Of marbles and matchsticks*

Harvey Lederman

First version: June 17, 2023
This version: September 2, 2023

Abstract

I present a new puzzle about choice under uncertainty for agents with preferences which are sensitive to multiple dimensions of outcomes in such a way as to be incomplete. In response, I develop a new theory of choice under uncertainty for incomplete preferences. I connect the puzzle to central questions in epistemology about the nature of rational requirements, and ask whether it shows that it is rationally required to have complete preferences.

1

Mira loves both marbles and matchsticks, but for very different reasons and in very different ways. She loves the smoothness of marbles, the chill they've assumed in the morning when she wakes, and the mystery of what life is like on the surface of those multicolored twists. She also likes the way that matchsticks look and feel—their slenderness, their splintery humility—but what she loves about them is different: it's their sad, glorious promise, their mute anticipation of their own bright end.

Mira has five marbles and five matchsticks to her name. It's not too few, but it's not so many either. If you offer her more marbles or more matchsticks (or both), at no cost, she'll gladly accept. If you offer to take away matches or marbles without payment, she'll angrily refuse.

Mira loves all of her marbles and matchsticks, but there are some prices at which she'd give one or more away. If you offer her a lot of marbles in exchange for a few matchsticks, or a lot of matches in exchange for a few marbles, she might well accept. For a hundred marbles, for instance, Mira will happily hand over a match. Mira is many things, but she isn't insane.

*Thanks to Calvin Baker, Adam Bales, Kyle Blumberg, Chris Bottomley, Sam Carter, Pietro Cibinel, Adam Elga, Nathan Engel-Hawbecker, Jane Friedman, Johan Gustafsson, Caspar Hare, Ben Holguín, Nico Kirk-Giannini, Jon Kleinberg, Sebastian Liu, Richard Pettigrew, Wlodek Rabinowicz, Dean Spears, Christian Tarsney, and Teru Thomas for conversations and correspondence; to audiences at TARK and the Global Priorities Workshop for questions and comments; to Cian Dorr, Brian Hedden, and Jake Nebel for extremely detailed comments on drafts, which helped me to improve the paper enormously; and finally to Chiara Damiolini for the figures and several key conversations.

Still, there some prices at which Mira finds herself stuck. If you offer her four matchsticks in exchange for two marbles, for instance, she finds there isn't a clear path to a choice. Sure, the trade would get her more matchsticks, and that's great (think of the anticipation!), but she'll lose two marbles, and that's not (two fewer twists!). The considerations in favor of the trade don't outweigh the considerations against; nor do those against outweigh those for.

If you ask Mira why she's stuck, she'll tell you that she doesn't prefer this trade to what she now has, and that she also doesn't prefer what she has now to the trade. It's also not, it seems to her, that she's exactly indifferent between the trade and what she's got. If she were, and you slightly sweetened the deal, then that would tip the balance: she should prefer the sweetened deal to her current stash. But she finds that she doesn't prefer the sweetening to what she has, either. If you offer Mira a choice between getting four matchsticks in exchange for two marbles, on the one hand, and getting *five* matchsticks in exchange for two marbles, she'd of course go for the five. As I told you, she isn't insane. But if you offered her just the five-two trade as opposed to what she has now, she'd still pause in doubt. She'd say she doesn't prefer the trade to what she's got, or what she's got to the trade. For Mira, these aren't ties to be settled any which way because it's all the same in the end. They are a different kind of hard choices to make.

2

If you're like me, you're in Mira's place more often than you might like to admit. Many of us get stuck picking between apples and oranges, carrots and cabbages, chalk and cheese, or, as I'm told the Serbians have it, between grandmothers and toads. (Though that one, I hasten to add, has never given me pause.) More often we're stuck making heavier tradeoffs of a more abstract kind. A large college offers a broader range of opportunities, but a smaller one provides a tighter community. One career has greater earning potential and the possibility of living close to family; another offers you a chance to do meaningful work you really love. One house offers more space but a longer commute; another's a bit cramped, but you can walk to work.

At least for some of us, some of the time, it seems that it's not that we're so repressed that we can't figure out what our deeper selves really want. It's not that if we thought about it more, or spent a few more years with an analyst, we'd realize that our values really favor one side. Instead, we just don't prefer one to the other, or the other to the one. And we aren't indifferent either: if you add some opportunities to the smaller college, add money to the meaningful career, or add a bit of space to one house, it's not that we all of a sudden prefer the sweetened deal. We just don't prefer one to the other, or the other to the one, and that's all there is to say.[1]

[1] "Prefer" without qualification should be understood from here on out as "weakly

This paper develops a new puzzle for preferences like this, preferences which are sensitive to various dimensions of options, in such a way as to be *incomplete*: they don't rank every pair of options with respect to one another. I'll show that having such preferences over options which do not involve uncertainty is incompatible with satisfying natural constraints on preferences over options which do involve uncertainty (§3 and §4). (I'll focus on preferences, but as I'll discuss at the end of §4, there's a version of the puzzle that arises in axiology as well.) I'll explain how the new puzzle differs from the phenomenon of 'opaque sweetening', discovered by Caspar Hare (§5), and then tentatively consider the ways forward for fans of incomplete preferences: either giving up what I call "Negative Dominance" (§6), or endorsing a new theory (presented here for the first time) which gives up Independence in some surprising new ways (§7).

Incomplete preferences and values have most often been discussed somewhere in-between action theory, ethics, and decision theory. But the problems they raise touch on central questions in epistemology, about the nature of "structural" rational requirements, and their relationship to features of the world. A conclusion (§8) draws out this issue, as part of my discussion of whether the puzzle supports the claim that rational preferences must be complete.

3

Suppose we offer Mira a game. We'll flip a coin. If the coin comes up Heads, we'll give her four matchsticks in exchange for two marbles. If the coin comes up Tails, we'll give her four marbles and take two matchsticks. The game is shown in the table below. I call it "The Hard Game".

The Hard Game

	Matchsticks	Marbles
Heads	4	-2
Tails	-2	4

Should Mira prefer playing this game to her current holdings, or should she prefer her current holdings to the game? A plausible argument can be given on either side. This pair of arguments will give us a first look at our puzzle.

I haven't told you much about Mira but it's important to know before we go on that her interest in marbles and matchsticks doesn't particularly

prefer", which allows for indifference. Throughout I'll understand preference not as a disposition to choose in a forced choice but as an attitude on a par with wanting, fearing or hoping. (Like all of these other attitudes, preferences may ultimately be some kind of disposition, but it's not as simple a disposition to choose in a forced choice.) My talk of preferences follows a standard practice in decision theory, but I recognize not everyone will accept that preferences play the central role in explaining choice as they are assumed to play in this framework. I invite skeptics (with whom I have considerable sympathy) to translate what I'll say into claims about other attitudes: what people want with various strengths, what they all-things-considered value, and so on.

abate if she gets more of them, at least at smaller scales. Sure, maybe after a hundred thousand marbles or matchsticks, she'd start to get bored, but with a gain of just four or for that matter ten, there's not much difference between how she values the tenth marble by comparison to the fourth. (In the lingo, she values them "linearly".) Mira also isn't opposed or attracted to games of chance for reasons other than what she might win by playing them. She doesn't shy from a game of chance just because she might lose by playing it, or seek out such games just for the thrill. (In the lingo, she's not "risk averse" or "risk prone".)

Our first argument is that, given how she is, Mira should strictly prefer the Hard Game. The game's *expected* value in matchsticks—its average yield, weighted by the probability of each outcome—is 1 (it's $\frac{1}{2} \cdot 4 = 2$ (if it's Heads), plus $\frac{1}{2} \cdot \text{-}2 = \text{-}1$ (if it's Tails)). And its expected yield of marbles is the same (it is $\frac{1}{2} \cdot \text{-}2$ (if Heads) plus $\frac{1}{2} \cdot 4$ (if Tails)). Since Mira isn't risk-averse or risk-prone, she should be indifferent between the Hard Game and a certain gain equal to its expectation, that is, a certain gain of one marble and one matchstick. And since Mira strictly prefers this certain gain to what she has now (a free marble! a free matchstick!), she should strictly prefer the game to what she's got.

Our second argument is that, given how she is, Mira should not strictly prefer the game. If Mira strictly prefers one game of chance over another, then since she doesn't care for games of chance as such, her strict preference for the first game must be explained by a strict preference for at least one of the prizes she could get by playing it, as opposed to one of the prizes she could get by playing the other. If she has no preference between any of the prizes of the first game and any prizes of the second game; what could possibly explain Mira's preference for the first game as a whole? As we saw, Mira does not strictly prefer any of the prizes of the Hard Game to the prize she would get by sticking with what she currently has (which I'll count as an honorary 'game of chance'). As I told you, Mira does not prefer giving up two marbles in exchange for four matchsticks, by comparison to sticking with what she's got. And Mira (as I did not tell you, but I'm telling you now) also does not prefer giving up two matchsticks in exchange for four marbles, by comparison to what she has now. As a result, it would be inexplicable for her to strictly prefer the Hard Game over her current holdings. Since she should not have inexplicable preferences, she should not prefer the game.

We can state the puzzle as a conflict between the following two principles:

Expectationalism: It's rationally required that: if Mira strictly prefers a certain gain of a particular bundle of marbles and matchsticks to her present holdings, she strictly prefers a game of chance which has an expected yield of that number of marbles and matchsticks to her present holdings.

Negative Dominance: It's rationally required that: if Mira strictly prefers

> one game of chance to another, she prefers one of the prizes that the first might yield, to one of the prizes that the second might yield.[2]

I've said that these arguments give rise to a puzzle, and ultimately, as I'll suggest later, I think they do. But when I first thought about this, and even now, when I see it just as a conflict between Expectationalism and Negative Dominance, it seems obvious to me which one we should reject. If Expectationalism is true, Mira's preferences would be bizarre. Suppose we offer Mira a choice between three options: (a) sticking with what she has; (b) trading in two matchsticks to get four marbles; or (c) trading in two marbles to get four matchsticks. In response, Mira shrugs, squints and throws up her hands; she doesn't have a preference any which way. But now, while she's stuck, we swoop in, and take those options off the table. In place of her initial choice, we offer her instead a choice between two options: the first is just (a) from before, sticking with what she has; the second choice is the Hard Game: we'll flip a coin, and if it's Heads, she'll get (b) (trading in two matchsticks to get four marbles), while if it's Tails, she'll get (c) (trading in two marbles to get four matchsticks). According to Expectationalism, this second choice is clear! But how could this be? Mira doesn't care for games of chance as such, and didn't have a preference in our first three-way choice. So her newfound preference here seems baseless, and bizarre.

It's not—I hasten to add—that I can't imagine *anyone* for whom this preference might make sense. Lots of people prefer to avoid decisions; they prefer a coin flip whenever it gets them out of having to decide. Other people just love coin flips, regardless of what the outcomes bring. For these people, there's no mystery here: they have a preference for the game of chance "as such". But I told you a moment ago that Mira isn't like this; she doesn't care one way or another about the fact that the choice involves chance. The pattern of actions or preferences Expectationlism predicts on its own isn't weird. What's weird is that Mira displays this pattern when she doesn't care one way or another for a coin-flip as such. In her case, the preference is baseless; it can't possibly be a paradigm of rationality. Instead, as I've said, it's bizarre.[3]

It's also pretty easy to see where Expectationalism might have gone off the rails. The expected value of the game as a whole 'forgets' how the original values were distributed across the prizes Mira might have won. It

[2] Negative Dominance is so-called because of its contrapositive. A weak version of Dominance says that it's rationally required that if Mira prefers every prize in one game of chance to every prize in another, then she prefers the first. Negative Dominance says (contrapositively) that it's rationally required that if there are no prizes in one game that Mira prefers to any prizes in a second game, then she does *not* prefer the first.

[3] If "it's rationally required" is replaced with "it's rationally permitted" in at most one of Expectationalism and Negative Dominance, the puzzle still arises. So, in particular, even if Negative Dominance leads to a permitted lack of preference, this still creates a problem if it's rationally required that: if a person values marbles and matchsticks linearly and is not risk averse or risk prone, they will be indifferent between games of chance and their expected values.

doesn't record that in the Hard Game, Mira gets a high value for marbles at the price of a low value for matchsticks, and vice versa. The expected value is the same as if she'd win four of both on Heads, and lose two of both on Tails. But we can't ignore this structure. In the Hard Game, Mira isn't going to win the expected value; she's going to win a high-low pair that she doesn't prefer, whether it's trading some marbles for some matchsticks, or the other way round.

4

The prizes in the Hard Game differ from each other both in the number of matchsticks and in the number of marbles they yield. When a game's prizes differ from each other in this way, the expected value can 'forget' important aspects of the structure of the game. But if a game's prizes differ from each other only in their number of matchsticks, and all agree in their number of marbles, or if they differ from each other only in their number of marbles, and all agree in their number of matchsticks, the prizes don't have so much structure; there's really nothing interesting to forget. So it seems plausible that, in such restricted cases, it would make sense to value the games in line with their expected values.

If we think of the matchsticks and marbles as different 'dimensions of value', the idea is to restrict attention to *unidimensional* games, where a game is unidimensional if its prizes are all the same in their number of matchsticks, or if its prizes are all the same in their number of marbles. The idea is that, in this special case, given how she is, Mira should treat the games as equivalent to their expected yield. More generally:

Unidimensional Expectations: It's rationally required that: if a person values marbles and matchsticks linearly, and is not averse or prone to risk, then they are indifferent between any unidimensional game and a certain gain of its expected value in marbles and matchsticks.

I've already told you that Mira herself values marbles and matchsticks linearly and is neutral about risk. Pretty much everyone agrees—and I'll assume for now, though I'll come back to it below—that doing so is rationally permissible. So I'll understand Unidimensional Expectations to imply that it's rationally required for Mira herself, given how she is, to be indifferent between unidimensional games and a certain gain of their expected value.[4]

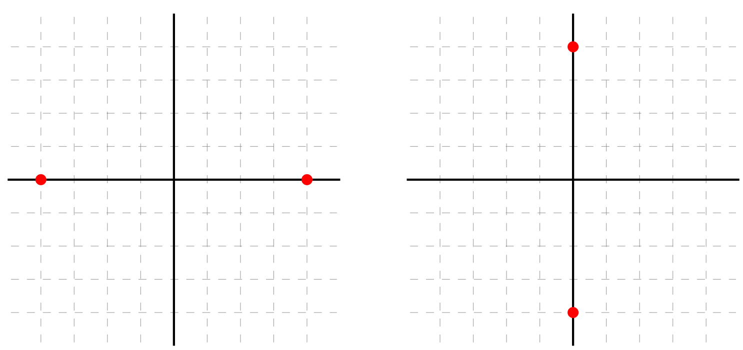
Unidimensional Expectations (as well as the consequence I just mentioned) is *much* weaker than Expectationalism. On its own it says nothing

[4]This would follow given a somewhat controversial form of detachment: that if it's rationally required that if a person Fs then they G, and it's rationally permitted for a person to F, then if they F, it's rationally required for them to G. But even if we reject this general principle, it's very natural to just assume that this requirement holds for Mira here, and I'll sometimes use "Unidimensional Expectations" to refer to such a requirement.

about how to value games whose prizes are not 'unidimensional' with each other, and so, it says nothing about the Hard Game. And it's not just logically weaker than Expectationalism; as I'll discuss in more detail later on (§6), it's directly motivated in a way that Expectationalism isn't, since it's basically an analytic truth: valuing unidimensional games as equivalent to their expectations is really just what it means to value marbles and matchsticks linearly while also being neutral about risk.

Even so, as we're about to see, with Unidimensional Expectations in the background, Negative Dominance conflicts with Independence (introduced below), which is another very plausible principle about choices under risk. This second conflict is the main puzzle of the paper.

In developing the argument for a conflict, it will be helpful to have some diagrams. I'll represent prizes in our games as points in the Cartesian plane, with the number of matchsticks Mira would gain or lose on the x-axis (x for "sti*cks*") and the number of marbles she would gain or lose on the y-axis (so her present holdings will be $(0, 0)$). In any given picture each (unshaded) point will be understood to occur with equal probability. Below, for instance, we have two unidimensional games of chance. On the left, there is a game where Mira would get four matchsticks with probability $\frac{1}{2}$ and would lose four matchsticks with probability $\frac{1}{2}$. On the right, is a second game, where Mira would get four marbles with probability $\frac{1}{2}$ and would lose four marbles with probability $\frac{1}{2}$.



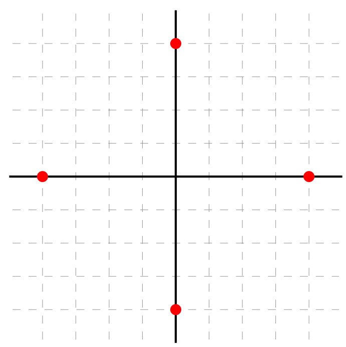
By Unidimensional Expectations, Mira should be indifferent between each of these games of chance and her present holdings.

If Mira is indifferent between her present holdings and each of these games, it's natural to think that she should also be indifferent between her present holdings and a game of chance which combines these two. Suppose we offer her a choice between sticking with what she has, and a new coin-flip. (I'm afraid there are going to be a few of these.) If the coin comes up Heads, we'll flip a second time, and award prizes as in the game on the left (it won't matter which is Heads or Tails). If the coin comes up Tails, we'll also flip a second time, and award prizes as in the game on the right. Plausibly, Mira should be indifferent between what she has now, and this sequence of flips.

But plausibly, too, the fact that there's a *sequence* of flips doesn't matter. It's not as if Mira likes coin-tosses and so prefers a sequence where there are more flips. So if we can produce, in a game that happens all at once, the same chances of the same prizes that she would get in the sequence of games, Mira should also be indifferent between the game that happens all at once, and what she presently has. It's not hard to make this happen at all once; we can do so with a four-side die. This new game, which I'll call "the Die Roll", is as follows: if our die comes up 1, we give Mira four matchsticks; if 2, she gives us four; if 3, we give her four marbles; if 4, she gives us four. The game is represented in the following table and picture.

The Die Roll

	Matchsticks	Marbles
One	4	0
Two	-4	0
Three	0	4
Four	0	-4



The Die Roll involves prizes which differ from each other in both their number of matchsticks and in their number of marbles. It's not a unidimensional game. But the Die Roll is composed of two unidimensional games, and since Mira is indifferent between both of those games and her current holdings, plausibly she should be indifferent between this new game and what she has now.

This informal justification seems plausible to me, and I hope it seems so to you too. But I'd like to be a bit more systematic here, and pin down a principle that suffices for this claim. In fact, we can derive Mira's indifference using the following principle, a version of which has been standard since Von Neumann and Morgenstern [1944]:

Independence It's rationally required that: if someone is not averse or prone to risk, they prefer a game of chance A to a game of chance B if and only if they also prefer a game consisting of a p chance of A and a $(1-p)$ chance of some other C to a game consisting of a p chance of B and a $(1-p)$ chance of the same C.

This principle is a bit of a mouthful, but we've already appealed to something close to it implicitly, and it's anyway very plausible when you break it down. If Mira prefers A to B, then, given that she's not averse or prone to risk, that preference should remain if we "mix in" a fixed chance of C to both of them. The only difference between the two new games is that one has an 'amount' of A, where the other has an amount of B, and given that she prefers A on its own to B on its own, she should prefer the A-mixture to the B one. Similarly, in the reverse direction: If Mira prefers the game which has A mixed with C to the one which has B mixed with C, then, given that she's not averse or prone to risk, it must be because she prefers A to B on their own, since that's the only difference between them.[5]

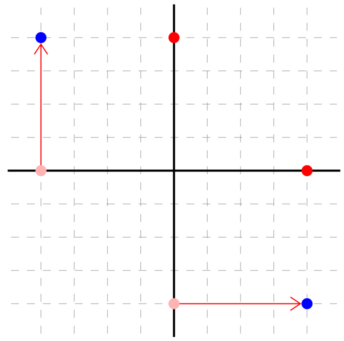
So Unidimensional Expectations and Independence, together with the claim that Mira is as rationality requires her to be, give us that Mira is indifferent between the Die Roll and what she has. If we improve the Die Roll significantly, then, she should prefer that improved game over her current holdings. So let's improve the game, in two different ways. First, if the die comes up Two, we'll still take away four matchsticks, but now we'll also also give Mira four marbles in recompense. Since getting four marbles makes the outcome better in a way and keeps it at least as good in every other way, this is an overall improvement in the outcome, and so, an overall improvement to the game. Second, if the die comes up Four we won't just take away four marbles, but we'll also give Mira four matchsticks. Once again, this change makes the outcome better in one way while keeping it at least as good in every other way, so again it is an overall improvement in the game. The table and picture below show the changes along with the new game. (Formally, we can justify both of these steps using Independence, but they're so plausible taken just on their own that I won't belabor the point.)[6]

The Improved Die Roll

	Matchsticks	Marbles
One	4	0
Two	-4	4
Three	0	4
Four	4	-4

[5]To see how the principle justifies our reasoning, note first that it implies that, since $(0,0)$ is indifferent to the game of chance giving equal probability of $(0,\text{-}4)$ and $(0,4)$, $(0,0)$ is also indifferent to a game of chance which yields $(0,0)$ with probability $\frac{1}{2}$ and yields $(0,\text{-}4)$ and $(0,4)$ each with probability $\frac{1}{4}$. Given that $(0,0)$ is indifferent to the game of chance giving equal probability of $(\text{-}4,0)$ and $(4,0)$, the principle implies that a game which gives $(0,0)$ with probability $\frac{1}{2}$ and yields $(0,\text{-}4)$ and $(0,4)$ each with probability $\frac{1}{4}$ is indifferent to a game which yields $(0,\text{-}4)$, $(0,4)$, $(\text{-}4,0)$, and $(4,0)$ each with probability $\frac{1}{4}$. Since the former is indifferent to $(0,0)$, the latter must be as well.

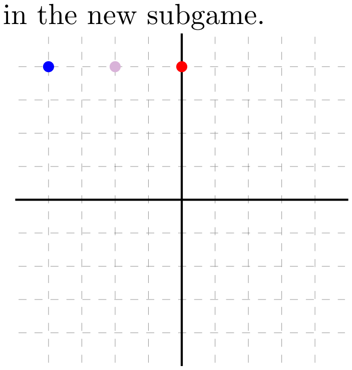
[6]In fact the weaker Stochastic Dominance would do as well (for a formal statement, see n. 19). A full proof is given in Lederman [2023].



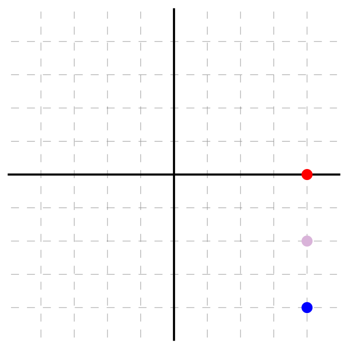
Given all that we've seen so far, we can conclude that Mira should strictly prefer the Improved Die Roll to what she has. We just saw that she should strictly prefer it to the Die Roll, and we already saw that she was indifferent between the Die Roll and what she currently has. So, she should strictly prefer the Improved Die Roll to her current holdings.

We're just one step away now from the promised conflict. The key last step will be to show that that Unidimensional Expectations and Independence imply that Mira should be indifferent between The Improved Die Roll and the Hard Game. Since we've already seen that she should prefer The Improved Die Roll to her present holdings, it will follow that she should prefer the Hard Game to her present holdings. And this, as we well know, conflicts with Negative Dominance.

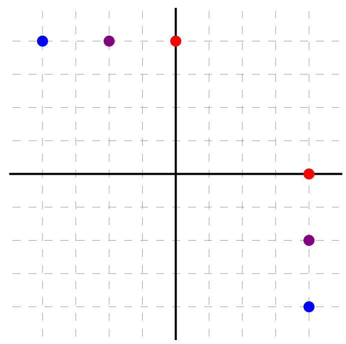
The argument for this last step can be broken into three parts. For the first part, let's consider first just the two points at the top left of the previous diagram, with coordinates $(\text{-}4, 4)$ and $(0, 4)$, which stand (in the first case) for Mira losing four matchsticks while gaining four marbles, and (in the second) for Mira keeping all her matchsticks while gaining four marbles. (These are the prizes if the die comes up Two and Three.) A game which yields each of these prizes with $\frac{1}{2}$ probability would be a unidimensional game, since both of its prizes agree in yielding four marbles, and only differ from each other in the number of matchsticks they yield. So according to Unidimensional Expectations, Mira should be indifferent between such a game and its expected value: that is, she should be indifferent between this game and a certainty of gaining four marbles, and losing two matchsticks (since $\frac{1}{2} \cdot \text{-}4 + \frac{1}{2} \cdot 0 = \text{-}2$). You can see all this in the figure, where the middle purple dot is the expected value, and the red and blue dots on either side are even-chance prizes in the new subgame.



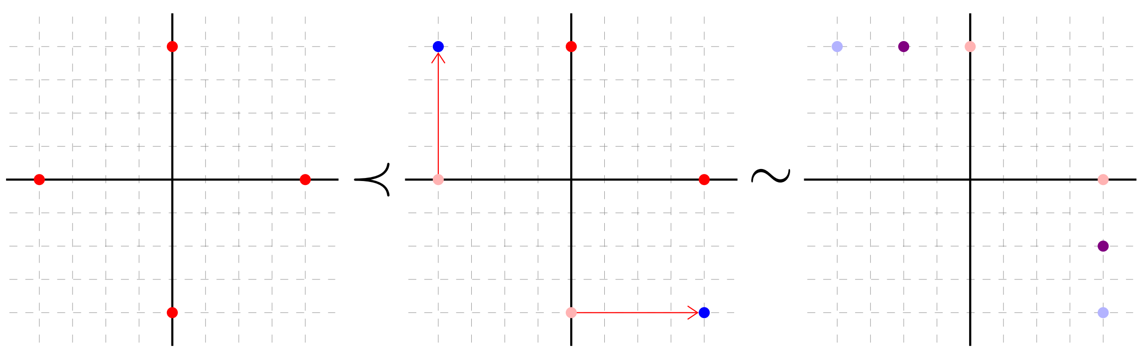
The second part of the argument is to observe that something similar holds for the two points on the lower right of the diagram, with coordinates $(4, 0)$ and $(4, \text{-}4)$, which stand (in the first case) for Mira's gaining four matchsticks while keeping all her marbles, and (in the second) for Mira's gaining four matchsticks, while losing four marbles. (They're the prizes if the die lands One or Four.) A game which yields these prizes with $\frac{1}{2}$ probability would be a unidimensional game, since all of its prizes agree in yielding four matchsticks, and differ from each other only in the number of marbles they yield. So according to Unidimensional Expectations, Mira should be indifferent between such a game, and its expected value: that is, she should be indifferent between this game and a certain prize where loses four marbles and gains two matchsticks (since $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 0 = 2$). You can see this again in the next figure, where again the central shaded purple dot stands for the expected value, and the red and blue dots are even chance prizes in our game.



The third part of the argument is to use these facts to connect the Improved Die Roll and the Hard Game. (The next diagram shows the Hard Game, with unshaded purple dots, and the Improved Die Roll, with shaded dots, both red and blue.) First, note that if Mira is indifferent between the red/blue games and their corresponding purple dots, she should also be indifferent between, on the one hand, a coin flip over the two red/blue games, and, on the other hand, a coin flip over the purple dots. These games give her equal chances of equally good prizes. Second, the fact that there's a sequence of coin-flips in the first of these two games, rather than the prizes being doled out all at once, shouldn't matter to Mira for reasons we went through above. A coin flip over the two red/blue games would give Mira a $\frac{1}{4}$ chance of each of the red or blue prizes, and she should be indifferent between this coin flip, and any game that gives her the same chances of the same prizes. In particular, she should be indifferent between this coin flip over red/blue games, and the Improved Die Roll, which gives a $\frac{1}{4}$ chance each at the same prizes. Finally, as we said, she should be indifferent between a coin flip over the red/blue games and a coin toss over the purple dots, so she should also be indifferent between the Improved Die Roll and a coin toss over the purple dots. But a coin toss over the purple dots is just the Hard Game. In it, Mira has a $\frac{1}{2}$ chance of gaining four marbles, and losing two matchsticks; she also has a $\frac{1}{2}$ chance of gaining four matchsticks, and losing two marbles. So we can conclude, as promised, that Mira should be indifferent between the Hard Game and the Improved Die Roll.[7] Since she should strictly prefer the latter to her current holdings, she should strictly prefer the Hard Game to her current holdings, in violation of Negative Dominance.



Putting the whole argument together in a single picture, we can see it as follows:



So, as promised, given Unidimensional Expectations, and the rational permissibility of Mira's state of mind, Independence conflicts with Negative Dominance.

In fact, these principles are much stronger than we need to generate a conflict. If we replace "it's rationally required" with "it's rationally permitted" in Negative Dominance, the resulting principle still conflicts with Unidimensional Expectations and Independence.[8] Moreover, while I've as-

[7] We can make this reasoning more formal as follows: given that (by Unidimensional Expectations) (-2, 4) is indifferent to a game with an even probability of (-4, 4) and (0, 4), Independence implies that the Hard Game, which gives an even probability of (-2, 4) and (4, -2), is indifferent to a game which yields (-4, 4) and (0, 4) each with probability $\frac{1}{4}$, and yields (4, -2) with probability $\frac{1}{2}$. Moreover, since (4, -2) is indifferent to a game with an even probability of (4, -4) and (4, 0), by Independence a game which yields (4, -2) with probability $\frac{1}{2}$ and yields (-4, 4) and (0, 4) each with probability $\frac{1}{4}$ will be indifferent to the Improved Die Roll, which yields (4, -4), (4, 0), (-4, 4), and (0, 4) each with probability $\frac{1}{4}$. Since the former is indifferent to the Hard Game, the latter is too.

[8] Similarly, if we delete "it's rationally required" in Independence, separate out the biconditional and insert "it's rationally permitted" in the consequent of each conditional,

sumed the permissibility of valuing marbles and matchsticks linearly, and of being neutral with respect to risk, there's a much more general mathematical result, which weakens these assumptions greatly (Lederman [2023, Proposition 1.3])). The more general result applies also to people who have (for example) "diminishing marginal utility" in both marbles and matchsticks, and also to people who have a wide array of "risk attitudes", including those who are quite risk averse. Even if one thought (in my view, implausibly) that Mira's state of mind makes it forbidden *for her* to value marbles and matchsticks linearly or to be neutral with respect to risk, that still wouldn't escape the problem: so long as *any* of these other ways of valuing marbles and matchsticks and/or risk is permitted for her, the problem still arises. Finally, the argument doesn't require that Mira be *exactly* indifferent between unidimensional games of chance, and their expected values. The problem still arises even if we allow that Mira considers unidimensional games of chance to be incomparable to their expected values, requiring only that if Mira strictly prefers the expected value of a game to a prize, she strictly prefers the game to the prize (and similarly for what she strictly disprefers) (Lederman [2023, Section 1.5]).

And, of course, the puzzle applies much more generally than our fictional Mira's odd love of marbles and matchsticks. There are many domains, including those with more 'dimensions' than two, where appropriate versions of all three of the principles seem plausible. In those settings, too, they can't all hold together.

In fact, the puzzle is not even restricted to the theory of rational preference; a version of it arises in axiology as well. Many believe that some things are better for some individuals than others are. Many are *pluralists* about this notion of betterness for a person; they hold that what is good for a person is sensitive to different dimensions of value: knowledge, friendship, happiness, achievement, love. Some pluralists are drawn to the thought that tradeoffs across different dimensions can lead to incompleteness in what's better for a person, so that some pairs of options are not ranked with respect to each other.

Suppose, then, in line with this picture, that instead of marbles and matchsticks, we understand our different dimensions as (say) achievement and happiness. (Achievement and happiness don't come in numbered units, but we can pretend that they do here; in Lederman [2023, §2], I show that this assumption is harmless.) Thinking of achievement as the x axis and happiness as the y, we can take Mira's *status quo* life to be $(0, 0)$ and consider again $(\text{-}2, 4)$ (a life with less achievement than the status quo, but greater happiness), and $(4, \text{-}2)$ (a life with greater achievement than the status quo, but less happiness). We can then show that the following three principles can't all be true:

Negative Dominance (Goodness): If one game of chance is better for Mira than another, then some prize in the first game of chance is

the result still conflicts with Unidimensional Expectations and Negative Dominance.

better for her than some prize in the second.

Unidimensional Expectations (Goodness): If every prize in a game of chance yields the same amount of achievement, then the game is exactly as good for Mira as a certain gain of its expected value achievement and happiness. Similarly, if every prize in a game yields the same amount of happiness, then the game is exactly as good for Mira as a certain gain of its expected value in achievement and happiness.

Independence (Betterness) A game of chance L is better for Mira than a game of chance L' if and only if a game consisting of a p chance of L and a $(1-p)$ chance of some other L'' is better for her than a game consisting of a p chance of L' and a $(1-p)$ chance of the same L''.[9]

The puzzle also arises also for consequentialists who accept a notion of overall betterness (say, of a state of the world), and hold that this notion of betterness can be incomplete owing to different dimensions of what is good (overall). In fact, if the relevant notions of betterness (for a person, or overall) give rise to corresponding "oughts" (a prudential ought, or a moral ought), there will also be corresponding puzzles for what Mira "ought" (in that sense to do). Appropriate versions of Unidimensional Expectations and Independence would imply that Mira ought prudentially/morally to choose an analogue of the Hard Game, while the appropriate version of Negative Dominance would imply that she is permitted not to.

That concludes my presentation of the main problem. In the next section, §5, I turn to a task I've already deferred too long: showing how Mira's problem differs from the phenomenon of "opaque sweetening", discovered by Hare [2010] (for discussion see, among others Hare [2013], Schoenfield [2014], Bales et al. [2014], Bader [2018], Doody [2019b,a, 2021], Rabinowicz [2021], Steele [2021], Bader [2023], Russell [forthcominga, §3.2]).

I then discuss two responses that proponents of incomplete preferences might adopt, primarily with the aim of arguing that the way forward for them is far from clear. I don't know of any theory which says that one is rationally forbidden from linearly valuing marbles and matchsticks, while at the same time being neutral with respect to risk, and, anyway, as I've said, claiming that Mira is forbidden from having such states on its own isn't even close to enough to escape the most general form of the problem. So I'll focus here on the prospects for denying Negative Dominance or Independence (assuming, for simplicity, both Unidimensional Expectations, and the permissibility of linear valuing and risk-neutrality). In §6 I discuss how an approach based on sets of utility functions implies Independence (and thus rules out Negative Dominance), and consider how the puzzle

[9] As I'll discuss at the end of §8, arguably the sense of betterness that applies to lotteries is different from the sense of betterness that applies to lives, but these principles don't require that these are the same notion; they just rely on the idea that betterness for lotteries interacts with betterness for prizes in the stated way.

presents a new problem for this view. In §7, I develop a novel, strong theory that rejects Independence, but also express some doubt about whether it is ultimately correct. In §8, at last, I come to the "nuclear option", of denying the rationality of incomplete preferences altogether.[10]

5

To see the difference between Mira's problem and the problem posed by opaque sweetening, it will be helpful to state Hare's original example. In that example, there are four outcomes, A, A^+, B, and B^+, with A^+ preferred to A, B^+ preferred to B, and the As and Bs incommensurable with each other. (In terms of marbles and matchsticks, we could think of A as $(3,1)$, A^+ as $(4,2)$, B as $(1,3)$ and B^+ as $(2,4)$.) Hare presents a choice between two games, L and L^+. A coin will be flipped. In the first game, L, Heads will yield A, while Tails will yield B. In the second game, L^+, Heads will yield B^+, and Tails will yield A^+. The games are depicted in the following table.

	L	L^+
Heads	A	B^+
Tails	B	A^+

Hare gives two arguments that one is rationally required to choose L^+, and two arguments that one is permitted to choose L (and thus not required to take L^+). To make my case that Mira's problem is different from Hare's, I'll recap one of each of these arguments, and argue that endorsing it would not on its own provide a resolution of Mira's problem. (In notes, I'll document how other arguments he and others have given, wouldn't settle it either.)

I'll begin with an argument that one is rationally required to take L^+. To state it, we first need a definition. If a game of chance G gives prizes

[10]This discussion of is not meant to be comprehensive. To take just one salient omission: in the main text, I won't discuss a response which rejects the possibility of the examples required to drive the axiological version of the puzzle, by postulating that balance is an extra dimension of value (along lines suggested by Hedden and Muñoz [2023] cf. Chang [2002, 2016]). I can't resist noting, though, that Brian Hedden (p.c.) has provided an example that illustrates severe challenges to making this response work. Suppose that there are two non-balance dimensions, x and y, each represented by $\mathbb{R}$, and model balance as -$|x-y|$, so that $\mathcal{O} \subset \mathbb{R}^3$, where the inclusion is strict, since each (x, y) pair is only matched with one (non-positive) z value. For instance, $(0,10)$ is $(0,10,\text{-}10)$, and $(10,0)$ is $(10,0,\text{-}10)$. A fair lottery on these two outcomes has an expectation of $(5,5,\text{-}10)$, which is worse on every dimension than $(6,6,0)$. But if, when a prize p is worse (but not sufficiently worse) along some dimensions than a prize p', and better (but not sufficiently better) along other dimensions than p', the two are incomparable, $(6,6,0)$ will be incomparable with $(0,10,\text{-}10)$ and $(10,0,\text{-}10)$. A s result Expectationalism would still force violations of Negative Dominance (and this result is preserved even if we take the balance to shift with the expectation, so that the expected value of the lottery would be $(5,5,0)$.

depending on the *states* Heads and Tails, which each receive probability $\frac{1}{2}$, its *twin* is a game that on Heads yields what L yields on Tails, and on Tails yields what L yields on Heads. (For the more technically inclined, this is a "probability-preserving statewise permutation.") The twin of L^+ yields A^+ if the coin lands Heads, and B^+ if the coin lands Tails. This twin is better in every state than L: it is better if the coin lands Heads (A^+ vs. A), and it is better if the coin lands Tails (B^+ vs. B). Since the twin is better in every state (it "state-wise dominates"), in a choice between L and the twin of L^+, it is rationally required to take the twin. Moreover, one should be indifferent between a lottery and its twin, so if it is rationally required to take the twin of L^+ in a choice between it and L, it is rationally required to take L^+ in a choice between L^+ and L. (This version of the argument is perhaps a bit closer to the use of SWITCH by Rabinowicz [2021, p. 207] than to Hare's original, but the difference won't matter here.)

	L	L^+	L^*
Heads	A	B^+	$A+$
Tails	B	A^+	B^+

This argument cannot be used to motivate the claim that Mira is required to play the Hard Game, or to motivate the rejection of Negative Dominance. Unlike Hare's example, Mira's problem is not sensitive to which 'states' we associate with which prizes. (To say it loud and clear: in my view, this is the most important difference between the examples.) We can't produce a game which is better in every state by permuting which outcome of the Hard Game Mira gets in which state. The Hard Game's twin (which yields (4, -2) on Tails and (-2, 4) on Heads) is no different from the Hard Game with respect to our problem; it too only has prizes which are not preferred to the status quo of $(0, 0)$.[11] So Hare's argument does not

[11] Similar points as those made in the main text apply to striking puzzles developed by Doody [2019b, 2021]. There, Doody provides a remarkable argument that fans of incompleteness must reject his "Principle of Predominance", and more strongly the principle NEVER BETTER, LIKELY WORSE, which says that one is permitted not to choose a game of chance if it is better in no state, and worse in states whose collective probability is greater than $\frac{1}{2}$. Doody's target examples contradict his principles, but probability-preserving statewise-permutations of them do not. Someone convinced of Doody's conclusion could thus still attempt to explain it away by claiming that the counterexamples are artifacts of the assignments of outcomes to states. In our richer setting, however, we can produce a counterexample to Doody's principles which does not have this feature. Consider a game of chance which gives probability $\frac{1}{2} + \epsilon$ to $(0, 0)$ and $\frac{1}{2} - \epsilon/2$ each to (9, -3) and (-3, 9). Assuming the latter are incomparable to $(1, 1)$, this lottery would be never better, and likely worse than $(1, 1)$. But its expectation is $(1.5 - 3\epsilon, 1.5 - 3\epsilon)$, which for small enough ϵ will be better than $(1, 1)$. So Expectationalism (and, through a more involved route, Unidimensional Expectations and Independence) would commit one to rejecting NEVER WORSE, LIKELY BETTER for reasons which go beyond Doody's. In this case, they cannot explain their rejection of the principle on the grounds that it arises from treating probability-preserving statewise permutations equivalently, and there is an equivalent lottery which does not directly give rise to the counterexample. It seems to me that this is an (even) worse result than Doody's extremely surprising one.

show that Mira should choose the Hard Game, or that we should give up Negative Dominance.

In fact, as I'll lay out in more detail at the end of this section, there are independent arguments for Negative Dominance, which are compatible with the claim that it is rationally required to take L^+ in Hare's case. So even those who accept Hare's arguments for preferring L^+ have good reason to endorse Negative Dominance.[12]

Let us now consider one of Hare's arguments that one is permitted to take L (and thus not rationally required to take L^+). The one I'll discuss turns on the following principle (restated to fit the terms of this paper):

> RECOGNITION: Whenever I have two options, and in every state, I would not prefer the prize of the one to the prize of the other, it is rationally permissible for me to take either.[13]

This principle again depends on what happens in which states. In each state, the prizes (outcomes) that L and L^+ yield are incomparable (that is, the decision-maker has no preference between them). So, RECOGNITION implies that it is permissible to take L.

[12] Hare's other argument for the conclusion that we should take L^+, because there are reasons to take L^+ that aren't reasons to take L, and no reasons to take L, which aren't reasons to take L^+, is harder to assess in our context, given widespread controversy about what counts as a reason, and what reasons people have. Suffice it to say, though, that whereas it's obvious that the "sweetening" is a reason to take L^+ and not a reason to take L, it's much less clear that there's a reason for Mira to take the Hard Game which isn't a reason to stick with what she has. Other arguments have a similar relationship to our puzzle. Bader [2018] argues that we should prefer the stochastically dominant of two options, and hence accept L^+. His argument also does not imply a rational requirement in our examples, since none of our lotteries stochastically dominates any of the others. I discuss this a bit more around n. 20, below. Finally, Rabinowicz [2021] gives an account of an axiological version of Hare's puzzle, providing a further argument that L^+ is better than L. He shows (§6) that in every preference which completes the given incomplete preferences on A, $A+$, B, $B+$, there will be at least as great ordinal distance between the "good" outcome of L^+ and the "bad" outcome of L than between the good outcome of L and the bad outcome of L^+, and that some completions will have a greater distance between the first pair than the second. (For instance: in the completion $A+ \succ A \succ B+ \succ B$, the ordinal distance between A^+ and B is greater than the ordinal distance between A and $B+$, intuitively because there are "more steps" between the former pair than between the latter.) Rabinowicz argues that if we understand incompleteness of betterness facts as due to the permissibility of diverse complete preference orders, then this fact about ordinal distances in the completions implies that L^+ is better than L. But no argument analogous to this one can be given for the betterness of the Hard Game. At one point, Rabinowicz suggests that there is an independent, *prima facie* case against an axiological version of Hare's RECOGNITION (see below in the main text) because of the complexity of the condition on completions of preferences which would be required to vindicate it (p. 214). But the analogous condition on completions that would vindicate Negative Dominance is not correspondingly complex. In our particular case, it would suffice to require that: if there are completions which rank $(0, 0)$ higher than $(-2, 4)$, and completions which rank $(0, 0)$ higher than $(4, -2)$, there are also completions which rank $(0, 0)$ higher than both $(-2, 4)$ and $(4, -2)$.

[13] See also Schoenfield [2014, p. 267]'s "LINK" Bales et al. [2014, p. 460]'s "COMPETITIVENESS", Rabinowicz [2021, p. 203]'s "Complementary (Statewise) Dominance".

This argument (and Hare's other one) has a very different relation to our puzzle than the first one. The first one did not settle which of our two key principles should be rejected. But the main premise of this argument implies that Mira is rationally permitted not to take the Hard Game, and to stick with the status quo. So, given the claim that if Mira strictly prefers one of two options, she is rationally required to choose it in a pairwise choice between them, the conclusions of these arguments are flatly incompatible with the combination of Unidimensional Expectations and Independence, since they imply a strict preference for the Hard Game.

We can make this point even more vivid. I could have developed my argument using a slightly different principle than Negative Dominance (replacing talk of required preferences, with talk of permitted actions), namely:

Negative Dominance (Action) If you have two options, and you have no preference between any prize of the one (regardless of what state it occurs in), and any prize of the other (regardless of what state it occurs in), then it is rationally permissible to take either.

This alternative principle is straightforwardly entailed by Hare's RECOGNITION. On its intended interpretation, using the language of states, Hare's principle licenses either of two actions when, in every state, the prize the first act yields *in that state* is not preferred to the prize the second yields *in that same state*. Negative Dominance (Action) licenses either of two actions when every prize of each action is not preferred to *any* prize of the other (regardless of which states they occur in). The latter condition is strictly more demanding than the former (if there's no preference between prizes regardless of what state yields them, then there's no preference between prizes which are given in the same states). So Hare's principle is stronger. And it's strictly stronger as well, since his game is precisely a case where RECOGNITION entails a permission but Negative Dominance (Action) does not. The fact that Negative Dominance is weaker than Hare's principle in this way is the central new feature of our puzzle.

Since those who accept Hare's arguments for RECOGNITION must endorse Negative Dominance (Action), they have in some sense a ready-made answer to Mira's problem. But Mira's problem still shows us something new about their view. Our puzzle here essentially implies that Negative Dominance (Action) commits one to the rejection of either Unidimensional Expectations or Independence. So it presents those who accept Hare's RECOGNITION with a new challenge: to develop a systematic theory of choice under uncertainty which either does not entail Unidimensional Expectations or does not entail Independence. As I'll discuss in a bit more detail in §7 (see n. 20), without these principles (and without others that RECOGNITION also forces us to abandon), it's very hard to know where to start.[14]

[14] Hare develops "deferentialism" as a systematic theory vindicating RECOGNITION, but in our setting, his theory implies very striking violations of Unidimensional Expectations,

I have emphasized that RECOGNITION entails Negative Dominance (Action) and so in a way is straightforwardly stronger than Negative Dominance—or at least the thought behind it. But this might make one wonder: is there any motivation for accepting Negative Dominance that isn't a motivation for accepting RECOGNITION?

There are. One of these is an argument I gave at the start. The main premise of that argument was that 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 any of the prizes of the second. This says nothing about which *states* the relevant prizes occur in. It thus gets us nowhere near the full strength of RECOGNITION. The latter principle (unlike Negative Dominance) applies to cases like Hare's, where it's only the prizes which occur in the same states that can't be compared. In those

---

which seem to me implausible.

Hare takes lotteries to be functions from a non-empty (and for our purposes, finite) set of states S to outcomes (prizes) O. (Actually, he does something even more sophisticated, with dependency hypotheses, but the difference won't matter here.) We assume in the background a probability p defined on S. Hare says that a utility function u *represents a coherent completion of* a transitive reflexive relation $\succeq$ if and only if, if $o \succ o'$ then $u(o) > u(o')$ and for all lotteries $L \in O^S$, $u(L) = \sum_{s\in S} p(s)u(L(s))$. Letting U be the set of functions which represent coherent completions of $\succeq$, a *regimentation* R of U is a subset of U which assigns some outcomes o, o' 0 and 1 respectively. (Here the aim is just to ensure that the functions are normalized to a common scale.) Hare's key idea is to consider, for a regimentation R, (what I will call) its "state-expansions", where a state-expansion $f : S \to R$ is a function from states to utility functions in the regimentation R. Such a state expansion delivers an expected value for every lottery, as $\sum_{s\in S} p(s)f(s)(L(s))$ (recall that $f(s)$ will be an element of U). But there are more state-expansions than there are coherent completions: the state-expansions allow us to "mix and match" coherent completions, choosing a different one for each state. In our terms Hare's deferentialism is:

DEFERENTIALISM It is permissible for an agent to choose a lottery if and only if, for some regimentation, R, of the set of utility functions that represent the agent's preferences, for some state-expansion f of R, no alternative has higher expected f-utility.

Idealizing for mathematical convenience, suppose the space of outcomes is $O = \mathbb{R}^2$ and $\succeq$ is defined so that $(x, y) \succeq (x', y')$ iff $x > x'$ and $y > y'$. Then all linear combinations of x and y (i.e. $u(x, y) = ax + by + c$, with $a, b > 0$) will be utility functions that represent coherent completions of the $\succeq$ relation. The regimentation R consisting of functions u such that $u((1, 1)) = 1$ and $u((0, 0)) = 0$ will at least include all functions u of the form $u((x, y)) = ax + by$ where $a, b > 0$ and $a + b = 1$. Now suppose $S = \{s_1, s_2\}$ with $p(s_1) = p(s_2) = \frac{1}{2}$. Suppose the agent faces a choice between L_1 and L_2 where $L_1(s_1) = (0, 100)$, $L_1(s_2) = (0,\text{-}1)$ and $L_2(s_1) = L_2(s_2) = (0, 0)$. Intuitively, it seems to me, choosing L_1 should be rationally required for someone like Mira. (If these numbers don't work for you, increase 100 however much you like.) But Hare's deferentialism does not deliver this result. Let u_1 be defined so that $u_1((x, y)) = (1 - \frac{1}{1000})x + \frac{1}{1000}y$, and u_2 be the function so that $u_2((x, y)) = \frac{1}{1000}x + (1 - \frac{1}{1000})y$, and let f be the state expansion defined so that $f(s_1) = u_1$ and $f(s_2) = u_2$. The f-expected utility of L_1 is $\frac{1}{1000} * 100 + (1 - \frac{1}{1000}) * \text{-}1 = .1 - .999 = -.899$, while the f-expected utility of L_2 is 0, so that it is permitted to choose L_2. More generally, for any three outcomes o_1, o_2, o_3, all unidimensional with each other, and such that $o_1 \succ o_2 \succ o_3$ there will be *no* lottery supported on o_1 and o_3 which it is rationally required to choose over a lottery supported only on o_2.

cases, Negative Dominance doesn't apply because some prizes, which occur in *different* states are comparable.

There's also a second motivation for Negative Dominance, that was implicit in my description of why Mira's behavior would be bizarre, if Expectationalism is true. There I said it would be bizarre if in a three-way choice Mira is rationally permitted to choose any of the certain options (a) (the status quo), (b) (four marbles gained, two matchsticks lost), (c) (four matchsticks gained, two matchsticks lost), but, in a two-way choice between (a) and a coin toss over (b) and (c), she is rationally required to choose the coin toss. More generally, say that a choice function—which maps sets of options to the subset of them which can be permissibly chosen—is *stochastically contractible* if, whenever it maps a set of certain options O to itself (so that all options are permitted, when all are on the table), it also maps any pair consisting of (i) one of the options in O and (ii) a lottery over some subset of O to itself (so that both options in the pair are permitted, when only two are on the table). The very plausible idea that rational preferences should determine a stochastically contractible choice function implies Negative Dominance. But it does not imply RECOGNITION.

So there are good reasons to accept Negative Dominance, even for those who endorse the claim that it is rationally required to take L^+, and who, in the face of Hare's puzzle, have come to terms with rejecting RECOGNITION.

To sum up: those who accept either of Hare's first two arguments (for the requirement to choose L^+) face a new choice-point here. Their endorsement of those arguments does not settle whether they should endorse Negative Dominance or Independence, and the present puzzle shows that they must choose. Those who accept one of Hare's second two arguments (for the permission to choose L), by contrast, are thereby committed to a principle which is very close to Negative Dominance. But the present conflict shows that, surprisingly, they must reject Unidimensional Expectations or Independence. The challenge they face is to provide a more systematic decision theory.

6

I now turn to responses to our problem beginning, in this section, with the possibility of upholding Independence, and rejecting Negative Dominance.

Before I say anything substantive about this approach, I want to reiterate and emphasize that Independence does not follow from Unidimensional Expectations, or the motivation for it. Unidimensional Expectations was motivated on the basis of two ideas: first, that Mira values matchsticks linearly, if we hold her stock of marbles fixed (and similarly for marbles, if we hold matchsticks fixed); and, second, that Mira not averse or prone to risk. These facts on their own say nothing about how Mira values games of chance over prizes that vary in both marbles and in matchsticks. As I'll show in the next section more formally, Unidimensional Expectations is consistent with denying Independence. Our two main assumptions—Independence and Negative Dominance—are independent of our background assumptions.

Perhaps the main argument in favor of upholding Independence (aside from its intuitive appeal) is that this approach allows us to avail ourselves of a well-developed framework for handling decisions under uncertainty, using sets of utility functions. Some hold that it is a rational requirement that a person's preferences can be represented by a set of utility functions, in the sense that they prefer one option to another if and only if every utility function in the relevant set accords the first at least as great expected utility (Seidenfeld et al. [1995], Shapley and Baucells [1986], Dubra et al. [2004], Nau [2006], Özgür Evren [2008], Evren and Ok [2011], Ok et al. [2012], Galaabaatar and Karni [2012, 2013], Riella [2015], Gorno [2017], McCarthy et al. [2021], Borie [2023]). Any person whose preferences can be represented by a set of utility functions in this way will satisfy Independence. So, given Unidimensional Expectations, they must violate Negative Dominance. Indeed, Unidimensional Expectations, together with the assumption that Mira's preferences can be represented by such a set of utility functions, implies Expectationalism (Lederman [2023, Proposition 4.2]).

The availability of this strong theory is an abductive reason for accepting Independence (and thereby for rejecting Negative Dominance). But it does not show us *why* Negative Dominance should be rejected, or undermine the appeal of this principle. To put it another way, there is no non-circular deductive argument from this theory to Independence: the arguments I know of that a rational person's preferences must be representable by a set of utility function all use Independence as a premise.[15]

What does it mean, then, to reject Negative Dominance? Earlier, I formulated this principle just as applying to Mira. A more general version (which I've been implicitly assuming) would say:

Negative Dominance It's rationally required that: if a person does not prefer games of chance for reasons other than the preferability (or not) of their prizes, then if they strictly prefer one game of chance over another, they must prefer one of the first game's prizes to one of the second game's prizes.

The antecedent of this expanded principle is supposed to apply to cases of people who prefer not to make a decision, or love coin flips and so have a preference for a game of chance "as such". But it is also intended to apply to those who may value other global features of games of chance, for instance, those who prefer games when their outcomes are relevantly

[15]Building on his discussion of opaque sweetening, Hare develops a formal theory, which he calls "prospectism", to vindicate the idea that a rational decision-maker is required to choose L^+. This theory amounts to the claim that a decision-maker should be representable by a set of utility functions. While I think this may be a reasonable place to end up, it does imply Independence, which Hare doesn't isolate or discuss. Mira's problem shows that this assumption is far from obvious in our context.

symmetric, so that they always prefer, say, a game over $(4, -2)$ and $(-2, 4)$ to a game over $(5, -1)$ and $(0, 3)$.

This fuller version of Negative Dominance casts our puzzle in a slightly different light. Fans of Independence *can* accept this new principle, even in Mira's case. But if they do, they must reject the claim that Mira overall state of mind is permissible. They will say that, if a person has preferences like Mira's, *and* satisfies Unidimensional Expectations, then they are required to prefer games of chance for reasons other than the preferability (or not) of their prizes. In other words, they will say that people with incomplete preferences must have preferences based on global features of lotteries.

Schoenfield [2014] criticizes those who reject (a version of) Hare's RECOGNITION on the grounds that they require "us to make choices that we are certain would lead to no improvement in value" and thus are "imposing requirements that transcend what we actually care about: the achievement of value" (p. 268). She accuses them of an "expected-value fetish". Bader [2018, §2.2] responds to this charge, showing that, since the requirement to take L^+ (and thus the rejection of RECOGNITION) are entailed by Stochastic Dominance alone (see next section), believing in such a requirement does not require an expected-value fetish. I'm convinced by Bader in Hare's original case. But I think that Schoenfield's diagnosis was prescience, since it applies exactly to Mira's problem, even if it doesn't apply to Hare's. Preserving Independence in response to Mira's puzzle does require that one endorse a rational injunction for Mira (and others like her) to value games of chance for reasons other than the preferability of their outcomes. It requires a fetish for expected-value.

This point is a bit abstract, so I want to bring it out by considering a putative counterexample to Negative Dominance, and showing why it fails. I'll spend a bit of time on the example because it's interesting in its own right, but the main goal is to explain how it could be that fans of Independence require people to value games of chance as such.

The example is inspired by contractualists who endorse an "*ex ante* Pareto" principle, stating that one is required to choose an action if it is better for each person in expectation than the alternatives (see, e.g. Diamond [1967], Frick [2015], with Taurek [1977], and against this, Broome [1984], and many others).[16] To make it concrete, I'll assume that we can speak of numbers associated with welfare levels. We'll understand the status quo for two people, Xander and Yara, as 0 (this needn't be "the neutral level"; it's just how their lives are now). In a first decision, we can choose either (a) to stick with the status quo of 0 for both people; (b) produce a situation where Xander has a welfare level of -2 and Yara a welfare level of 4; or (c) produce a situation where Xander has a welfare level of -2, while Yara has a welfare level of 4. Contractualists will say that we are not required to choose either (b) or (c) in this three-way choice (in fact, they

[16] A different kind of example, related to critical level utilitarianism (as discussed in e.g. Gustafsson [2020], Thornley [2022]), is discussed in Lederman and Spears [2023].

may say that we are not even permitted to choose them): if we chose (b) or (c), the person who has a welfare level of -2 would have a legitimate complaint against us, since we made them worse off than they could have been. But contractualists who endorse *ex ante* Pareto will hold that, if we instead have a choice between (i) the status quo, and (ii) a fifty-fifty lottery between (b) and (c), then we are required to choose the lottery: since it gives each individual a benefit in expectation (the expected value is 1 for Xander and 1 for Yara), no individual can complain, and it's better for each of them. If we reframe this talk of permitted and required actions in terms of preference, and think of each individual as one of our dimensions, it seems that the pattern is exactly what Expectationalism recommends. We are required to prefer the Hard Game (which is equivalent to a lottery over (b) and (c)), even if we don't have a preference over a three-way choice between (a), (b), or (c). So it might seem that the contractualist has not only bought into violations of a generalization of Negative Dominance, they also have a good explanation for why it fails.

This putative counterexample is, I think, only apparent. There are two ways this kind of contractualism might be understood, and on neither understanding does it yield a counterexample to the fuller version of Negative Dominance. First, a contractualist might holds that there is a morally-relevant value associated with lotteries as such (when they're fair). If contractualism is understood in this way, Negative Dominance won't apply non-trivially to people whose preferences are exactly in line with moral value. They will value games of chance for reasons which don't have to do with the prizes of those games, but have to do with global properties (in this case, they fact that they make it so that no one has a complaint). Second, a contractualist might hold that there is a morally-relevant value associated with how distributions over welfare-levels are obtained. As a result, outcomes ("prizes") can't be represented just by pairs of numbers assigned to Xander and Yara. When we obtain $(4, \text{-}2)$ as a result of our deciding to harm Yara and benefit Xander, Yara is entitled to complain. But when we obtain this result through a chance process, no one can complain. A perspicuous representation of these outcomes would include this difference. But once it's included, it's clear that the example is not a counterexample to Negative Dominance: $(4, \text{-}2)$ *when obtained through a fair lottery* is better (preferred) to $(0, 0)$. So the preference for the lottery in this case is, after all, explained by a preference for its prizes.

Both of these ways of understanding contractualism offer a way of responding to the axiological version of the puzzle in cases where dimensions are understood as the welfares of different individuals. But neither of them offers a promising way of responding to every putative instance of the axiological puzzle. Abstract dimensions like achievement or happiness don't have complaints. So instances of the puzzle that arise from pluralism about well-being or value in general can't obviously be explained away using the contractualist machinery. And, more importantly, the example doesn't provide a counterexample to Negative Dominance, properly understood.[17]

What it does show (and this was our main reason for discussing it) is how an "expected-value fetish" might allow us to escape the puzzle. Our first kind of contractualist claimed that morality "cares" about global features of lotteries, assigning greater expected value of the game its own distinctive value. In much the same way (though more surprisingly), fans of Independence claim that there is a much more general rational requirement that: if one has incomplete preferences like Mira's, values marbles and matchsticks linearly and is neutral with respect to risk, one must value global properties of lotteries for their own sake (the expected value). This seems to me a very surprising, almost magical result. Why must one care about these global features of the lotteries, when what one cares about is not the expected value, but the values one will get in the end? The view which upholds Independence not only has this strange consequence. It also implies the (to me) counterintuitive verdict that Mira must strictly prefer the Hard Game. For me, these two reasons are more than enough to wonder whether this view can be right. I'm open to the idea that it may be the correct view in the end, but, at this point, it is hardly the obvious choice.[18]

7

This brings us, next, to the possibility of rejecting Independence. As I said earlier, beyond the intuitive plausibility of Independence, perhaps the most important argument in favor of it is an abductive one, based on the fact that it's implied by the claim that rational people's preferences can be represented by a set of utility functions.

Can proponents of Negative Dominance provide a comparably strong

[17] There is also some evidence of fairly widespread ethical preferences *against* randomizing, though the examples aren't exactly analogous to ours: see Meyer et al. [2019].

[18] By squinting at the theory of Levi [1986] we can develop a theory of incomparability, which uses the machinery of sets of utility functions, but which has a very different overall shape. Levi himself aims to develop a theory of how one should choose when one is uncertain about the true objective "value function", and is seeking to make decisions in the face of that uncertainty. But if we think of his value-functions instead as (something like) dimensions of preference, we can use his framework in our context. At its most general level, Levi's proposal, v-max, says that an agent is permitted to choose an option only if it is the best of her options with respect to one of her value-functions. Levi further holds that value-functions must be mixture-preserving (like expectational utility functions) and, moreover, that if two value-functions are in the agent's set of value-functions, any linear combination of them is also in. If Mira's value-functions are all and only linear combinations of our dimensions (i.e. functions of the form $v((x, y)) = ax + by + c$, with $a, b > 0$), this implies that: (i) in a pairwise choice between $(-2, 4)$ and $(0, 0)$, Mira is permitted to choose either; (ii) in a pairwise choice between $(0, 0)$ and $(4, -2)$, Mira is permitted to choose either; but (iii) in a three-way choice between $(-2, 4)$, $(0, 0)$, and $(4, -2)$, Mira is required to choose the first or the third. Given this (iii), it is perfectly reasonable that Mira is required to prefer the Hard Game over $(0, 0)$. But (iii) seems to me implausible given (i) and (ii). If Mira is permitted to choose $(0, 0)$ in both pairwise choices, why wouldn't she be permitted to choose it when all three are on the table?

theory for choice under uncertainty? I have some good news on this front. Consider the following principle, which implies Unidimensional Expectations:

Good Expectations It's rationally required for Mira that: if every prize in a game of chance is comparable for her (i.e. preferred or dispreferred) to every other, then Mira is indifferent between the game of chance and its expected value.

To see why this principle is stronger than Unidimensional expectations, consider what I'll call the Easy Game. In this game, if a fair coin lands Heads, Mira will get four matchsticks and four marbles; if it lands Tails Mira will give away two of both.

The Easy Game

	Matchsticks	Marbles
Heads	4	4
Tails	-2	-2

Unidimensional Expectations says nothing about this game. But Good Expectations directly implies that Mira should be indifferent between it and its expected value, since every prize ($(4, 4)$, $(\text{-}2, \text{-}2)$) is comparable to its expected value ($(1, 1)$). Like Unidimensional Expectations, however, Good Expectations does not entail that Mira should be indifferent between the Hard Game and *its* expected value. It sees a difference between the Easy Game and the Hard Game, in spite of the fact that their expected values are exactly the same, because the prizes in the latter game are not comparable with its expected value. This seems to me the right result: the Easy Game is an obvious choice; it's the Hard Game that's hard.

The good news is that Good Expectations is consistent with Negative Dominance and the basic shape of Mira's preferences. In fact, it is consistent with them, even if we add Stochastic Dominance, that is, roughly, the claim that it's a rational requirement that if for each outcome o, L offers at least as great a probability of outcomes Mira prefers to o, and there are some outcomes o' so that L offers a greater probability of outcomes Mira prefers to o', then Mira strictly prefers L (Lederman [2023, Proposition 5.1]).[19]

[19] "Roughly", because, as Russell [forthcomingb] observes, this standard definition of Stochastic Dominance goes awry when incompleteness is in play. A more exact characterization is as follows. A *lottery* is a function from some set of outcomes O to probabilities. A *generalized* lottery is a set $X \subset O \times [0, 1]$. A generalized lottery L^* is equivalent to a lottery L iff for every outcome o, $\sum_{x \in L^*, \pi_1(x) = o} \pi_2(x) = L(o)$ (where π_1 and π_2 are projection operations on pairs, taking the first and second coordinate of an ordered pair, respectively, so that $\pi_1((x, y)) = x$, and $\pi_2((x, y)) = y$.). A lottery L stochastically dominates a lottery L' iff there is a generalized lottery L^* equivalent to L, a generalized lottery L'^* equivalent to L', and a bijection f between them, such that for all $x \in L^*$ $\pi_2(x) = \pi_2(f(x))$ and $\pi_1(x) \succeq \pi_1(f(x))$, and there is some $x \in L^*$ such that $\pi_1(x) \succ \pi_1(f(x))$.

In fact an even stronger principle is consistent with our package. A lottery L existen-

In my view, this is an important result, for at least two reasons. First (least important), it shows formally that Unidimensional Expectations is independent of Independence, since Unidimensional Expectations holds in this theory but Independence does not. Second, the result confirms a claim I made earlier about the possibility of motivating Negative Dominance independently from Hare's RECOGNITION. As Bader [2018] emphasizes, Hare's L^+ stochastically dominates L. So, an appropriate generalization of Stochastic Dominance implies that it is a rational requirement that Hare's decision-maker take L^+, and thus implies the negation of RECOGNITION. The present consistency result shows that this important argument, just like others mentioned earlier, does not force Mira to choose the Hard Game.[20] Third, and most importantly, it shows that Negative Dominance is at least in principle compatible with a strong, simple theory for making decisions under uncertainty. This fact brings proponents of Negative Dominance closer to an even footing with proponents of Independence, who will naturally hold that rational people's preferences are representable by a set of utility functions, and thereby have a strong theory of how people with incomplete preferences should make decisions. The consistency result shows that proponents of Negative Dominance have some hope of a comparably general theory.

I think this is all good news, and I've sometimes been attracted to seeing the puzzle of this paper as an argument for the claim that Independence must fail in new ways if preferences can be incomplete. But I myself am not (yet?) convinced; here I'll say why.

A first issue is that I'd like to have a broader story about rationality which jointly implies Negative Dominance, Stochastic Dominance and Good Expectations. The result above is a consistency result, but the axioms it proves consistent aren't a unified theory. Perhaps relatedly, the axioms also aren't quite as domain-general as the idea that preferences should be representable by a set of utility functions, since Good Expectations requires that the notion of an "expectation" be defined, and so really only works in a setting like that of marbles and matchsticks. Without a more domain-general theory for fans of Negative Dominance, there is still *some* abductive support for Independence.

A second issue—which could be seen, in a way, as the flip-side of the first—concerns the violations of Independence. Any plausible theory that reconciles Negative Dominance and Unidimensional Expectations by rejecting Independence must either hold that the Die Roll is not indifferent to $(0, 0)$, or that the Improved Die Roll is not indifferent to the Hard

tially stochastically dominates a lottery L' iff there is a generalized lottery L^* equivalent to L, a generalized lottery L'^* equivalent to L', and a bijection f between them, such that for all $x \in L^*$ $\pi_2(x) = \pi_2(f(x))$ and either $\pi_1(x) \succeq \pi_1(f(x))$ or $\pi_1(x)$ is incomparable with $\pi_2(x)$, and there is some $x \in L^*$ such that $\pi_1(x) \succ \pi_1(f(x))$. We can also show that adding a corresponding rational requirement is consistent with our other principles.

[20] By the same token, since Stochastic Dominance is incompatible with RECOGNITION, the theory does not help resolve the challenge that I articulated for proponents of that principle: to develop a systematic theory of choices under uncertainty.

Game. But rejecting either of these claims strikes me as implausible. Before I accept this theory, I'd want to understand why we should *expect* Independence to fail in this way.

Unfortunately, I don't see how the proponent of Negative Dominance can take advantage of the venerable tradition of rejecting independence on the basis of the Allais paradox (Allais [1953], cf. Machina [1982], Buchak [2013]). The main reason is that even fans of the rationality of risk aversion typically do not think that rationality *requires* risk aversion. They say that rational people are permitted to be risk-neutral, and thus permitted to satisfy Independence. But an agent with incomplete preferences who is risk-neutral and satisfies Independence will run afoul of Negative Dominance. So in response to the present puzzle, we would need a story about why incompleteness (in the multidimensional setting) *requires* failures of Independence even for intuitively risk-neutral agents, who satisfy Unidimensional Expectations. It's not clear how the precedent of theories of risk aversion can help.[21]

So accepting Negative Dominance is not yet obviously on a better footing than accepting Independence. The consistency result shows that there is some hope for developing a theory which endorses Negative Dominance. But until we know whether that hope bears fruit, it's hard to bet on Negative Dominance.

8

The core of this paper is a new problem for preferences which are sensitive to different dimensions of prizes in such a way as to be incomplete. The problem suggests that we must either reject a new dominance-like principle, Negative Dominance, or that we should reject Independence. I've focused on the stylized example of Mira's love for marbles and matchsticks, but

[21] It is still interesting to ask whether the formal tools developed by proponents of risk-aversion might help us replace Independence with a restricted but still powerful principle. Unfortunately, the theory which is most prominent for philosophers—that of Buchak [2013]—crucially uses the notion of an outcome's *rank* in the agent's preference order, and thus is not obviously well-defined in our setting, where preferences are incomplete (so that outcomes can be tied in ordinal rank without being indifferent). Considered on its own Buchak's Comonotonic Sure-Thing Principle (p. 107; which, although developed in a Savage Framework, can be re-interpreted in ours) would not license either of the key steps for which Independence was used in the argument of section 4. But it's unclear whether this principle can be embedded in a reasonably strong theory which is compatible with incompleteness, given the importance of the rankings of outcomes to the formal implementation of her theory.

Recently, Chris Bottomley and Timothy Williamson (Bottomley and Williamson [forthcoming]) have advocated a different sort of theory, which does not require a complete ranking of outcomes. Their principle of Betweenness licenses the conclusion that the Die Roll is indifferent to $(0, 0)$, but it does not license the conclusion that the Improved Die Roll is indifferent to an even lottery over $(4, \text{-}2)$ and $(\text{-}2, 4)$. I don't yet know (though I'd be very curious to know!) whether Betweenness can be consistently added to the package described above. But even if it can be, this will not give us the explanation I think we should hope for, of why Independence fails here.

the problem is much more general. Indeed, I've argued that it extends not only to a wide array of rational preferences but to problems in the theory of well-being and overall goodness as well.

My own impulse is to see the puzzle as elucidating surprising features of the structure of incomplete preferences and betterness. If rational preferences or betterness can be incomplete, then one of Unidimensional Expectations, Negative Dominance, and Independence must be false. I hope that the challenge of discovering which one of these principles fails, and why, will help to deepen our understanding of the structure of incompleteness.

Throughout the paper I've been assuming that rational preferences can be incomplete. I started by saying that, at least in some cases, it seems that we don't have preferences between two options, and that's not because we're ignorant of what we prefer, or because our minds are divided. But one response to the puzzle is to see it as evidence against the claim that rational preferences can be incomplete owing to their sensitivity to different dimensions of preference. Indeed, since arguably the best motivation for the rationality of incomplete preferences is the idea that preferences can be sensitive to different dimensions in this way, one might also see it as an argument against incompleteness itself.[22]

Officially, I'm open to this response. I do believe the puzzle should make us at least a little more confident that it is rationally required to have complete preferences. But I want to close with a few, more speculative remarks about why I myself don't see it this way. My main reason is roughly that I don't see the puzzle as strong evidence that betterness or betterness-for-a-person is complete, and I think that if betterness or betterness-for-a-person are incomplete, then clearly it's not a rational requirement that preferences be complete.[23]

[22]Mira's puzzle might be added to other considerations—e.g. Gustafsson [2022]'s moneypump, or Dorr et al. [forthcoming], Dorr et al. [2021]'s linguistic arguments—which tell against the rationality or even possibility of incomplete preferences (though cf. Eva [manuscript] on the latter). The axiological version might be added to considerations—e.g. those of Broome [1997, 2021] or again of Dorr et al. [forthcoming, 2021]—which tell against the possibility of incomplete betterness.

[23]Bradley [forthcoming] suggests a different sort of argument. Suppose we accept the view of Worsnip [2021] that "a set of attitudinal mental states is jointly incoherent iff it is (partially) constitutive of (at least some of) the states in the set that any agent who holds this set of states has a disposition, when conditions of full transparency are met, to revise at least one of the states" (p. 133). Then we might take the fact that people do not seem disposed to "fill in" their preferences as evidence that completeness is not a structural requirement (i.e. a requirement of coherence), and hence not a requirement at all.

Another route to this conclusion might be based on disanalogies between completeness and other structural requirements. Both the putative requirement to have a complete confidence ranking, and the requirement to have complete preferences, are "wide-scope" putative requirements. But there is a difference between these putative wide-scope requirements and more typical examples, for instance, the requirement that if one believe that p, and believe that if p then q, then one believe q, or the requirement that if one believe that taking an action is the best feasible way to achieve what one most desires, then one intend to take that action. Those who violate these latter, paradigmatic requirements (and recognize that they do) can remedy their situation by dropping one or

First, then, let us suppose that betterness (for a person, or overall) is incomplete, and imagine that Mira knows all the facts about what's better than what, knows that she knows this, and faces a choice in which there is nothing special at stake for her or others except what's better (for her, or overall). To adapt our stylized example: she knows that gaining four marbles at the expense of two matchsticks is not better or worse or the same for her as what she has now. But if it is a rational requirement to have complete preferences, then apparently she must, even so, have preferences between these options.

This seems to me odd. If Mira does not have a preference between these, it is unclear whether she could come to have one, given that she knows that there's no reason to prefer one to the other. Even supposing that she could come to have such a preference, perhaps by bringing a reason beyond those offered by objective betterness into existence (Chang [2002]), it is unclear what reasons we could give to convince her to bring this reason into existence. Indeed, if we recommend that she come to have such a preference, she has quite a compelling reason to reject that recommendation: she knows that neither of the options is better than the other. If this is good enough for objective betterness, it is good enough for her.

So, if betterness is incomplete, I find it hard to understand how there could be a further rational requirement for completeness of preferences.[24] As a result, the puzzle in the present paper should make us more confident that preferences are required to be complete only if it should make us more confident that betterness is complete.

This conclusion itself brings out something important about rational requirements, which may be worth pausing on. In the terminology of "structural" vs. "substantive" requirements, completeness seems to fall on the "structural" side if it falls anywhere. It is a requirement of "coeherence" on the pattern of attitudes we have, not a requirement on how our attitudes respond to the world or our evidence about it. It is often held that structural requirements of this kind have less of a relationship to the way the world is, than so-called "substantive" requirements. But the present example highlights that this cannot mean that they do not have any such relationship at all. As we have seen, it is a requirement that preferences be complete, only if betterness is also complete. This is not a unique feature of the requirement of completeness. For instance, the requirement not to believe p while also believing not p also depends on the claim that there are

more of the attitudes stated in its antecedent. It has often been held that we are more able to directly control whether we stop believing something we already believe than to directly control whether we begin believing something we do not yet believe. Whether or not that's true, it does seem somehow that the reasons for ceasing to believe can be somehow more unspecific than the reasons for having one. The key point is then that, unlike paradigmatic wide-scope requirements, there is no way to satisfy the completeness requirements by giving up an attitude. This difference is some evidence against it being a genuine requirement.

[24] I agree with Worsnip [2021, §9.2] that money-pumps don't provide the right kind of reason for such a requirement.

no true contradictions, so that it's never true that p and $\not p$. But the case of completeness brings out the connection between the requirement and the world especially clearly, perhaps because, whereas it's obvious that contradictions can't be true, it's not at all obvious whether betterness is complete.

Still, we are left with the question of whether the puzzle in this paper should make us substantially more confident that betterness itself is incomplete. Those who think it should may still see our puzzle as supporting a rational requirement of completeness. But I myself have two reasons for thinking that it shouldn't.

First, as we saw in §6, contractualism offers an ethical worldview according to which the axiological version of the puzzle does not arise, either because lotteries can have moral value in and of themselves, or because a distribution of well-being brought about via a lottery may have different value than if it were brought about by a certain choice. If this view is correct, betterness can be incomplete, without giving rise to any particularly deep problem: the contractualist has a clear explanation of how our principles all hold in their case. (As I've said, the corresponding way of resolving the puzzle about preferences in cases like Mira's, which concern 'dimensions' of value, not people—by holding that those with incomplete preferences are rationally required to value lotteries as such, strikes me as much less plausible.)

Second, I am not particularly moved by an abductive argument that holds that, since this puzzle would not arise if betterness were complete, we have reason to believe that it is. The main reason is that I think, metaphysically, the property of betterness is in the first instance a property of certain outcomes—whether those are states of the world, situations, lives, or whatever—not a property of games of chance. There is clearly a sense in which one game of chance may be better for a person than another, but this is not the same sense in which one life (for instance) is better than another. It is the friendship, happiness, achievement, or love in a life, that are better for a person in the first instance, not the expected friendship, happiness, achievement or love, which a lottery might provide. The idea that betterness as a property of lives or states of the world should be simple and logically well-behaved is compelling to me. But I am much less moved by the idea that betterness as a property of lotteries need be well-behaved. It is no objection to a theory of gravity that it does not tell us how choose in the face of uncertain options, if we happen to want to bring about as much gravity in the world as we can, in the face of our uncertainty. We should not be misled by the fact that it is natural to use the same word "better" to apply to lotteries and lives or states of affairs to think that betterness is importantly different. So, while I do believe it is *some* evidence for the completeness of betterness that it would allow us to maintain all three of our plausible-seeming principles for betterness, I don't see this as a strong argument for completeness, since it concerns in part the simplicity of the relationship between betterness as a property of certain options, and betterness as a property of uncertain ones.

As I've said, all of this is much more speculative than what's gone before. My goal has been to present the puzzle, and I myself am not sure what to think. The options I've mentioned seem to me better along some dimensions, and worse on others. As a result, perhaps irrationally, I don't have a preference among them. But I hope that, in this case more analysis will lead to a resolution.

References

M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546, 1953.

Ralf Bader. Choice under incompleteness. Unpublished Book MS, 2023.

Ralf M Bader. Stochastic dominance and opaque sweetening. *Australasian Journal of Philosophy*, 96(3):498–507, 2018.

Adam Bales, Daniel Cohen, and Toby Handfield. Decision theory for agents with incomplete preferences. *Australasian Journal of Philosophy*, 92(3):453–470, 2014.

Dino Borie. Expected utility in Savage's framework without the completeness axiom. *Economic Theory*, 76(2):525–550, 2023.

Christopher Bottomley and Timothy Luke Williamson. Rational risk-aversion: Good things come to those who weight. *Philosophy and Phenomenological Research*, forthcoming. doi: 10.1111/phpr.13006.

Richard Bradley. Review of *Fitting Things Together: Coherence and the Requirements of Structural Rationality* by Alex Worsnip. *Economics and Philosophy*, pages 1–6, forthcoming.

John Broome. Uncertainty and fairness. *The Economic Journal*, 94(375):624–632, 1984.

John Broome. Is incommensurability vagueness? In Ruth Chang, editor, *Incommensurability, Incomparability, and Practical Reason.* Harvard University Press, 1997.

John Broome. Incommensurateness is vagueness. In Henrik Andersson and Anders Herlitz, editors, *Value Incommensurability: Ethics, Risk, and Decision-Making*, pages 29–49. Routledge, 2021.

Lara Buchak. *Risk and rationality.* Oxford University Press, 2013.

Ruth Chang. The possibility of parity. *Ethics*, 112(4):659–688, 2002.

Ruth Chang. Parity: An intuitive case. *Ratio*, 29(4):395–411, 2016.

Peter A Diamond. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of political economy*, 75(5):765–766, 1967.

Ryan Doody. Opaque sweetening and transitivity. *Australasian Journal of Philosophy*, 97(3):559–571, 2019a.

Ryan Doody. Parity, prospects, and predominance. *Philosophical Studies*, 176: 1077–1095, 2019b.

Ryan Doody. Hard choices made harder. In Henrik Andersson and Anders Herlitz, editors, *Value Incommensurability: Ethics, Risk, and Decision-Making*, pages 247–266. Routledge, 2021.

Cian Dorr, Jacob M. Nebel, and Jake Zuehl. Consequences of comparability. *Philosophical Perspectives*, 35(1):70–98, 2021. doi: 10.1111/phpe.12157.

Cian Dorr, Jacob M. Nebel, and Jake Zuehl. The case for comparability. *Noûs*, forthcoming. doi: 10.1111/nous.12407.

Juan Dubra, Fabio Maccheroni, and Efe A Ok. Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1):118–133, 2004.

Benjamin Eva. The case for incomparability. Unpublished MS, manuscript.

Özgür Evren and Efe Ok. On the multi-utility representation of preference relations. *Journal of Mathematical Economics*, 47(4-5):554–563, 2011.

Johann Frick. Contractualism and social risk. *Philosophy and Public Affairs*, 43 (3):175–223, 2015. doi: 10.1111/papa.12058.

Tsogbadral Galaabaatar and Edi Karni. Expected multi-utility representations. *Mathematical Social Sciences*, 64(3):242–246, 2012.

Tsogbadral Galaabaatar and Edi Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

Leandro Gorno. A strict expected multi-utility theorem. *Journal of Mathematical Economics*, 71:92–95, 2017. ISSN 0304-4068.

Johan E. Gustafsson. Population axiology and the possibility of a fourth category of absolute value. *Economics and Philosophy*, 36(1):81–110, 2020. doi: 10.1017/s0266267119000087.

Johan E. Gustafsson. *Money-Pump Arguments*. Cambridge: Cambridge University Press, 2022.

Caspar Hare. Take the sugar. *Analysis*, 70(2):237–247, 2010.

Caspar Hare. *The limits of kindness*. Oxford University Press, 2013.

Brian Hedden and Daniel Muñoz. Dimensions of value. *Noûs*, 2023.

Harvey Lederman. Multidimensional preferences, incompleteness, and negative dominance. University of Texas at Austin, June 2023.

Harvey Lederman and Dean Spears. A note on different number incompleteness. In progress, 2023.

Isaac Levi. *Hard choices: Decision making under unresolved conflict*. Cambridge University Press, 1986.

Mark J. Machina. "Expected utility" analysis without the independence axiom. *Econometrica*, 50(2):277–323, 1982.

David McCarthy, Kalle Mikkola, and Teruji Thomas. Expected utility theory on mixture spaces without the completeness axiom. *Journal of Mathematical Economics*, 97:102538, 2021.

Michelle N. Meyer, Patrick R. Heck, Geoffrey S. Holtzman, Stephen M. Anderson, William Cai, Duncan J. Watts, and Christopher F. Chabris. Objecting to experiments that compare two unobjectionable policies or treatments. *Proceedings of the National Academy of Sciences*, 116(22):10723–10728, 2019.

Robert Nau. The shape of incomplete preferences. *The Annals of Statistics*, 34 (5):2430 – 2448, 2006.

Efe A Ok, Pietro Ortoleva, and Gil Riella. Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica*, 80(4):1791–1808, 2012.

Özgür Evren. On the existence of expected multi-utility representations. *Economic Theory*, 35(3):575–592, 2008. ISSN 09382259, 14320479.

Wlodek Rabinowicz. Incommensurability meets risk. In *Value Incommensurability*, page 201. Routledge, 2021.

Gil Riella. On the representation of incomplete preferences under uncertainty with indecisiveness in tastes and beliefs. *Economic Theory*, 58(3):571–600, 2015.

Jeffrey Sanford Russell. On two arguments for fanaticism. *Noûs*, forthcominga.

Jeffrey Sanford Russell. Fixing stochastic dominance. *The British Journal for the Philosophy of Science*, forthcomingb.

Miriam Schoenfield. Decision making in the face of parity. *Philosophical Perspec-*

tives, 28:263–277, 2014.

Teddy Seidenfeld, Mark J. Schervish, and Joseph B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23:2168–2217, 1995.

Lloyd S. Shapley and Manel Baucells. Multiperson utility. Tech Rep. 779 UCLA Department of Economics, 1986.

Katie Steele. Incommensurability that can (not) be ignored. In *Value Incommensurability*, page 231. Routledge, 2021.

John M Taurek. Should the numbers count? *Philosophy & Public Affairs*, pages 293–316, 1977.

Elliott Thornley. Critical levels, critical ranges, and imprecise exchange rates in population axiology. *Journal of Ethics and Social Philosophy*, 22(3):382–414, 2022. doi: 10.26556/jesp.v22i3.1593.

John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1944.

Alexander Worsnip. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. Oxford University Press, New York, 2021.