Of marbles and matchsticks*

Harvey Lederman

Submitted: September 2, 2023
Accepted: December 11, 2023
Corrected version: June 4, 2024

Abstract

I present a new puzzle about choice under uncertainty for agents whose preferences 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 preferences are rationally required to be complete.

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 recompense, she'll angrily refuse.

Mira loves all 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, Pietro Cibinel, Adam Elga, Nathan Engel-Hawbecker, Kit Fine, Jane Friedman, Johan Gustafsson, Caspar Hare, Ben Holguín, Nico Kirk-Giannini, Jon Kleinberg, Sebastian Liu, Tiankai Liu, Aidan Penn, Richard Pettigrew, Wlodek Rabinowicz, Dean Spears, Una Stojnič, Christian Tarsney, Teru Thomas, and Patrick Wu for conversations and correspondence; to audiences at Melbourne, MIT, NYU, TARK, and the Global Priorities Workshop for questions and comments; to Ralf Bader, Sam Carter, Cian Dorr, Jeremy Goodman, Brian Hedden, and Jake Nebel for detailed comments on drafts; and finally to Chiara Damiolini for several key conversations.

Still, there are 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 say that she doesn't prefer the trade to what she has, and that she doesn't prefer what she has to the trade. It's also not, it seems to her, that she's exactly indifferent between them. If she were, and you slightly sweetened the deal, then that would tip the balance: she would prefer the sweetened deal to what she has. But she finds that she doesn't. Of course she'd prefer to get *five* matchsticks in exchange for two marbles, than to get four matchsticks at the same price. But in a choice between the five-two trade and what she has now, she'd still find herself stuck. She'd say that she doesn't prefer the trade to what she has, and also doesn't prefer what she has 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 hard choices of a different kind.

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, it seems that 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 small 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.

This paper develops a new puzzle for preferences of this kind, preferences which are sensitive to various dimensions of options, in such a way as to be *incomplete*: they don't rank every option with respect to every other. 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 (§3 and §4). (As I discuss at the end of §4, a version of the puzzle arises in axiology as well.) I'll explain how the puzzle differs from the phenomenon of 'opaque sweetening', described by

Caspar Hare (§5), and then consider the ways forward for fans of incomplete preferences: either giving up what I call "Negative Dominance" (§6), or endorsing a new theory 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 these connections, as I consider to what extent 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 away 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 this game to what she has, or not? There's a plausible argument on either side.

I haven't told you much about Mira but it's important to know now, before we go on, that her interest in marbles and matchsticks doesn't particularly 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 these facts, Mira should 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 values marbles and matchsticks linearly, and isn't averse or prone to risk, 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. Since Mira prefers this certain gain to what she has now (a free marble! a free matchstick!), she should prefer the game to what she's got.

Our second argument is that, given how she is, Mira should not prefer the game. If Mira prefers one game of chance to another, then since she doesn't care for games of chance as such, her preference for the game must be explained in part by a preference for one of its prizes. If she did not prefer any of the prizes of the first game to *any* of the prizes of the second, what could explain her preference for the first? Now in fact, Mira does not prefer any of the prizes of the Hard Game to sticking with what she has (a 'game of chance' with just one prize). As I told you before, 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. Given that she does not prefer either of the the Hard Game's prizes to what she has, a preference for the Hard Game would be inexplicable. Since she should not have inexplicable preferences, she should not prefer the Hard Game.

The arguments reveal a conflict between two claims:

Expectationalism: It's rationally required that: if Mira prefers a certain gain of a particular bundle of marbles and matchsticks to her present holdings, she 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 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.[1]

I've said that there's a puzzle here, and ultimately, as I'll describe in the next section, I think there is. 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. It's pretty plausible, just on its face, that Mira is *permitted* not to prefer the Hard Game; that's a first strike against Expectationalism. But more than this: 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) giving up two matchsticks to get four marbles; or (c) giving up 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 (and she isn't indifferent either). But now, while she's stuck, we swoop in, and take this choice off the table. In its place, we offer her a choice between two options: the first is just (a) from before, sticking with what she has; the second 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, she didn't have a preference in our first three-way choice, and wouldn't have a preference in choices

[1]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 to every prize in another, then she prefers the first. The contrapositive of Negative Dominance says that it's rationally required that if there are no prizes in one game that Mira prefers to any of the prizes in another, then she does *not* prefer the first.

over any of its pairs. So a preference for the game would be 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 might be. 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 Hard Game involves chance. The pattern of preferences that Expectationalism predicts on its own isn't weird. What's weird is that Mira displays this pattern when she doesn't care for coin-flips as such. It's in her case that the preference would be baseless, not in all.

On top of all this, there's an easy diagnosis of where Expectationalism might have gone off the rails. The expected value of the Hard Game 'forgets' how marbles and matchsticks were distributed across its prizes. It doesn't record that in this game, Mira will get a high value for marbles only at the price of a low value for matchsticks, and vice versa. The expectation of this game is the same as that of one where Mira would win four marbles and four matchsticks on Heads, and lose two of each on Tails. But the structural difference between these games matters; it shouldn't be forgotten. 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.

If we think of matchsticks and marbles as 'dimensions' of the prizes, the idea is to restrict attention to *unidimensional* games, where a game is unidimensional if all of its prizes yield the same number of matchsticks (and so differ only in the number of marbles they yield), or if all of its prizes yield the same number of marbles (and differ only in the number of matchsticks they yield). In this special case, given how she is, Mira should treat the games as equivalent to their expected yield:

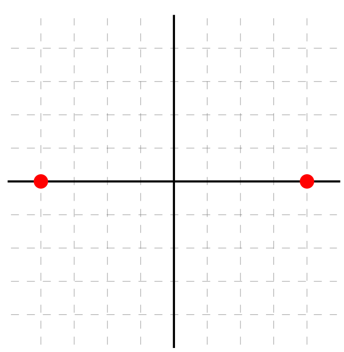
Unidimensional Expectations: It's rationally required that: if a person values marbles and matchsticks linearly, and is not averse or prone to risk, then they are indifferent between any unidimensional game and a certain gain of its expected value in marbles and matchsticks.

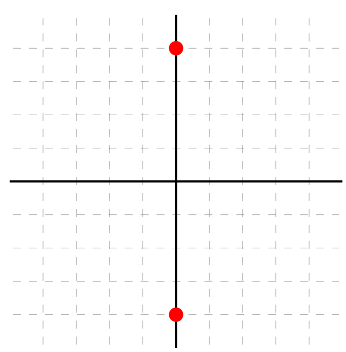
I've already told you that Mira values marbles and matchsticks linearly and is neutral about risk. Pretty much everyone agrees—and I'll assume for now—that it's rationally permitted to do so. So I'll understand Unidimensional Expectations to imply that it's rationally required for Mira herself, given how she is, to be indifferent between unidimensional games and a certain gain of their expected value.[2]

Unidimensional Expectations (as well as the consequence I just mentioned) is *much* weaker than Expectationalism. It says nothing about how to value games whose prizes are not 'unidimensional' with each other, so it says nothing about the Hard Game. The principle is also directly motivated in a way that Expectationalism isn't. In fact, it's basically true by definition: 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.

But with Unidimensional Expectations in the background, Independence—a plausible, standard principle that I'll state in a moment—implies a requirement that Mira prefer the Hard Game, and so rules out Negative Dominance. It's this result that convinced me that Mira's choice presents a genuine puzzle, that there's a real question about what she should prefer.

In presenting the argument for this result, it will be helpful to have some diagrams. I'll represent prizes 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 the diagrams each dot will be understood to occur with equal probability. Below, for instance, we have two unidimensional games of chance. On the left is a game where Mira would get four matchsticks with probability $\frac{1}{2}$ and would lose four matchsticks with probability $\frac{1}{2}$. I'll call it "Only Matchsticks", since only matchsticks are at stake. 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 $\frac{1}{2}$. This one is "Only Marbles", since only marbles are at stake.





By Unidimensional Expectations, Mira should be indifferent between each

[2] 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 one rejects this general principle, it's natural to just assume that this requirement holds for Mira here, and I'll sometimes use "Unidimensional Expectations" to refer to such a requirement.

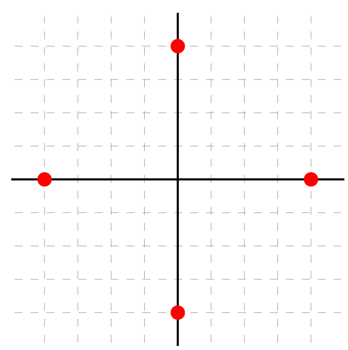
of these two games and her present holdings (which aren't represented, but would be a single dot at $(0, 0)$, the origin).

If Mira is indifferent between her present holdings and each of these two games, she should also be indifferent between her present holdings and a new game, which combines the old ones in an appropriate way. Suppose we offer Mira 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 play Only Matchsticks: flipping a second coin, and awarding prizes as in the figure on the left (it won't matter which is Heads or Tails). If the coin comes up Tails, we play Only Marbles: flipping a second time, and awarding prizes as in the figure 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 isn't important. As I said above, Mira doesn't care for coin-tosses as such, so it doesn't matter to her that a sequence involves a certain number of flips. If a game that happens all at once offers the same chances of the same prizes that Mira would get in our sequence, Mira should be indifferent between the game that happens all at once, and the sequence. And so, she should also be indifferent between the game that happens all at once, and what she presently has. We can create such an "all-at-once" game, which offers the same chances of the same prizes as our sequence, using a fair four-sided die. In this game, which I'll call "the Die Roll", if our die comes up One, we give Mira four matchsticks; if Two, she gives us four; if Three, we give her four marbles; if Four, she gives us four marbles. (The game is represented in the following table and picture.) We've just seen that Mira should be indifferent between what she has and this new game.

The Die Roll

	(Matchsticks, Marbles)
One	(4, 0)
Two	(-4, 0)
Three	(0,4)
Four	(0,-4)



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, so I'll state a general principle that entails the claim that Mira should be indifferent between the Die Roll and what she has. Using "weakly prefers" for "prefers or is indifferent between", we can derive this claim using the following principle, a version of which has been standard since Von Neumann and Morgenstern [1944]:

Independence It's rationally required that: if a person is not averse or prone to risk, they weakly prefer a game of chance A to a game of

> chance B if and only if, for any p between 0 and 1, and any game C, they weakly prefer the p-mixture of A with C (a game offering p times A's chances of A's prizes and $1-p$ times C's chances of C's prizes) to the p-mixture of B with C (which offers p times B's chances of B's prizes and $1-p$ times C's chances of C's prizes).

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. Consider first (as we did above) two sequences of games. In the A/C-sequence, we flip a coin with a p chance of Heads and, if it comes up Heads we play A, while if it's Tails, we play C. In the B/C-sequence, we flip the same coin, but if it's Heads we now play B, and if it's Tails we still play C. Plausibly, Mira should weakly prefer A to B if and only if she prefers the A/C-sequence to the B/C-sequence. The only difference between the two sequences is that the first has a fixed chance of yielding A, where the second has the same chance of yielding B. So if Mira weakly prefers A on its own to B on its own, she should weakly prefer the A/C-sequence to the B/C one. And similarly in the other direction: If Mira weakly prefers the A/C-sequence to the B/C-sequence, it must be because she weakly prefers A to B, since a change of A for B is the only difference between them.

Independence extends this claim about the *sequences* to "all-at-once" games that give the same chances of the same prizes. The p-mixture of A with C is an all-at-once game that gives the same chances of the same prizes as the A/C-sequence. And the p-mixture of B with C similarly gives the same chances of the same prizes as the B/C-sequence. Given that Mira doesn't care for coin-flips as such, she should be indifferent between these mixtures and the corresponding sequences. And since, as we saw, Mira should weakly prefer A to B if and only if she weakly prefers the A/C-sequence to the B/C-sequence, she should also weakly prefer A to B if and only if she prefers the p-mixture of A with C to the p-mixture of B with C—which is just what Independence says.[3]

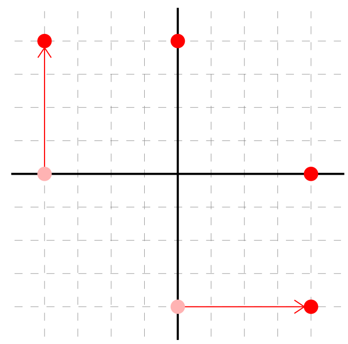
Unidimensional Expectations and Independence imply that Mira should be indifferent between the Die Roll and what she has. Accordingly, if we improve the Die Roll, she should prefer the improved game to her current holdings. So let's now do just that, improving the game in two different ways. First, if the die comes up Two, we'll still take away four matchsticks, but now we'll also give Mira four marbles in recompense. Since getting four marbles makes the prize better in a way and keeps it at least as good in every other way, this is an improvement in the prize, and so, an

[3]To see how Independence implies that Mira should be indifferent between the Die Roll and what she has, note first that it implies that, since Mira is indifferent between $(0,0)$ and Only Marbles, Mira should also be indifferent between $(0,0)$ and a $\frac{1}{2}$-mixture of $(0,0)$ and Only Marbles, that is, a game which offers a $\frac{1}{2}$ probability of $(0,0)$ and a $\frac{1}{4}$ probability of each of $(0,\text{-}4)$ and $(0,4)$. Given that Mira is also indifferent between $(0,0)$ and Only Matchsticks, Independence implies that Mira should be indifferent between a $\frac{1}{2}$-mixture of $(0,0)$ with Only Marbles, a game which offers a $\frac{1}{2}$ chance of $(0,0)$ and a $\frac{1}{4}$ chance of each of $(0,\text{-}4)$ and $(0,4)$, on the one hand, and, on the other, the Die Roll, a $\frac{1}{2}$-mixture of Only Matchsticks and Only Marbles, which offers $\frac{1}{4}$ chance of each of: $(0,\text{-}4)$, $(0,4)$, $(\text{-}4,0)$, and $(4,0)$. Since she should be indifferent between $(0,0)$ and the former, she must be indifferent between $(0,0)$ and the latter, as well.

improvement in the game. Second, if the die comes up Four, we'll still take away four marbles, but now we'll also give Mira four matchsticks in recompense. Once again, since this change makes this prize better in one way while keeping it at least as good in every other way, it is an improvement in the prize, and so, an improvement in the game. The table and picture below show the changes and the new game.[4]

The Improved Die Roll

	(Matchsticks,Marbles)
One	(4,0)
Two	(-4,4)
Three	(0,4)
Four	(4,-4)



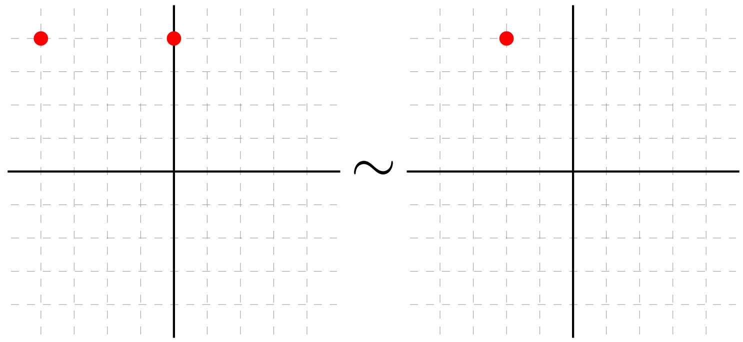
In general, if a person should prefer a first game to a second and should be indifferent between the second and a third, they should prefer the first to the third. Mira should prefer the Improved Die Roll to the Die Roll, and she should be indifferent between the Die Roll and what she has. So, she should prefer the Improved Die Roll to what she has.

We're now just one step away from the promised conflict. The last step is to show 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.

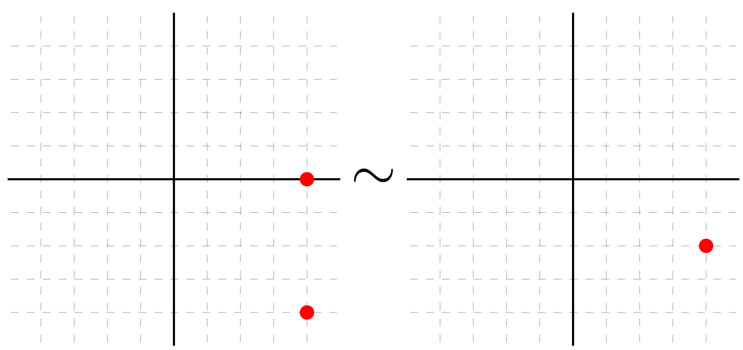
The argument for this last step can be broken into three parts. The first begins with a fair coin-flip—I'll call it the "Top Left Flip"—over the outcomes shown in the top left of the previous diagram, with coordinates $(\text{-}4, 4)$ and $(0, 4)$, which stand (in the first case) for Mira losing four matchsticks while gaining four marbles, and (in the second) for Mira keeping all her matchsticks while gaining four marbles. The Top Left Flip is a unidimensional game, since both of its prizes agree in yielding four marbles. So, by Unidimensional Expectations, Mira should be indifferent between it and a certain gain of its expected value, that is, a certain gain of four marbles at the cost of two matchsticks (since $\frac{1}{2} \cdot \text{-}4 + \frac{1}{2} \cdot 0 = \text{-}2$). You can see all this below, where the left-hand figure depicts the Top Left Flip, the figure on the right depicts a game that yields the expected value of the Top

[4] Formally, we can justify this step using Independence, but it's so plausible taken on its own that I won't belabor the point. (In fact the weaker Stochastic Dominance would do as well (for a statement, see n. 19).) A detailed proof of the main result here can be found in Lederman [2023].

Left Flip with certainty, and the symbol $\sim$ indicates that Mira should be indifferent between them.

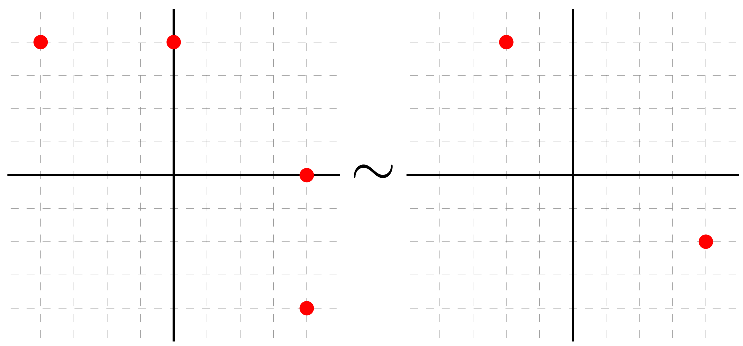


The second part of the argument begins with a similar coin-flip—let's call it the "Bottom Right Flip"—over the two outcomes depicted at the bottom right of the figure showing the Improved Die Roll. These outcomes have coordinates $(4, 0)$ and $(4, \text{-}4)$, which stand (in the first case) for Mira's gaining four matchsticks while keeping all her marbles, and (in the second) for Mira's gaining four matchsticks, while losing four marbles. The Bottom Right Flip is a unidimensional game, because all of its prizes agree in yielding four matchsticks. So by Unidimensional Expectations, Mira should be indifferent between it and a certain gain of its expected value, that is, a certain gain of four matchsticks at the cost of two marbles (since $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 0 = 2$). You can again see all this in the next figure, where the left-hand figure depicts the Bottom Right Flip, and the right hand figure depicts a game which yields the expected value of the Bottom Right Flip with certainty.



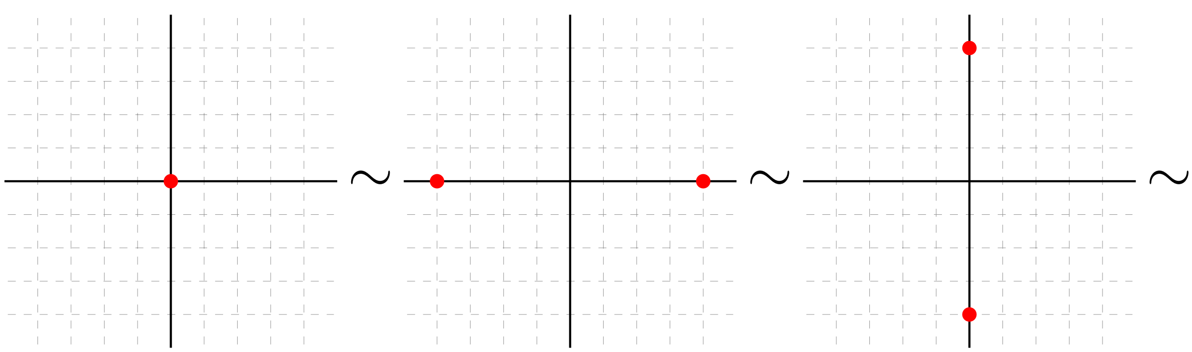
The third part of this step uses these facts, together with Independence, to conclude, as promised, that Mira should be indifferent between the Improved Die Roll and the Hard Game. (The next diagram shows the Improved Die Roll on the left, and the Hard Game on the right.) Independence implies that it's rationally required that if a person is (i) neutral with respect to risk, (ii) indifferent between a game of chance A and a game a, and (iii) indifferent between a game of chance B and a prize b, then they are indifferent between any p-mixture of A with B (on the one hand) and the same p-mixture of a with b (on the other). We've seen (first part) that

Mira should be indifferent between the Top Left Flip and a certain outcome of (-2, 4). We've also seen (second part) that Mira should be indifferent between the Bottom Right Flip and a certain outcome of (4, -2). So, by the above consequence of Independence, Mira should be indifferent between a $\frac{1}{2}$-mixture of the Top Left Flip with the Bottom Right Flip (on the one hand), and a $\frac{1}{2}$-mixture of (-2, 4) with (4, -2). These two mixtures are, in the first case, equivalent to the Improved Die Roll, and, in the second case, equivalent to the Hard Game. So, Mira should be indifferent between the Improved Die Roll itself, and the Hard Game.[5]

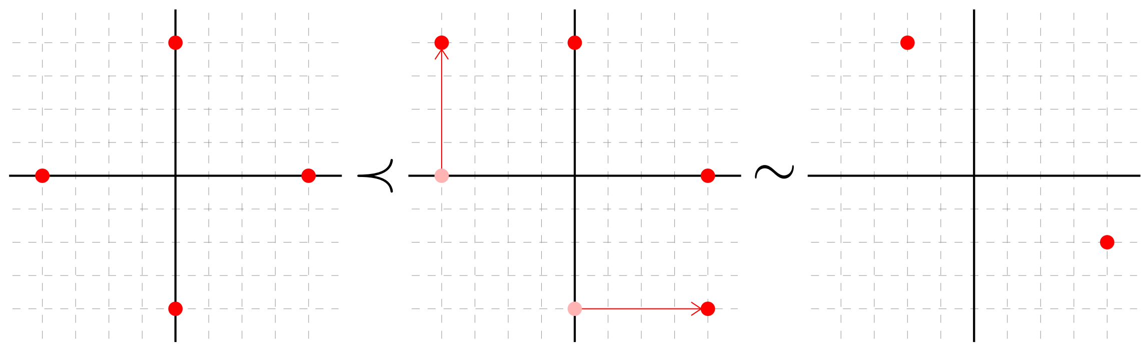


This completes the argument for the promised conflict. Mira should be indifferent between the Hard Game and the Improved Die Roll, and should prefer the Improved Die Roll to her current holdings. So she should prefer the Hard Game to her current holdings, in violation of Negative Dominance.

The whole argument can be summarized in a single picture (with $\prec$ indicating preference):



[5] In more detail, the argument is as follows: given that Mira is indifferent between (-2, 4) and the Top Left Flip, Independence implies that she should be indifferent between the Hard Game, which gives an even probability of (-2, 4) and (4, -2), on the one hand, and, on the other, the $\frac{1}{2}$ mixture of the Top Left Flip with (4, -2), that is, 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 Mira is indifferent between (4, -2) and the Bottom Right Flip, by Independence she should be indifferent between a $\frac{1}{2}$ mixture of (4, -2) with the Top Left Flip, that is a game which yield (4, -2) with probability $\frac{1}{2}$ and (-4, 4) and (0, 4) each with probability $\frac{1}{4}$ (on the one hand), and (on the other) the Improved Die Roll, which is a $\frac{1}{2}$ mixture of the Bottom Right Flip and the Top Left Flip, a game yields (4, -4), (4, 0), (-4, 4), and (0, 4) each with probability $\frac{1}{4}$. Since she's indifferent between the former and the Hard Game, she should be indifferent between the latter and the Hard Game as well.



Unidimensional Expectations and Independence imply that Mira is required to prefer the Hard Game, and so, that Negative Dominance must be rejected. This is the main puzzle of the paper.

In fact, a conflict can be generated with even weaker premises. 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. Moreover, while I've assumed the permissibility of valuing marbles and matchsticks linearly, and of being neutral with respect to risk, there are more general mathematical results, which weaken these assumptions greatly (Lederman [2023, Propositions 2.4 and 2.5, §2.5])). And the problem can be extended to any number of 'dimensions'; it's not restricted to cases with just two.

Beyond preferences, a version of the puzzle arises in axiology as well. Many are *pluralists* about what is better or worse for individuals, or better or worse overall. They hold that what is good for a person (or overall) is sensitive to different dimensions of value, for instance, perhaps, knowledge, friendship, happiness, achievement, love. Some pluralists hold that tradeoffs across these different dimensions can lead to incompleteness in what's better for a person, or in what's better overall, so that the relevant form of betterness doesn't rank some pairs of outcomes (Raz [1985, 1986], Chang [1997, 2002], Hedden and Muñoz [2023]). Such pluralists will face a version of our puzzle for the relevant notion of betterness. And if their notion of betterness (for a person, or overall) gives rise to a corresponding (prudential or moral) "ought", a corresponding puzzle arises for what one "ought" (in that sense) to do.

To make this general point concrete (though the point is, really, quite general), suppose that instead of marbles and matchsticks, we understand our dimensions as achievement and happiness. (Achievement and happiness don't come in numbered units, but let's assume that they do here; in Lederman [2023, Appendix], I show how to do away with the assumption.) With achievement on the x axis and happiness on the y, we can take Mira's *status quo* life to be $(0, 0)$ and consider again a version of the Hard Game: a coin-flip between (-2, 4) (a life with less achievement, but greater happiness, which we assume is not better or worse, or equally good for her), and (4, -2) (a life with greater achievement, but less happiness, which again is not better, nor worse, nor equally good). Then the following three principles can't all be true:

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

Unidimensional Expectations (Goodness): If the prizes in a game of chance vary only in a single dimension of value (for instance, achievement), and are exactly alike in every other dimension, then the game is exactly as good for Mira as a certain gain of its expected value.

Independence (Goodness) A game of chance A is at least as good for Mira as a game of chance B if and only if, for any p between 0 and 1, and any game C, the p-mixture of A with C is at least as good for her as the p-mixture of B with C.

Much more could be said here, but I'll leave it at that. I'll turn now, in the next section, §5, to a task already deferred too long: of showing how Mira's problem differs from the phenomenon of "opaque sweetening", discovered by Hare [2010] (for discussion see Hare [2013], Schoenfield [2014], Bales et al. [2014], Bader [2018], Doody [2019a,b, 2021], Rabinowicz [2021], Steele [2021], Bader [2023], Russell [forthcominga, §3.2], Fine [2024]).

I then discuss two responses to the puzzle, primarily with the aim of arguing that the way forward isn't clear. I don't know of a plausible theory which says that it is rationally forbidden to value marbles and matchsticks linearly, while at the same time being neutral with respect to risk, and, anyway, as I've said, claiming that this is forbidden isn't enough to escape the problem. So I'll focus here on the prospects of denying either Negative Dominance or Independence, while upholding our other assumptions. In §6 I discuss how an approach based on sets of utility functions implies Independence (and thus rules out Negative Dominance), but say why I'm not completely satisfied with this approach. In §7, I develop a novel, strong theory that rejects Independence, but again express some doubt about whether it too is ultimately correct. In §8, at last, I come to the "nuclear option": of denying the rationality of incomplete preferences altogether.

5

To introduce the difference between Mira's problem and Hare's, I'll first lay out Hare's example. In this example, there are four prizes, A, A^+, B, and B^+, with A^+ preferred to A, B^+ preferred to B, and no preference (or indifference) between the As and Bs. (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^+. In each, a coin will be flipped. In the first, L, Heads will yield A, while Tails will yield B. In the second, L^+, Heads will yield B^+, and Tails will yield A^+. The games are shown in the following table.

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

Hare presents two arguments that it is rationally required to choose L^+, and two arguments that it is rationally permitted to choose L (and thus not required to choose L^+). To make my case that Mira's problem is different from Hare's, I'll discuss one each of these pairs of arguments, and argue that endorsing it would not on its own provide a resolution to Mira's problem.[6]

I'll start with an argument that it is rationally required to take L^+. If a game of chance G gives prizes depending on the *states* Heads and Tails, each of which occurs with probability $\frac{1}{2}$, its *twin* is a game that on Heads yields what G yields on Tails, and on Tails yields what G yields on Heads. The twin of L^+, which I'll call L^*, is depicted in the table below: it yields A^+ if the coin lands Heads, and B^+ if the coin lands Tails. In every "state" this twin gives a prize that is preferred to the prize of L: it is preferred if the coin lands Heads (A^+ vs. A), and preferred if the coin lands Tails (B^+ vs. B). Since L^* is preferred in every state (it "state-wise dominates"), Hare argues that in a choice between L and L^*, it is rationally required to take L^*. Moreover, given that one should be indifferent between a lottery and its twin, one should be indifferent between L^+ and L^*, and so, it is rationally required to take L^+ in a choice between L^+ and L.

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

This argument does not support the claim that Mira is required to play the Hard Game (or, for that matter, to stick with what she has). Unlike Hare's problem, the problem posed by the Hard Game is not sensitive to which 'states' we associate with which prizes. (In my view, this is the key difference between the examples.) In particular, permuting which outcome of the Hard Game Mira gets in which state does not produce a game which is preferred in every state. Just like the Hard Game, the Hard Game's twin (which yields $(4,\text{-}2)$ on Tails and $(\text{-}2,4)$ on Heads) also only has prizes which are not preferred to what she has. So this argument of Hare's cannot show that Mira should choose the Hard Game, or, for that matter, that we should reject Negative Dominance.[7]

[6]Manzini and Mariotti [2008, pg. 310] independently identified a fact closely related to Hare's puzzle, when they remarked that, given a suitably rich domain of outcomes, Stochastic Dominance (see below n. 19) conflicts with their "Vagueness Sure Thing" principle. This Vagueness Sure Thing principle is stronger than (a natural formal version of) Hare's RECOGNITION, which he uses to argue that it is permitted to take L. My comments below about the relationship between RECOGNITION and Negative Dominance will apply *mutatis mutandis*, to this Vagueness Sure Thing principle.

[7]Related points apply to arguments developed by Doody [2019b, 2021], that fans of incompleteness must reject (among other things) the principle "Never Better, Likely

As I'll describe in more detail at the end of this section, there are 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 argument for taking L^+ have good reason to endorse Negative Dominance. This argument for a requirement to take L^+ does not settle whether Mira should prefer the Hard Game.[8]

Let's now turn to an argument 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 (which I've restated to fit the terms of this paper):

RECOGNITION Whenever I have two options, and in every state, I would not prefer the prize the one option yields in that state to the prize the other yields in that state, it is rationally permissible for me to take either.[9]

In each state, the decision-maker does not prefer the prize L yields in that state to the prize L^+ yields in that same state. So, RECOGNITION implies that it is permissible to take L. (To foreshadow a bit, Negative Dominance does not imply this conclusion, because there *are* prizes of L^+ that are preferred to prizes of L; they just occur in different states.)

This argument has a very different relationship to our puzzle than the first one we considered. That first one did not settle which of our two key principles should be rejected. But the main premise of this second argument implies that, when given a choice between what she has and the Hard Game, Mira is rationally permitted to stick with what she has. So, given the claim that if Mira prefers one of two options, she is rationally required to choose it when choosing just between those two, the conclusion of this argument is incompatible with the combination of Unidimensional Expectations and Independence, since together those principles imply a strict preference for the Hard Game.

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

Negative Dominance (Action) Whenever a person has two options, and has no preference between any prize of the one (regardless of

Worse", which states 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 counterexample to this principle is sensitive to how prizes are assigned to states. But in the present setting, we can produce counterexamples to it which are not. Consider for instance a game of chance which gives probability $\frac{1}{2}+\epsilon$ to $(0,0)$ and $\frac{1}{2}-\epsilon/2$ each to $(9,\text{-}3)$ and $(\text{-}3,9)$. Assuming the latter are not better than (preferred 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 $\epsilon < \frac{1}{6}$) will be better than (preferred to) $(1,1)$. So Expectationalism (and, through a more involved route, Unidimensional Expectations and Independence) would commit one to rejecting Never Better, Likely Worse for reasons which go beyond Doody's.

[8]The same goes for other prominent arguments for this conclusion. Neither Hare's second argument, nor appropriate versions of the arguments from Bader [2018] or Rabinowicz [2021], settle whether Mira should prefer the Hard game.

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

> what state it occurs in), and any prize of the other (regardless of what state it occurs in), then it is rationally permissible for them to take either.

This alternative principle is straightforwardly entailed by (a general version of) Hare's RECOGNITION. 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). But if there's no preference between prizes regardless of what state yields them (so that Negative Dominance (Action) comes into play), then there's no preference between prizes which are given in the same states (so that RECOGNITION comes into play): every case where Negative Dominance (Action) applies is a case where RECOGNITION does as well. And there are cases—like Hare's own example—where RECOGNITION applies, but Negative Dominance (Action) does not. So Negative Dominance (Action) is strictly weaker than RECOGNITION.

Since those who accept RECOGNITION must endorse Negative Dominance (Action), they have a ready-made response to Mira's problem. In fact, their response to Hare's problem already commits them to rejecting Independence.[10] So in a sense, Mira's problem offers this second camp nothing new; its novelty depends on what it offers the first (who endorse a requirement to take L^+).

But the framework we've developed to present Mira's problem does offer something new, even for this second camp. In this setting, RECOGNITION is in conflict not just with Independence, but also with Unidimensional Expectations just on its own. To me, this fact provides a powerful new argument against RECOGNITION. The conflict can be seen by considering the following two games:

	Only Matchsticks (Matchsticks, Marbles)	Only Marbles$^+$ (Matchsticks, Marbles)
Heads	(-4,0)	(0,-2)
Tails	(4,0)	(0,4)

Only Matchsticks, which we saw at the start of §4, is a unidimensional game with expected value $(0, 0)$. Only Marbles$^+$, an improvement of another game we saw there, is a unidimensional game with expected value $(0, 1)$. Unidimensional Expectations requires that Mira be indifferent between these games and their expected value, and thus that she prefer Only Marbles$^+$ to Only Matchstick. But, plausibly, in every state, Mira does not prefer the prize Only Marbles$^+$ yields in that state, to the prize that Only Matchsticks yields in that state. She plausibly does not prefer the prize of Only Marbles$^+$ in Heads to the alternative, since the difference between the two prizes is the same $((4, \text{-}2))$ as the difference between one

[10] In our setting Independence implies Stochastic Dominance (see around n. 19), and Bader [2018] observes that Stochastic Dominance suffices for a requirement to take L^+.

of the Hard Game's prizes, and what she has. She also plausibly does not prefer the game's prize in Tails, since its difference is worse ($(-4, 4)$) than the difference between the other of the Hard Game's prizes and what she has. So RECOGNITION implies that Mira is permitted to take either of these two games. This verdict conflicts with Unidimensional Expectations, since if she prefers Only Marbles$^+$, she would not be permitted to take Only Matchsticks in a pairwise choice. As I said, this seems a strong argument against RECOGNITION: the principle rules out a standard treatment of even unidimensional games, where incompleteness is not in play. But even if the argument doesn't move you, it at least provides a new, surprising constraint on what sorts of theories could vindicate RECOGNITION.[11]

RECOGNITION entails Negative Dominance (Action) and so is in an im-

[11]The best-worked out theory which vindicates RECOGNITION, Hare's "deferentialism" (which he presents but ultimately rejects) predicts even starker violations of Unidimensional Expectations than the one in the main text. Deferentialism implies that Mira is permitted to choose her present holdings, over a coin-flip in which Heads yields a hundred matchsticks, while Tails costs one matchstick, and in which neither outcome changes anything about her collection of marbles. More generally, it says that for any three prizes p_1, p_2, p_3, all "unidimensional" with each other, and such that p_1 is preferred to p_2 which is in turn preferred to p_3 Mira is permitted to choose *any* game of chance yielding prizes p_1 and p_3 over a certain gain of p_2.

To see this requires a bit of setup. 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 more sophisticated, with dependency hypotheses, but the difference won't matter here.) We assume in the background a probability p defined on S. A utility function u *represents a coherent completion of* a transitive asymmetric relation $\succ$ on O^S 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. (This ensures 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.

For simplicity let's suppose the space of outcomes is $O = \mathbb{R}^2$ and $\succ$ is defined so that $(x, y) \succ (x', y')$ iff $x > x'$ and $y > y'$ (with indifference only as required by reflexivity). All linear combinations of x and y (i.e. $u((x, y)) = ax + by + c$, with $a, b > 0$) represent coherent completions of this $\succeq$ relation. The regimentation R consisting of functions u such that $u((1, 1)) = 1$ and $u((0, 0)) = 0$ includes 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 that Mira faces a choice between L_1 and L_2 where $L_1(s_1) = (0, 100)$, $L_1(s_2) = (0,\text{-}1)$ and $L_2(s_1) = L_2(s_2) = (0, 0)$. 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, as claimed above, Mira is permitted to choose L_2. This argument generalizes straightforwardly to the class of lotteries described in the third sentence of this note.

portant sense stronger than Negative Dominance—or at least the thought behind that principle. But, one might reasonably wonder: is there any motivation for accepting Negative Dominance that isn't a motivation for accepting RECOGNITION?

There is. A first such motivation is an argument I gave at the start. There I argued 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 one of the prizes of the second. This says nothing about which *states* the relevant prizes occur in. So the argument motivates Negative Dominance, without motivating RECOGNITION.

A second motivation for Negative Dominance was implicit in my description of why Mira's behavior is "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 outcomes (a) (what she has), (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. 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 *prizes* P to itself (so that choosing any of the prizes is permitted, when all are on the table), it also maps any pair consisting of (i) one of the prizes in P and (ii) a game of chance with prizes drawn from P, to itself (so that both the prize and the lottery are permitted, when only these two are on the table). The plausible idea that rational preferences should determine a stochastically contractible choice function implies Negative Dominance. But it does not imply RECOGNITION.[12]

So, there are good reasons to accept Negative Dominance, even for those (like me) who hold that it is rationally required to take L^+ in Hare's case, and who, accordingly, reject RECOGNITION.

To sum up this section: those who accept Hare's arguments for the requirement to choose L^+ face a new choice-point here. Their endorsement of those arguments does not settle whether they should accept Negative Dominance or Independence, and Mira's puzzle shows that they must choose. Those who accept one of Hare's second two arguments (for the permission to choose L), by contrast, endorse RECOGNITION, which effectively implies Negative Dominance, so they already have a take on Mira's puzzle. But the setting of Mira's puzzle reveals a new and serious problem for their view, since RECOGNITION rules out Unidimensional Expectations just on its own. To me, this problem contributes to a case that erstwhile fans of RECOGNITION would do better to reject that principle and endorse Negative Dominance instead. But whether or not one accepts this conclusion, the problem displays a striking new obstacle to developing a systematic decision theory which vindicates RECOGNITION.

[12] Bader [2023, §1.2.2] identifies and provides important independent arguments for a related property, RATIFIABILITY UNDER EQUIPROBABLE PICKING, which would also imply a permission to stick with what Mira has in this case.

6

I now turn to responses to our problem beginning, in this section, with the possibility of upholding Independence, and thus endorsing the claim that Mira is rationally required to prefer the Hard Game.

Independence does not follow from Unidimensional Expectations, or the motivation for that principle. 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 her stock of matchsticks fixed); and, second, that Mira is 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 discuss in the next section, Unidimensional Expectations is in fact consistent with the failure of Independence in such cases. So Independence requires support that goes beyond the support for Unidimensional Expectations.

Arguments for Independence are, in fact, not far to seek. Just on its own, the principle has great intuitive appeal. Moreover, upholding the principle allows us to avail ourselves of a well-developed framework for handling decisions under risk, using sets of utility functions. It can be shown that, if a person's (incomplete) preferences satisfy Independence together with other standard axioms then their preferences can be represented by a set of utility functions: they prefer one option to another if and only if, according to every utility function in this set, that option has greater expected utility.[13] This representation provides a strong, tractable framework for reasoning about choices under risk. Since anyone whose preferences can be represented in this way must satisfy Independence, if they also satisfy Unidimensional Expectations, they must violate Negative Dominance. Indeed, Unidimensional Expectations, together with the assumption that Mira's preferences can be represented by a set of utility functions, implies Expectationalism (Lederman [2023, Proposition 3.2]).

The fact that Independence makes this strong theory available is a reason, beyond its intuitive plausibility, for accepting it. But I'm not quite happy to leave it at that. I'd be happier if I had a positive explanation of *why* Mira is not permitted to stick with what she has, in a choice between it and to the Hard Game, ideally in a way that undermines the great intuitive appeal of this judgment, and the appeal of Negative Dominance.

Let me explain. Earlier, I formulated Negative Dominance 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 prefers games of chance only because of the preferability (or not) of their prizes, then if they prefer one game to another, they prefer one of the prizes of the first to one of the prizes of the second.

[13] For results of this kind, see Seidenfeld et al. [1995], Baucells and Shapley [1998], Dubra et al. [2004], Nau [2006], 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].

The antecedent of this expanded principle is intended to rule out people who prefer not to make a decision, or love coin flips and so have a preference for a game of chance "as such".[14] It is also intended to rule out those who value other global features of games, for instance, who prefer games which have relevantly symmetric outcomes, preferring a coin-flip over $(4,\text{-}2)$ and $(\text{-}2,4)$ to a coin-flip over $(5,-1)$ and $(0,4)$, just because the former is symmetric. Beyond such clear cases, this antecedent remains vague, but there are enough clear cases for us to know, more or less, when it applies.

Given this official version of Negative Dominance, fans of Independence shouldn't reject the principle itself. (So, their official stance will differ a bit from the way I put it above.) Instead they should reject the claim that Mira is permitted to satisfy its antecedent, given the way the rest of her preferences are. In particular, they should say that if a person has preferences like Mira's, *and* satisfies Unidimensional Expectations, then it's rationally required for them to prefer games of chance for reasons other than the preferability (or not) of their prizes.

Schoenfield [2014] criticizes those who reject 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 goes on to accuse 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) is entailed by Stochastic Dominance alone (which we'll come to around n. 19), believing in such a requirement does not mean that one fetishizes expected value. I'm convinced by Bader's arguments. But I think that Schoenfield's diagnosis was prescient, since it applies to Mira's problem, even if not to Hare's. Those who uphold Independence in response to Mira's case must hold that it is rationally required that: if a person has incomplete preferences like Mira's, they must prefer games of chance for reasons other than the preferability or not of their prizes. Or, to put it another way, they hold that such a person must value global features of these games. This seems to me a very surprising result, a fetish of a different but still striking kind. It's this general requirement that Independence imposes—this fetish for global features of games—that I hope to see not just argued for, but, well, explained.

If you call something a "fetish" in philosophy, it's pretty much always a bad thing. But I don't think every fetish is bad, and there's at least one kind of view that makes this particular fetish more palatable to me. Suppose that the relation of betterness (for a person) is *complete* (for any two options, if each is at least as good as itself, then one is at least as good as the other), and that apparent examples of incompleteness in objective betterness are in fact due to incompleteness in *determinate* betterness. For instance, in Mira's case, suppose that it is not determinately the case that

[14] For models of such agents in the recent philosophical literature, see Stefánsson and Bradley [2015], Bradley [2016, 2017], Stefánsson and Bradley [2019], Goldschmidt and Nissan-Rozen [2020], Cohen et al. [2022]. There is also some evidence of fairly widespread ethical preferences in certain cases *against* randomizing, though the examples aren't exactly analogous to ours: see Meyer et al. [2019].

(4, -2) is better for Mira than what she has, and also not determinately the case that (-2, 4) is better for Mira than what she has. But suppose, even so (and this is the crucial idea), that it *is* determinate that the Hard Game is better than what Mira has (perhaps this is most natural if we also assume that it is determinately the case that at least one of (4, -2) and (-2, 4) is better for Mira than what she has, even though it's not determinate which). If requirements on Mira's preferences (in general or in this case) are constrained by what is *determinately* better, she could be required to prefer the Hard Game to what she has (since it's determinately better than what she has), and at the same time not required to prefer either of its prizes (since neither of them is determinately better). Determinate betterness would be a 'global feature' of this game, since the game would be determinately better though none of the prizes would be. So Mira's preferences would exhibit this pattern because she fetishizes a global feature. But a fetish for determinate betterness just doesn't seem so bad.[15]

This sketch of an explanation, of course, leaves quite a lot to be filled in. Most obviously there's the question: why should the Hard Game be determinately better for Mira than what she has, even though neither of its outcomes is?[16] But even if these details can be filled in, the full vindication would still have at best a narrow scope. Many attracted to the idea that rational people's preferences can be incomplete will also be attracted to the idea that betterness (and not just determinate betterness) is incomplete. They won't be able to accept the explanation above, no matter what its final form might be.

[15] A different way of explaining a requirement to prefer games of chance on the basis of global features is salient in distributional and population ethics. Many in this area are attracted to "*ex ante* Pareto" principles (e.g. Harsanyi [1955], see now e.g. Frick [2015], McCarthy et al. [2020]). As argued in Tarsney et al. [2023], a very weak such principle, stating that one lottery is better than another if it is stochastically dominant for each person, conditional on their existence, can lead to counterexamples to Negative Dominance (Goodness) (for views which accept incompleteness, but reject Negative Dominance, see Gustafsson [2020], Thornley [2022]). But even if it is accepted, the appeal of this explanation for the betterness of the lottery (because it is better for each person, *ex ante*) does not carry over to cases like those we are considering here, where the relevant "*ex ante*" principle concerns abstract dimensions of value, not people. The idea that the *ex ante* "interests" of such dimensions should be normatively relevant does not have the same appeal.

[16] Here's my best attempt. If determinate betterness satisfies Independence, continuity in probabilities, and a standard Archimedeanness condition, it is guaranteed to be representable by a set of expected value functions (for exact statements of the conditions, see McCarthy et al. [2021]; the claim follows from their Theorem 2.4, together with the fact that "Countable Domination" implies Archimedeanness). If determinate betterness moreover satisfies a version of Unidimensional Expectations, together with the claim that one prize is determinately better than another if and only if it offers more marbles and more matchsticks, then the relevant value functions in our case are all and only functions of the form $v((x, y)) = ax + by + c$ for positive a, b, c (for an exact statement, see Lederman [2023, Proposition 3.2]). The expected value of the Hard Game is greater than that of $(0, 0)$ on all of these functions, so the Hard Game would be determinately better than what Mira has—even though none of its prizes would be. This line of thought is an advance over simply assuming that the Hard Game is determinately better. But key assumptions still require defense, most obviously the Archimedeanness condition, which is violated in a wide array of simple, natural models.

Proponents of Independence must accept the (to me) surprising verdict that Mira is required to prefer the Hard Game, and so, that she's not permitted to prefer games of chance solely because of the preferability or not of their prizes. Perhaps the motivations for Independence are enough to make this conclusion acceptable. But I'd be happier if we had some kind of explanation for why she would be required to fetishize global features of the game in this way. Fans of Independence can go some way toward explaining this conclusion, if they hold that betterness is complete, and rational incompleteness in preferences is connected to incompleteness in determinate betterness. But this explanation is limited, since key motivations for incompleteness of preferences also motivate incompleteness of betterness. All of this should make us wonder if those who believe in rational incomplete preferences can do better, either by offering a more general explanation for why people are required to fetishize global features of games, or of finding a fetish-free version of the view. In the next section, I'll explore the second path.[17]

7

What if we uphold Negative Dominance and the claim that Mira's state of mind is rationally permitted, while rejecting Independence? As I said earlier, beyond the intuitive plausibility of Independence, an important argument in favor of it is that it is part of a strong simple theory: that rational people's preferences can be represented by a set of utility functions.

Can proponents of Negative Dominance provide a comparably strong theory of choice under uncertainty? I have some good news on this front. The following principle implies Unidimensional Expectations, but is significantly stronger:

Good Expectations It's rationally required that: if for every pair of possible prizes in a game of chance, a person weakly prefers one to the

[17] A version of the view which offers a clearer explanation for a preference for the Hard Game might be inspired by Levi [1986, Ch. 5]. Levi develops 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. 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. If Mira's value-functions are (say) 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 $(0, 0)$ (in particular, there is no violation of stochastic contractibility). 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? (The case is one in which Levi's choice rule violates both conditions β and γ from Sen [1969]; for discussion see Levi [1986, Ch. 6.7-8] and now Bader [2023, Ch. 1.2.1].) Needless to say, I'll follow a different route below.

other, then the person is indifferent between the game of chance and its expected value.[18]

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 she prefers (and so weakly prefers) one of its prizes to the other ($(4, 4)$ to $(\text{-}2, \text{-}2)$), and there are just these two. Crucially, however, Good Expectations still does not entail that Mira should be indifferent between the Hard Game and *its* expected value, because neither of the prizes in the Hard Game is weakly preferred to the other. This seems to me exactly the result a fan of Negative Dominance should want: the Easy Game is an obvious choice; it's the Hard Game that's hard.

My good news concerns not just Good Expectations, but a further strong axiom, which has been on the edges of our discussion for quite some time now. A game of chance A *stochastically dominates* a game of chance B for a person if and only if for each prize o, A offers at least as great a probability of prizes the person weakly prefers to o, and there are some

[18] This principle isn't as domain-general as the idea that preferences should be representable by a set of utility functions, since its consequent requires that the notion of an "expectation" be well-defined, and so only works in a setting like that of marbles and matchsticks, where there are objective quantities. But a more general principle is available:

Comparable Representation It's rationally required that, for every set of prizes (i.e. outcomes) that are totally ordered by a person's weak preferences, there is a single utility function u such that, for every pair of lotteries L, L', the utility assigned to the lottery L is its expected utility and the agent weakly prefers L to L' if and only L has at least as great utility as L' according to u.

In the setting and terminology of Lederman [2023], given Pareto, Converse Pareto and Stochastic Dominance (see below), a person's preferences satisfy Unidimensional Expectations and Comparable Representation if and only if they satisfy Good Expectations. The core idea of the proof is to use Stochastic Dominance and Unidimensional Expectations to show that, for any lottery supported on two outcomes (x, y), (x', y') that are comparable in the person's weak preferences, the lottery must be weakly preferred to the outcome (x^*, y^-), where x^* is the first-coordinate expectation of the lottery and y^- is the minimum value of y, y'. Similarly, we can show that the point (x^+, y^*) is weakly preferred to the lottery, where x^+ is the maximum of x, x', and y^* is the second-coordinate expectation. Given Pareto and Converse Pareto, the only outcome satisfying these two constraints is (x^*, y^*). Every point on the line between $(x, y), (x', y')$, including (x^*, y^*) must be comparable to every other, so Comparable Representation implies that the lottery must be exactly as good as some point on this line. But the only candidate is (x^*, y^*). (This argument can be extended from two-outcome lotteries to n-outcome lotteries, again using Comparable Representation.)

prizes o' for which A offers a greater probability of prizes the person prefers to o'.[19] It's extremely plausible that there's rational requirement to prefer one game to another if the first stochastically dominates the second (Bader [2018], Tarsney [2020]):

Stochastic Dominance It's rationally required that if one game of chance stochastically dominates another for a person, then the person prefers the first to the second.

The promised good news is that Good Expectations and Stochastic Dominance are consistent with Negative Dominance, and the basic shape of Mira's preferences (Lederman [2023, Proposition 3.1]).

In my view, this is an important result, for at least three reasons. First (least important), it shows formally (as advertised at the beginning of the previous section) that Unidimensional Expectations is independent of Independence, since Unidimensional Expectations holds in this theory but Independence does not. Second (more important), the result provides further support for a claim I made earlier, that Negative Dominance is interestingly weaker than Hare's RECOGNITION. As Bader [2018] emphasizes, Hare's L^+ stochastically dominates L for the decision-maker in that example. So, Stochastic Dominance implies a rational requirement to take L^+ in Hare's case, and thus is incompatible with RECOGNITION. The above result shows that Stochastic Dominance is, however, compatible with Negative Dominance even with a fairly strong theory in the background. Third (most important), the result delivers our good news: it shows that Negative Dominance is compatible with a strong, general theory for making decisions under uncertainty, and so puts proponents of Negative Dominance on something like an even footing with proponents of Independence.

This is, as I've said, good news for those who want to uphold Negative Dominance and the permissibility of Mira's preferences, while rejecting Independence. But I myself am not yet convinced that this is the way to go. The reason is parallel to the dissatisfaction I expressed in the previous section. There I said I hoped not just for an argument against Negative Dominance, but for a clearer explanation of why it would fail. Similarly here, I would like not just an argument against Independence, but an explanation for why *it* fails. Any plausible theory that reconciles Negative Dominance and Unidimensional Expectations by rejecting Independence will stop the argument of §4 in one of two places: by rejecting the claim

[19] Russell [forthcomingb] argues convincingly that this standard definition of stochastic dominance goes awry when incompleteness is in play. The details won't matter in the main text here, but a more exact characterization, that is apt for our purposes, would be as follows. A *lottery* is a function from some set of outcomes O to probabilities. A *generalized* lottery is a set $X \subset O \times [0,1]$ such that, $\sum_{x \in X} \pi_2(x) = 1$ (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 generalized lottery L^* is equivalent to a lottery L iff for every outcome o, $\sum_{x \in L^*, \pi_1(x)=o} \pi_2(x) = L(o)$. 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 $\pi_1(x) \succ \pi_1(f(x))$.

that Mira should be indifferent between the Die Roll and her current holdings; or (more plausibly) by rejecting the claim that she should be indifferent between the Improved Die Roll and the Hard Game. Neither of these strikes me as particularly appealing. Before I accept this theory, then, I'd like an explanation for why we might *expect* Independence to fail in this way.

Things would be much easier if the proponent of Negative Dominance could take advantage of the venerable tradition of rejecting Independence on the basis of the Allais paradox (Allais [1953]; in the case of incompleteness, Maccheroni [2004], Karni [2020], Karni and Zhou [2021]). But unfortunately I don't see how they can. Those who think that risk aversion is rationally permitted typically do not think that risk aversion is rationally required. They say that risk-neutrality is rationally permitted, and thus, that satisfying Independence is, as well. But a fan of Negative Dominance needs a story about why incompleteness in our setting *requires* failures of Independence even for intuitively risk-neutral agents, who satisfy Unidimensional Expectations. So it's not clear how the precedent set by theories of risk aversion can help.[20]

So accepting Negative Dominance is not obviously on a better footing than accepting Independence. In each case, we have strong theories which are consistent with Mira's preferences. But in each case we also have surprising results—a requirement to prefer games on the basis of their global features, or striking violations of Independence—which leave me uncertain which (if any) to prefer.

8

The core of this paper is a new puzzle for people with preferences which are sensitive to different dimensions of prizes in such a way as to be incomplete. The puzzle suggests that those who think such preferences can be rational must either reject a new dominance-like principle, Negative Dominance, or reject Independence. I've focused on a stylized example—

[20]Could the formal tools developed by proponents of risk-aversion at least help us replace Independence with a restricted but still powerful principle? The theory which is most prominent among 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 can be incomplete. Considered on its own an appropriate version Buchak's Comonotonic Sure-Thing Principle (p. 107) would not license either of the key steps for which Independence was used in the argument of §4. But it's unclear whether this principle can be embedded in a reasonably strong theory which is compatible with incompleteness.

The theory of Chew [1979, 1983] (see now Bottomley and Williamson [forthcoming]) does not require a complete ranking of outcomes (Karni and Zhou [2021] provide a characterization). The principle of Weak Substitution (which Bottomley and Williamson call "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 know whether Weak Substitution can be consistently added to the package described above. But even if it can be, this would not yet give us the explanation I hope for here, of why Independence fails in the way it would still have to.

of Mira's preferences for marbles and matchsticks—but the problem for incomplete preferences is much more general. Indeed, it extends beyond rational preferences to views in axiology according to which betterness for a person, or overall, can be incomplete, 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 (the relevant version 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.

But some may see the puzzle in a quite different light. I started by saying that, at least in some cases, it seems that people don't have preferences between two options, and that's not because they're ignorant of what they prefer. But one might respond to the puzzle by seeing it as evidence against the claim that rational preferences can be incomplete owing to their sensitivity to different dimensions of prizes. Since arguably one of the best motivations for the rationality of incomplete preferences is the idea that preferences can be sensitive to different dimensions in this way, one might see it as an argument against incompleteness full stop.[21]

I certainly think that, since the puzzle arises for incomplete preferences and not for complete ones, it should make us *somewhat* more confident that it is rationally required that preferences be complete. But I don't see the puzzle as particularly strong evidence for that conclusion, and I'll close, in the rest of this section, by saying why.

The main reason is that I don't see the puzzle as strong evidence that betterness (for a person, or overall) is complete, and I think that if betterness is incomplete, there's no rational requirement that preferences be complete.[22]

[21] Mira's puzzle might be added to other considerations—e.g. Dorr et al. [forthcoming], Dorr et al. [2021]'s linguistic arguments, or Gustafsson [2022]'s moneypump—which tell against the rationality or even possibility of incomplete preferences (though see Worsnip [2021, §9.2], with which I have some sympathy, 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.

[22] 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 Bradley [forthcoming] suggests that 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. The putative requirement to have complete preferences is a putative "wide-scope" requirement. But there is a difference between it and more paradigmatic such requirements, for instance, the requirement that if one believe that p, and believe that if p then q, then one believe q. Those who violate such paradigmatic requirements (and recognize that they do) can remedy their situation by giving up one or more of the attitudes described in their antecedents. But unlike paradigmatic wide-scope requirements, there is no way to complete one's preferences (or for that matter comparative confidence rankings) by giving up an attitude.

To motivate the second claim, 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 nothing is at stake except what's better (for her, or overall). For instance, perhaps she knows that gaining four marbles at the expense of two matchsticks is not better or worse or exactly as good for her as what she has now, and there's nothing else at stake. If it is a rational requirement to have complete preferences, then apparently Mira must, even so, have preferences between these options.

This seems to me odd. If Mira does not have a preference between the options, it is unclear what reasons we could give to convince her to form one. If we recommend that she come to have such a preference, she has a compelling reason to reject our advice, since she knows that neither of the options is better than the other. If this is good enough for betterness, she might say, it is good enough for her.

If betterness is incomplete, then, it seems unlikely to me that there would be a further rational requirement that preferences be complete. And so, Mira's problem should make us significantly more confident that preferences are required to be complete only if it should make us even more confident that betterness is complete.

This conclusion 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. It is a requirement of "coherence" on the pattern of attitudes we have, not a requirement on how our attitudes respond to the world or our evidence about it. It can often seem that structural requirements of this kind have a less intimate relationship to the way the world is, than so-called "substantive" requirements. But the present example highlights that this does not mean that they do not have any such relationship at all. As we have seen, it is plausibly a rational requirement that preferences be complete only if betterness is also complete. The (structural) rational requirement would thus make a certain demand on the world. Many structural requirements make such demands. Plausibly, it is a rational requirement that one not believe p while also believing $\neg p$ only if, in fact, it is never the case that p and $\neg p$. But the case of completeness brings out the connection to constraints on the world especially clearly perhaps because, while it is obvious that contradictions can't be true, it is not obvious that betterness is complete.

Should Mira's problem make us significantly more confident that betterness is complete? I think it should make us at least a little more confident, but I am not sure how much. The reason—in an even more speculative key—is as follows.

In my view, there are two relevant relations expressed by the word "better". A first, more basic relation, holds among prizes or outcomes—whether those are states of the world, situations, lives, or something else. A second, less basic relation, holds between games of chance. One game of chance may, in a sense, be better for a person than another, but this is not the same sense in which one life is better for that person than another. It is the friendship, happiness, achievement, or love in a life, that are (in the basic sense) better for a person, not the expected friendship, happiness, achievement or love, which a game of chance provides. In my view, betterness in the first sense—the relation between lives or states of the world—should be simple and logically well-behaved. But it is not at all clear that betterness in the second sense will.

An example may help to motivate this idea. The relation of "worth at least as much as" that holds between bundles of standard one-dollar bills is easy to characterize: it just depends on how many are in each bundle. By contrast, the relation of "worth at least as much as" that holds between different dollar-denominated gambles in a casino is hard to characterize, and, in fact, many doubt that it exists. Some say that what different gambles are worth varies from person to person, depending on (for instance) their risk attitudes, so that it doesn't make sense to talk about which is worth more *simpliciter*. But even those who think we can speak of the comparative worth of gambles often hold that the relation on gambles behaves quite differently than the relation over bundles of one-dollar bills. For instance, the relation among gambles might be incomplete, though the relation on bundles is complete. Whether or not such positions are correct, the relation on bundles of dollar bills is clearly distinguishable from the relation on gambles, and, plausibly, these relations are in fact distinct, as well.

Our example serves to make a second important point, as well. In the example, considerations from the theory of gambles do not provide evidence for or against a particular theory of the worth of the bundles. The fact that it is hard to settle the comparative value of gambles does not cast any doubt on our original theory of the comparative value of the bundles themselves.

In much the same way, it may be that complexity in the relation of betterness among games of chance does not provide evidence, one way or the other, about the theory of betterness defined on outcomes. If there is incompleteness in the relation over outcomes due to different dimensions of those outcomes, then one of Negative Dominance and Independence plausibly must fail, so the relation of betterness on games of chance must be, to some extent, poorly behaved. But this fact about the relation of betterness on games of chance may not provide strong evidence against incompleteness in the betterness relation on outcomes. The proponent of incompleteness can naturally hold that the betterness relation over outcomes does not "care" whether life is easy or hard for those who face games of chance over outcomes; betterness is, at its most basic level, a relation only on the outcomes themselves.

In short: the puzzle is certainly *some* evidence that betterness is complete, since if it is, we could maintain all three of the plausible-seeming principles. But I don't see the puzzle as strong evidence for this claim, because there is a natural position on which the ill-behavedness of the betterness relation on games of chance is simply irrelevant to questions about structural features of the (different) betterness relation on outcomes.

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

References

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

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

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

Manel Baucells and Lloyd S. Shapley. Multiperson utility. Tech Rep. 779 UCLA Department of Economics, 1998.

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.

Richard Bradley. Ellsberg's paradox and the value of chances. *Economics & Philosophy*, 32(2):231–248, 2016.

Richard Bradley. *Decision theory with a human face.* Cambridge University Press, 2017.

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. 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. Incommensurability, incomparability, and practical reason. In Ruth Chang, editor, *Incommensurability, Incomparability, and Practical Reason.* Harvard University Press, 1997.

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

SM Chew. *Alpha-nu choice theory: A generalization of expected utility theory.* British Columbia, 1979.

Soo Hong Chew. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica*, pages 1065–1092, 1983.

Haim Cohen, Anat Maril, Sun Bleicher, and Ittay Nissan-Rozen. Attitudes toward risk are complicated: Experimental evidence for the re-individuation approach to risk-attitudes. *Philosophical Studies*, 179(8):2553–2577, 2022.

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.

Cian Dorr, Jacob M. Nebel, and Jake Zuehl. The case for comparability. *Noûs*, forthcoming.

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

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

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

Kit Fine. Parity under risk. Manuscript, 2024.

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

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.

Zeev Goldschmidt and Ittay Nissan-Rozen. The intrinsic value of risky prospects. *Synthese*, 198(8):7553–7575, 2020.

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

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

Johan E. Gustafsson. *Money-Pump Arguments*. 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.

John C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321, 1955.

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

Edi Karni. Probabilistic sophistication without completeness. *Journal of Mathematical Economics*, 89:8–13, 2020.

Edi Karni and Nan Zhou. Weighted utility theory with incomplete preferences. *Mathematical Social Sciences*, 113:116–135, 2021.

Harvey Lederman. Incompleteness, independence, and negative dominance. MS, 2023. URL https://arxiv.org/abs/2403.17641.

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

Fabio Maccheroni. Yaari's dual theory without the completeness axiom. *Economic Theory*, 23:701–714, 2004.

Paola Manzini and Marco Mariotti. On the representation of incomplete preferences over risky alternatives. *Theory and Decision*, 65:303–323, 2008.

David McCarthy, Kalle Mikkola, and Joaquin Teruji Thomas. Utilitarianism with and without expected utility. *Journal of Mathematical Economics*, 87:77–113, 2020.

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

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

Joseph Raz. Value incommensurability: some preliminaries. In *Proceedings of the Aristotelian Society*, volume 86, pages 117–134. JSTOR, 1985.

Joseph Raz. *The Morality of Freedom*. Clarendon Press, 1986.

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 Perspectives*, 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.

Amartya Sen. Quasi-transitivity, rational choice and collective decisions. *The Review of Economic Studies*, 36(3):381–393, 1969.

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

H Orri Stefánsson and Richard Bradley. How valuable are chances? *Philosophy of Science*, 82(4):602–625, 2015.

H. Orri Stefánsson and Richard Bradley. What is risk aversion? *The British Journal for the Philosophy of Science*, 70(1):77–102, 2019.

Christian Tarsney. Exceeding expectations: Stochastic dominance as a general decision theory. Global Priorities Working Paper, 2020.

Christian Tarsney, Harvey Lederman, and Dean Spears. Share the sugar. MS, 2023. URL https://arxiv.org/abs/2403.17641.

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

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.