) caught between two imperatives. The actual example Sartre gives is complicated in interesting ways, but we'll work

with a very simple version of the story. Pierre lives in occupied France during WWII, and feels torn between his duty to care for his ailing mother, and his duty to fight for his occupied country. What should he do?

The case is extremely underdescribed, but the following verdicts have seemed plausible to many people. First, Pierre can rationally, and morally, choose either option. Caring for his mother and fighting Nazis are both noble goals, and it's fine to pursue either. Second, and this follows from the first and the fact that the case is underdescribed, the options are not equally good. After all, a small improvement to either would not break a tie between them.

A third intuition is more contentious, but perhaps plausible: there is something wrong about Pierre going back and forth between the two choices; he should make a choice and stick to it. On this view, there is something intrinsically good about settling on a choice and sticking to it. What makes this intuition less than fully clear is that oscillation would be practically bad in any realistic version of the case. He could spend the whole war travelling between England and France as he changes his mind on where he should be, and that would be bad. The intuition I have in mind is that there is something good about taking a stance and committing to it, even outside of the practical costs of changing one's mind.[21]

A very simple model for these intuitions is that Pierre's situation is surprisingly like the person playing Table 4. There are two options here, and either is acceptable, but once one is chosen, it becomes the preferred one. There are two good values here, caring for family and caring for country, and Pierre's fundamental choice is to adopt one of these as his value. As Chang (2024) puts it, he chooses to put 'his whole self' behind one of those values. Once that choice is made, his preferences and his actions follow naturally.

It is easy enough to reject the third intuition. Perhaps Pierre could rationally, as Chang (2024) puts it, *drift* into one choice. Perhaps he could adopt one path and sometime later rationally regret his choice, because the other value strikes his later self as more important. Still, if one holds all three intuitions, toy models like Table 4 capture a surprising amount of what's going on with Pierre. And those models are incompatible with choiceworthiness being binary.

18. β and Incompleteness

The Pierre case in Section 17 is, even without the third intuition, a straightforward counterexample to β. Just to spell it out,

    x = Help mother
    y = Fight Nazis
    z = Help mother plus one extra ration book

If the choices are *x* and *y*, either is acceptable. If the choices are *x*, *y* and *z*, *x* alone is unchoiceworthy. So β fails.

---

21. As Moss (2015) points out, it is less clear this is intrinsically bad if there is more time between the reconsiderations; it makes more sense to change one's mind than to rapidly flip-flop. Conversely, if Pierre resembles the young Thomas Schelling (as discussed by Holton (1999)), firmly committing to one plan and then another over the course of successive nights, he's doing something wrong even if it has no practical downsides.

This is the Small Improvement argument, and it is often thought to be an argument against completeness, i.e., (4). The point of this section is to explain why the argument against (4) might fail, even though the argument against β succeeds.

Imagine someone were convinced by the arguments in Dorr, Nebel, and Zuehl (2023) that (4) must be true. Their arguments turn on semantic properties of comparatives; they claim that since R is a comparative, it must be trichotomous, as all comparatives are.[22] Now it would be very strange if the semantics of comparatives in natural language entailed that some intuitions about three-way choice had to be false. And in fact these claims about semantics are consistent with β failing in cases like Pierre's.[23]

The following view is coherent, and its coherence shows that Dorr et al.'s view of comparatives does not entail β.

- β fails in cases like Pierre's.
- An option is choiceworthy for Pierre iff no option is *determinately* preferred to it.
- To make a choice, Pierre must determine which value is really his.
- Once he does that, his preferences will satisfy (4).
- Before he does that, there are two possibilities. One is that his preferences aren't even defined over these options, so asking which option is preferred is like asking whether the number 7 is taller, shorter, or the same height as justice. Another is that it is vague what Pierre's preferences are, but any resolution of the vagueness makes (4) true. This latter option fits nicely with the idea that C should satisfy Aiz, since it should be determined by a set of orderings, each of them the possible precisifications, or determinations, of his current state.[24]

One objection to this view is that it seems to imply that Pierre could rationally choose one option while he prefers, but does not determinately prefer, another. But this isn't what the view implies. Once Pierre makes the choice, he must, if he is rational, determine that his preferences match it. Preference, as I argued in Section 16, is fundamentally an ex post notion. That is, (14) is true, while (15) is false.

(14) Once Pierre chooses an option, there must not be some other option he strictly prefers to it.
(15) Before an option is choiceworthy for Pierre, it must be better or just as good (given his preferences) as any other option.

What (14) says is that if Pierre is rational, he will choose what he prefers. That does not mean, as (15) claims, that rational choices are those that are preferred prior to the choice. Choice, on this picture, is prior to preference, both analytically and, in this case, causally.

When I say this is coherent, I don't mean to half-heartedly say that it is correct. My preferred view is that Pierre could rationally drift (in Chang's sense) into either option, and if he does, (4) would fail even ex post. All I mean to argue here is that the case against β doesn't turn on this view,

---

22. For what it's worth, I think this argument fails because of the case of 'stronger' in logic. They address this case, but I don't think their response works. But it would be a huge digression to follow that thread through.

23. I'm focussing here on the argument in Dorr, Nebel, and Zuehl (2023), but a similar response works if someone is convinced of (4) by the argument in Broome (1997).

24. This view of preference mirrors the view of credence defended by Carr (2020).

and one can coherently reject β while endorsing trichotomy.

19. Bad Compromises

My version of the Pierre example was very simple, but it allows for some interesting complications. As stated, you might think Pierre isn't thinking through his choices well enough. He should join the local resistance, so he can stay close enough to his mother to help, while also fighting the Nazis.

But maybe that's a terrible option. We can easily imagine that the resistance is either so useless that it does practically nothing, or so good at recruiting that it has little useful work. It's just as easy to imagine that it creates busywork that dramatically reduces how much he can care for his mother, while not doing much to help the war effort. At risk of trivialising the issue, we can imagine that Pierre's options look like this, where the two columns represent how much each option respects/promotes the relevant value.

Table 6: Pierre's options, including resistance

	Caring value	Fighting value
Stay Home	10	0
Join resistance	1	1
Join Free French	0	10

Intuitively, if these are Pierre's options he should not join the resistance; it's almost the worst of both worlds. We don't want to rest on an intuition, and I'll argue for the view that Pierre should shun the resistance in the next two sections. If it is irrational to join the resistance in Table 6, this is another counterexample to γ. To see this, consider these two variants of the case.

> **No exit**: While Pierre is deliberating, he hears that the options for getting to the Free French have been decisively cut off. (In the original, he worries this might happen.) Now his only two options are to stay home, or to join the resistance.

> **Promise**: Pierre's brother Jean is fighting the Nazis. Pierre has promised Jean that if Jean is killed, Pierre will take up the fight in some way. Sadly, Jean is killed, and Pierre regards this promise as binding. Now his only options are to join the resistance, or join the Free French.

In either case, joining the resistance seems choiceworthy. In both cases, it is the option that does best of the remaining choices on one of the criteria. Pierre could decide he endorses that criterion as his own, and acts accordingly. So the resistance is choiceworthy amongst either pair of options, but not amongst their union.

20. Levi and Sen

In *Hard Choices*, Isaac Levi (1986) defended a view where the choiceworthy options are only those that maximize value on some resolution of the incompleteness in the agent's values. (He doesn't use this language, but so far this is similar to the multi-utility approach discussed in Section 11.) Levi also had views about what further constraints there should be on choice, so he did not defend the view I've been discussing where any option that is maximal on any resolution is choiceworthy. But his views are still relevant here, because this requirement that a choice be maximal on a resolution meant that he was committed to γ failing, and choices not being binary.

A central example which he uses, and which Sen (2004) picks up on, involves an executive looking to hire a secretary. I'll follow Sen and call the executive Ms. Jones.[25] She is looking for a secretary with good typing skills and good stenography skills. (This is the 1980s.) We'll conceive of these skills, a bit arbitrarily, as distinct values. There are three candidates: Jack, Danny, and Luke, and their value on each measure is given in Table 7. (I've slightly adjusted the numbers to match the earlier examples.)

Table 7: Three candidates for a secretarial position

	Typing Skill	Stenography Skill
Jack	10	0
Danny	1	1
Luke	0	10

Levi argues that if the numbers are like this, and Danny is barely better than the worst of the other two on each metric, he is not choiceworthy. That's true even though he might be choiceworthy if one or the other candidates was unavailable.

Sen argues that the right choice theory in this case should be "inarticulate", and say that any of the three is choiceworthy. He responds to the intuition Levi presents with a dilemma.

On the first horn, we understand these numbers as representing an objective measure of the skills of the candidates at each of the tasks. As Sen points out, it's easy to imagine situations where someone who is not abysmal at either half of the job is more useful than someone who is an expert on one, and abysmal on the other.

On the other horn, we measure the "importance" (Sen 2004, 53) of each skill for the task at hand. Sen argues, by analogy with the difficulty in establishing a social welfare function out of the welfare of each individual, that there will be no way to do this. Let's turn to how we might go about it.

---

25. I've also changed the secretaries' names.

21. Lotteries, Choices, and Values

In this section I'll offer a response, on Levi's behalf, to Sen's dilemma. The example will draw heavily on recent work by Harvey Lederman (forthcoming).[26] We'll start by imagining that Ms. Jones might have a choice not of secretaries, but of agencies, and she has a reasonable credal distribution over the skills of the people a particular agency might send. That is, each choice of agency will be a choice of a *lottery*, where she doesn't choose a package of skills, but a probability distribution over some outcomes, where each outcome is a secretary with a numerical skill on each attribute.

To set this up, I'll need three new bits of notation. I'll write *Lxy* for a lottery that has equal chance of returning outcomes *x* and *y*, where these might be secretaries or further lotteries. I'll write $x = y$ for what I've previously written as *xEy*. It means that *x* and *y* are equally good by Ms. Jones's lights. It's perhaps suboptimal to introduce new notation for an old concept, but the mix of L's and E's became hard to read. Finally, I'll write $\langle x, y \rangle$ for a secretary with skill *x* at typing and skill *y* at stenography.

If the numbers in Table 7 measure importance, then Ms. Jones's preferences should satisfy (a special case of) what Lederman calls **Unidimensional Expectations**.

Unidimensional Expectations (UE)

$$L\langle x_1, y \rangle \langle x_2, y \rangle = \langle (x_1 + x_2)/2, y \rangle$$
$$L\langle x, y_1 \rangle \langle x, y_2 \rangle = \langle x, (y_1 + y_2)/2 \rangle$$

That is, the value of a lottery where the possible outcomes agree on one dimension also has the same value on that dimension, and has the expected value of the other dimension. If the lottery does not involve value conflict, old-fashioned expected value maximisation is the way to go.

This is enough to rule out the example Sen has in mind on the first horn, where a secretary with skill 1 on either dimension is more than half as valuable as a secretary with skill 10. But it's not enough to say that Ms. Jones should not hire Danny. Unidimensional Expectations is consistent with one resolution of his indeterminate value being that the value of $\langle x,y \rangle$ is *xy*. If that's a permissible resolution, then there will be a resolution on which Danny is maximally valuable. But there are further principles that do rule out Danny. The most intuitive argument I know uses the following.

Substitution of Identicals (SI) If $x = y$, then $Lxz = Lyz$ and $Lzx = Lzy$.
No Trade-Off (NT) $L\langle x, x \rangle \langle y, y \rangle = \langle (x + y)/2,(x + y)/2 \rangle$
Rearrangement of Outcomes (RO) $L(Lxy)(Lvw) = L(Lxw)(Lvy)$
Weak Independence (WI) If $x = Lxy$ then $x = y$

Substitution of Identicals follows naturally from the idea that the outcomes are truly equal, so it doesn't matter whether a lottery has an outcome that results in one or the other. No Trade-Off, like Unidimensional Expectations, says that when we are just considering strictly better and worse options, so there are no relevant complications about resolving indeterminacy, we're back in the

---

26. See also Lederman (2023), and Tarsney, Lederman, and Spears (forthcoming). This is not to say Lederman would endorse anything like this response; as we'll see in the next section he is sympathetic to Sen's view that the multi-utility approach is mistaken. But every principle I'll use here is discussed, one way or another, in his work, and I've drawn heavily on that discussion in what follows.

land of expected utility maximisation. Rearrangement of Outcomes follows from the fact that the compound lotteries on either side of the identity sign each have probability 1/4 of returning one of those four outcomes. And Weak Independence, which is probably the most contentious of the lot, says that if *y* is not exactly as good as *x*, then a 1/2 chance of *y* should not be exactly as good as *x*. Given these principles, we can argue as follows.

1.	$\langle 5, 5 \rangle = L\langle 10, 5 \rangle\langle 0,5 \rangle$	UD
2.	$\langle 10, 5 \rangle = L\langle 10, 10 \rangle\langle 10,0 \rangle$	UD
3.	$\langle 0, 5 \rangle = L\langle 0, 10 \rangle\langle 0,0 \rangle$	UD
4.	$\langle 5, 5 \rangle = L(L\langle 10, 10 \rangle\langle 10,0 \rangle)\langle 0,5 \rangle$	1,3 SI
5.	$\langle 5, 5 \rangle = L(L\langle 10, 10 \rangle\langle 10,0 \rangle)(L\langle 0, 10 \rangle\langle 0,0 \rangle)$	2, 4 SI
6.	$\langle 5, 5 \rangle = L(L\langle 10, 10 \rangle\langle 0,0 \rangle)(L\langle 0, 10 \rangle\langle 10,0 \rangle)$	5 RO
7.	$\langle 5, 5 \rangle = L\langle 10, 10 \rangle\langle 0,0 \rangle$	NT
8.	$\langle 5, 5 \rangle = L\langle 5, 5 \rangle(L\langle 0, 10 \rangle\langle 10,0 \rangle)$	6, 7 SI
9.	$\langle 5, 5 \rangle = L\langle 0, 10 \rangle\langle 10,0 \rangle$	8 WI

To complete the argument, we need two plausible principles. First, *x* is not choiceworthy from *S* when *x* is not choiceworthy from the set consisting of *x* and *Lyz*, for *y*, *z* in *S*. Second, *x* is not choiceworthy from *S* when an option that is strictly better on every dimension is in *S*.

There are several assumptions here, and any one of them could be the subject of a whole paper. But they each seem very plausible, and they show how we can ground the intuition that the multi-utility approach is on the right track, and sometimes an option is not choiceworthy because it is close to the worst option on every dimension we care about.

22. Negative Dominance

Harvey Lederman (forthcoming) notes that the picture I've developed, where Danny is unchoiceworthy, appears to violate a plausible principle which he calls **Negative Dominance**. Lederman gives a few versions of this principle, the following is the version most relevant to the case.[27]

Negative Dominance (Goodness) If one game of chance is better for [Ms. Jones] than another, some prize in the first game is better for her than some prize in the second. (Lederman forthcoming, 13)

The main application of this is to reject the idea that *L*(Jack)(Luke) is better than Danny.

I said the picture from previous sections 'appears to violate' Negative Dominance because one more assumption is needed to generate the tension. The assumption is that when Chooser has only two options, preference and choiceworthiness are closely related. In particular, *xPy* iff *C*({*x*, *y*}) = {*x*}. I've been arguing against principles linking preference and choice throughout this paper, but I've mostly accepted that they are tightly connected when there are only two options on the menu. Still, without some assumption linking choice and preference, the picture in the previous

---

27. In all the quotes I'll change the names and example to match the one I'm using.

sections implies nothing about preference, hence it says nothing that contradicts Negative Dominance. For now, let's assume xPy iff $C(\{x, y\}) = \{x\}$, and I'll return at the end to what happens without that assumption.

Negative Dominance seems like a very plausible principle if (and I think only if) one thinks that the role of decision theory is to come up with coherence constraints on *preferences*. Centering decision theory on preference follows naturally from the idea that values and choices are ultimately grounded in preferences. As Lederman argues, preferences about lotteries have to be grounded in something about their prizes, and if preferences are fundamental, presumably they have to be grounded in preferences over their prizes. The key response I'm making is that if choiceworthiness is prior to preference, this last step doesn't follow.

So when Lederman says,

> a strict preference for one game of chance over another must be explained by a strict preference for one of the *prizes* of the first, by comparison to one of the prizes of the second (Lederman forthcoming, 18, emphasis in original)

we should question the uses of 'one'. What's true is that attitudes towards games of chance should be somehow explained by attitudes towards the prizes, but these attitudes need not be *preferences*. For instance, the fact that L(Jack)(Luke) is better than Danny could be grounded in the fact that Danny is not choiceworthy from {Jack, Danny, Luke}.

But there are more complicated cases where Lederman's challenge of how to properly ground attitudes to lotteries is more pressing. The following case mixes Levi's case with the main example in Tarsney, Lederman, and Spears (forthcoming). Ms. Jones is now trying to hire a programmer, and she has four candidates, each of which has the skills in the four languages she cares about shown in Table 9.

Table 9: Four programmers, and their skills

	Python	Java	C	Ruby
Jane	6	6	0	0
Dolly	0	0	6	6
Lily	5	0	5	0
Suzy	0	5	0	5

When the menu consists of any set of the programmers, all the options are choiceworthy. But Ms. Jones strictly prefers L(Jane)(Dolly) to L(Lily)(Suzy), since the former lottery is better in expectation on all four dimensions. Lederman is right that this needs to be explained, that it should be explained in terms of evaluative features of the prizes (i.e., the programmers), and if the explanation uses expected value, we should explain why expected value matters. No explanation in terms of the choiceworthiness of some options will work to explain why one lottery is strictly better than another.

The fact to be explained is that when L(Jane)(Dolly) and L(Lily)(Suzy) are available, only the former is choiceworthy. Here's how I explain it:

1. Ms. Jones has four values, and it is indeterminate how they should be balanced. This means both that she hasn't decided how to balance them, and maybe it is unnecessary, or even inadvisable, to balance them.
2. Given Unidimensional Expectations and No Trade-Off (as discussed in Section 21), the only permissible balancings are linear mixtures of the values.
3. Given the result from Pearce (1984) discussed in Section 13 (his Lemma 3), a lottery is best on no linear resolution of the indecision in point 1 iff some available lottery over other choices is better in expectation on every value.
4. A lottery is choiceworthy from a menu of other lotteries (or options) iff it is optimal on some permissible resolution of this indecision.

If L(Lily)(Suzy) were choiceworthy, by 4 it would have to be best on some resolution of Ms. Jones's values. By 1 and 2, this means that it is best on some linear mixture of these values. By 3, that means it is better in expectation on one of these values. But it is not; on all four dimensions L(Jane)(Dolly) has expected value 3, and L(Lily)(Suzy) has expected value 2.5. Even though Ms. Jones has not resolved the indeterminacy in her values, the fact that any resolution would mean she prefers the first lottery is enough reason to prefer the first lottery.

In short, the focus on expected values comes not from any particular importance on expectations as such, but from the thought that permissible reactions to indeterminacy in values are constrained by permissible reactions to resolutions of that indeterminacy, combined with (a) constraints on resolutions like Unidimensional Expectations and No Trade-Off, and (b) Pearce's result linking expected value to linear mixtures of values.

I'll end with two related objections to this reasoning. One comes from Jamie Dreier (2022). He is discussing which options are better and worse, not more or less preferred, but a translated version of this objection has bite. He writes,

> To say, "Right, but no matter how we got rid of parity [i.e., the denial of trichotomy] in favor of the usual relations, this prospect would turn out to be better than that prospect, so really it must just *be* better," just seems to be a non sequitur. It's as though, having heard that the number six has no spatial location at all, someone replied, "Right, but no matter how we assigned it a spatial location, it would be reachable from Tucson in some finite amount of time by a light wave, so we can conclude that it just is reachable from Tucson in a finite amount of time." (Dreier 2022: 124)

Relatedly, we might ask why the fact that Ms. Jones would prefer one lottery to another on any way of balancing her values should mean that she has that preference now. Why should features of some other value function, one not her own, constrain what she now values?[28]

My reply is that choice functions are meant to play a certain role, they are meant to guide action. But they can't do this on their own. If x and y are both in $C(S)$, and Chooser must choose from S, Chooser needs something more. Chooser needs a *plan*. Ideally she would have what Gibbard (2003) calls a *hyperplan*. A hyperplan H is a function that takes a menu of options S, and returns a member of S. It is a very plausible constraint on C such that there is some coherent H such that for

---

28. Compare the objection to supervaluationism in Fodor and Lepore (1996).

all S, $H(S) \in C(S)$. I conjecture, though I don't have a complete proof of this, that there are plausible constraints on H, similar to Unidimensional Expectations and No Trade-Off, which imply that H is coherent only if there is some utility function such that $H(S)$ is an element of S with maximal expected utility. That's far from a complete argument, but it's the direction I think a reply to Dreier's good objection should take.

The other objection also comes from Dreier's work. He notes that principles like that some choice is choiceworthy iff it is optimal on some permissible resolution of the incompleteness in the values are plausible if we view incompleteness in value as a kind of indeterminacy. So the explanation of Ms. Jones's choice dispositions I gave in Section 22 goes through very smoothly if we say it is indeterminate how she values different skills.

On this way of thinking, it's natural to reject the assumption with which I started this section, i.e., xPy iff $C(\{x, y\}) = \{x\}$. What's more plausibly true is that x is **determinately** preferred to y iff $C(\{x, y\}) = \{x\}$. But if $C(\{x, y\}) = \{x, y\}$, then it is indeterminate what preferences Ms. Jones has.

This leads back naturally to positions similar to the one discussed in Section 18. For example, we could say that preferences are defined not in terms of choice functions, but in terms of hyperplans. For a chooser with choice function C and hyperplan H, they prefer x to y iff $H(\{x, y\}) = x$. For more normal choosers, who have not determined a hyperplan, it is indeterminate what their preferences are; the preferences don't exist until (that part of) the hyperplan does.

On this view Ms. Jones does not violate Negative Dominance because it isn't true that she is indifferent between the four programmers. Rather, it is indeterminate which of them she prefers. What's true of her preferences is simply what is true on all ways of turning her choice function into a hyperplan. That will include that her preferences are trichotomous.[29]

The striking thing is that, like in Section 18, we can say all this while saying all the things which trichotomy was supposed to rule out. Ms. Jones can rationally select any of the programmers, and this remains true even if any of them improves by ε on any skill, but she cannot rationally choose L(Lily)(Suzy) when L(Jane)(Dolly) is available. The resulting position is one where preferences are largely disconnected from rational choice.

The position set out in the last two paragraphs is not a compulsory consequence of the arguments in the rest of the paper. We could simply reject Negative Dominance (Goodness) and say that Ms. Jones is indeed indifferent between the programmers. But we, as theorists, have some options here. Once we say that preferences are analytically posterior to choiceworthiness judgments, there are a lot of ways to understand what preferences are. Some important arguments, like Lederman's, are not quite arguments against a particular view of choice, but the conjunction of that view with a way of understanding preference. It could be we're best off simply understanding preference a different way.

---

29. Indeed, the way I've set things up, her preferences satisfy $xPy \vee yRx$; there are various ways to allow for equality in this framework, and hence preserving trichotomy while rejecting $xPy \vee yRx$.

23. Conclusion

This paper has been ultimately about the grounding of facts about rational choice. I've been mostly concerned to argue against a popular, if largely implicit, view: rational choice is grounded in rational preference. If Chooser wants a holiday, and is choosing where to go, which destinations are rationally choiceworthy is grounded in Chooser's (rational) preferences over pairs of choices. I've rejected this primarily for two reasons:

1. As argued in Section 15 and Section 16, choiceworthiness is an ex ante concept, and preference is an ex post concept, and hence choiceworthiness is analytically prior to preference, so should not be grounded in preference.
2. Any choice function that violates γ cannot be generated from a preference relation, and there are many reasons for endorsing choice functions which violate γ.

If rational choiceworthiness is not grounded in preferences, what is it grounded in? There are two natural options here.

A subjectivist theory, as flagged in Section 3, says that norms on choiceworthiness are just coherence norms. In particular, the key norms are α and Aiz. What makes this choiceworthiness judgment rational just is that it fits well with the other choiceworthiness judgments. There are tricky metaphysical questions about just how instances of coherence norms are grounded, and whether it will mean we have to give up widely accepted principles like that grounding is acyclic. But these questions aren't different in kind from ones that we'd face on the more mainstream view that decision theory is largely about preferences, and norms on preference are coherence norms.

A more objectivist view is also possible, and I'll end by just stating it. The world is full of values, many of them. All of these values are orderings (or perhaps semiorderings) of outcomes. An option is rationally choiceworthy iff it does best, given Chooser's evidence, on some permissible mixture of some of these values. If all goes well, the values can be measured numerically, and the permissible mixtures are linear mixtures, but as noted in Section 2, we need some good arguments about why they should be numerical. If they are, the argument in Section 21 should generalize to argue that the permissible mixtures are linear mixtures. Whether those last two sentences work or not, the picture is that the rationality of a choice is grounded in something external to the agent, i.e., values in the world.

I'm going to leave arguments about which of these views is correct, or which positions between them might be preferable to another day. All I hoped to do in this final section is to sketch ways in which the core thesis of the paper, that the rationality of choices is prior to the rationality of preferences, could be true.

References

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.

Beiser, Frederick C. 2024. "Gustav Theodor Fechner." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2024. https://plato.stanford.edu/archives/sum2024/entries/fechner/; Metaphysics Research Lab, Stanford University.

Bernheim, B. Douglas. 1984. "Rationalizable Strategic Behavior." *Econometrica* 52 (4): 1007–28. https://doi.org/10.2307/1911196.

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.

Broome, John. 1997. "Is Incommensurability Vagueness?" In *Incommensurability, Comparability and Practical Reason*, edited by Ruth Chang, 67–89. Cambridge, MA: Harvard University Press.

Carr, Jennifer Rose. 2020. "Imprecise Evidence Without Imprecise Credences." *Philosophical Studies* 177 (9): 2735–58. https://doi.org/10.1007/s11098-019-01336-7.

Chang, Ruth. 1997. "Introduction." In *Incommensurability, Incomparability and Practical Reason.*, edited by Ruth Chang, 1–34. Cambridge, MA: Harvard University Press.

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

———. 2024. "What's so Hard about Hard Choices?" *Erasmus Journal for Philosophy and Economics* 17 (1): 272–86. https://doi.org/10.23941/ejpe.v17i1.872.

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

Debreu, Gerard. 1960. "Review of *Individual Choice Behavior: A Theoretical Analysis*, by R. Duncan Luce." *American Economic Review* 50 (1): 186–88.

Dorr, Cian, Jacob M. Nebel, and Jake Zuehl. 2023. "The Case for Comparability." *Noûs* 57 (2): 414–53. https://doi.org/10.1111/nous.12407.

Dreier, Jamie. 2022. "Blessed Lives, Bright Prospects, Incomplete Orderings." *Oxford Studies in Normative Ethics* 12: 105–26. https://doi.org/10.1093/oso/9780192868886.003.0006.

Evren, Özgür, and Efe A. Ok. 2011. "On the Multi-Utility Representation of Preference Relations." *Journal of Mathematical Economics* 47 (4): 554–63. https://doi.org/10.1016/j.jmateco.2011.07.003.

Fara, Delia Graff. 2001. "Phenomenal Continua and the Sorites." *Mind* 110 (440): 905–36. https:

//doi.org/10.1093/mind/110.440.905.

Fechner, Gustav. 1860. *Elemente Der Psychophysik*. Leipzig: Breitkopf und Härtel.

Fodor, Jerry A., and Ernest Lepore. 1996. "What Cannot Be Valuated Cannot Be Valuated, and It Cannot Be Supervaluated Either." *Journal of Philosophy* 93 (10): 516–35. https://doi.org/10.5840/jphil1996931013.

Gallow, J. Dmitri. 2020. "The Causal Decision Theorist's Guide to Managing the News." *The Journal of Philosophy* 117 (3): 117–49. https://doi.org/10.5840/jphil202011739.

———. 2024. "It Can Be Irrational to Knowingly Choose the Best." *Australasian Journal of Philosophy* 103 (2): 540–46. https://doi.org/10.1080/00048402.2024.2310197.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

———. 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. 2009. "Preference-Based Choice Functions: A Generalized Approach." *Synthese* 171 (2): 257–69. https://doi.org/10.1007/s11229-009-9650-5.

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.

Holton, Richard. 1999. "Intention and Weakness of Will." *The Journal of Philosophy* 96 (5): 241–62. https://doi.org/10.2307/2564667.

Lederman, Harvey. forthcoming. "Of Marbles and Matchsticks." *Oxford Studies in Epistemology*, forthcoming.

———. 2023. "Incompleteness, Independence, and Negative Dominance." https://philpapers.org/archive/LEDIIA.pdf.

Lehrer, Keith, and Carl Wagner. 1985. "Intransitive Indifference: The Semi-Order Problem." *Synthese* 65: 249–56. https://doi.org/10.1007/BF00869302.

Levi, Isaac. 1986. *Hard Choices*. Cambridge: Cambridge University Press.

Lewis, David. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy* 59 (1): 5–30. https://doi.org/10.1080/00048408112340011.

Luce, R. Duncan. 1956. "Semiorders and a Theory of Utility Discrimination." *Econometrica* 24 (2): 178–91. https://doi.org/10.2307/1905751.

———. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.

Moss, Sarah. 2015. "Time-Slice Epistemology and Action Under Indeterminacy." *Oxford Studies in Epistemology* 5: 172–94. https://doi.org/10.1093/acprof:oso/9780198722762.003.0006.

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.

Pearce, David G. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52 (4): 1029–50. https://doi.org/10.2307/1911197.

Peterson, Martin. 2017. *An Introduction to Decision Theory*. Second. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316585061.

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.

Sartre, Jean-Paul. 1946/2007. "Existentialism Is a Humanism." In *Existentialism Is a Humanism*, translated by Annie Cohen-Solal, 17–72. New Haven: Yale University Press.

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.

———. 2004. "Incompleteness and Reasoned Choice." *Synthese* 140 (1-2): 43–59. https://doi.org/10.1023/B:SYNT.0000029940.51537.b3.

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

Skyrms, Brian. 1982. "Causal Decision Theory." *Journal of Philosophy* 79 (11): 695–711. https://doi.org/10.2307/2026547.

Spencer, Jack. 2023. "Can It Be Irrational to Knowingly Choose the Best?" *Australasian Journal of Philosophy* 101 (1): 128–39. https://doi.org/10.1080/00048402.2021.1958880.

Tarsney, Christian, Harvey Lederman, and Dean Spears. forthcoming. "A Dominance Argument Against Incompleteness." *Philosophical Review*, forthcoming. https://doi.org/10.48550/arXiv.2403.17641.

Weatherson, Brian. 2024. *Knowledge: A Human Interest Story*. Cambridge: Open Book Publishers. https://doi.org/10.11647/obp.0425.