Dice Ex Machina

Ryan Doody
Brown University

Abstract

A familiar way of deciding between indifferent options is to use a randomizing device (like the toss of a coin, or the roll of a die) to aid in the selection. The central aim of this paper is to demonstrate that such behavior is actually deeply puzzling. I argue that tossing a coin is in tension with a number of frameworks for rational choice. I then develop a solution to the puzzle, which posits that tossing a coin allows us to express our indifference between the options, and that we sometimes have good reasons—reasons regarding self- and other- understanding—to want to express such things.

1 Introduction

Which color should we paint the wall? You study the two remaining color swatches, and can't decide. What should we order for dinner tonight? You want either the Pad Thai or the Drunken Noodles, but you're struggling to choose. Where should we vacation this summer: Baghdad or Samarra? It's not clear which place to go.

Temporarily paralyzed by indecision, you reach into your front pocket and—to your great relief—you find a fair coin. Rather than select one of the two options outright, there's now something else you can do: you can let the coin decide. This is welcome news. Your indecision has been resolved: you elect to flip the coin.

I trust this phenomenon—of resolving indecision by coin-toss, dice-roll, or other external, randomizing aid—is a familiar one. It's not uncommon, at least concerning fairly trivial matters, to use a randomizing device in decision-making.[1] Without loss of generality, let's call the phenomenon a *dice ex machina*.

This paper has two goals. The first is to demonstrate that, despite their popularity, *dice ex machina*s—resolving indecision with, e.g., the toss of a coin—are actually deeply puzzling. The second is to offer an account of the phenomenon that answers the puzzle.

[1] For experimental results attesting to a "preference for randomness" in certain cases, see Agranov and Ortoleva (2017, p. 54), Dwenger et al. (ms). Keren and Teigen (2010) found, however, that experimental subjects are strongly opposed to resolving indecision by coin-toss when matters of great importance are at stake (especially, in cases of life-and-death).

In order to bring out what's puzzling about deciding by coin-toss, it'll be helpful to consider different accounts of the sorts of quandaries that open the paper. I'll start with the view most familiar to economists: expected utility theory. I'll then consider other views, which weaken some of the assumptions of expected utility theory. I'll argue that, in each case, there's something puzzling about using a randomizing device to decide.

2 Orthodoxy: Expected Utility Theory

According to expected utility theory, you rationally ought to take the option, out of those available to you, that maximizes expected utility. The expected utility of an option is the weighted average of its outcomes' values, where the weights correspond to the probabilities of the outcomes. On this view, your motivations are represented with a utility-function, which assigns numbers to every potential outcome—the higher the number, the more strongly you prefer it. Your uncertainty, if you have any, is represented with a subjective probability-function, or credence. Together, these two combine to rank your options in terms of their *expected utilities*. You ought to take the option with the highest expected utility. If there are multiple options tied for "highest", you ought to take one of these, but which one in particular is left up to you—it's permissible to take any.

In order, then, to apply expected utility theory in any particular case, we need to know something about your *preferences* and your *credences*. What about the cases that open the paper?

Those examples all have the following structure. You have two options: A and B. You neither prefer A to B, nor do you prefer B to A. According to expected utility theory, then, you are *indifferent* between A and B—in other words, the expected utility you assign to A is equal to the expected utility you assign to B.[2] Flipping a fair coin to decide between your options is a gamble that, e.g., results in A if the coin lands heads and results in B if the coin lands tails. Because the coin is (correctly believed by you) to be fair, the coin-toss option corresponds to a 50/50 lottery between option A and option B.

The examples that open the paper suggest something further as well—namely, that there is something special about the coin-toss: it's the option, out of those available, that you most prefer. Here are two pieces of evidence. The first is *phenomenological*: the discovery of the coin in your pocket comes as welcome news.

[2] Expected utility theory offers a representation of your preferences. The expected utility you assign to X, $eu(X)$, is higher than the expected utility you assign to Y, $eu(Y)$, if and only if you prefer X to Y. In these examples, because you don't prefer A to B, $eu(A) \not> eu(B)$; because you don't prefer B to A, $eu(B) \not> eu(A)$; and, because expected utility theory involves assigning numbers to all of your options, it must be $eu(A) = eu(B)$.

This suggests that you would rather decide from the menu that includes the coin-toss option than from the menu that does not—you prefer the former to the latter. And *that* suggests that you prefer the coin-toss to A and B.The second is *behavioral*: the introduction of the coin-toss option resolves your indecision in its favor—once you have the option to toss the coin, you avail yourself of it. In fact, let's stipulate—plausibly enough—that, in the kinds of cases under consideration, when flipping a coin becomes an option, you are disposed to take it. It's not the case, for example, that were you to face the same decision problem again and again, about one-third of the time you would take A, one-third of the time you would take B, and the remaining third of the time you would toss the coin—instead, you would always toss the coin. This, too, suggests that you prefer the coin-toss to A and B.[3]

Conjecture 1: You prefer the coin-toss to A and you prefer the coin-toss to B.

Suppose I told you that there is an option, Z, that I am happy to have on the menu and that, when it is on the menu, I am strongly disposed to select. It would be reasonable of you to conclude that I prefer option Z to the other options on the menu. It appears as though the coin-toss is just such an option.

But it's hard to reconcile **Conjecture 1** with expected utility theory. Here's why. According to expected utility theory, you should prefer one option to another only if the former has greater expected utility than the latter. But it's hard to see how the coin-toss could have *greater* expected utility than the options it's being used to decide between. As mentioned, flipping the coin to decide between A and B corresponds to a 50/50 lottery between option A and option B—that is, it pays out option A with probability 0.5 and pays out option B with probability 0.5. The expected utility of the coin-toss is, then, the *average* of the expected utilities of A and B:

$$\begin{aligned} eu\left(Toss\right) &= \tfrac{1}{2} \cdot eu\left(A\right) + \tfrac{1}{2} \cdot eu\left(B\right) \\ &= \frac{eu\left(A\right) + eu\left(B\right)}{2} \end{aligned}$$

Whatever those expected utilities happen to be, it's clear that $eu\left(Toss\right)$ cannot be *greater* than both $eu\left(A\right)$ and $eu\left(B\right)$—for the simple reason that the *average* of two numbers can never be greater than *both* of the numbers it's the average

[3] On some views—some interpretations of *revealed preference theory* (Little, 1950; Samuelson, 1938), for example—*preference* is analyzed in terms of *choice*: roughly, you prefer one option to another just in case you are disposed, when given the choice between the two, to choose the former over the latter. There are compelling reasons to resist such an analysis (Hausman, 2012), which is I why I say 'suggests' rather than 'means' or 'shows'.

of.[4] Furthermore, if $eu(A) = eu(B)$ (as I suggested, according to expected utility theory, if you don't prefer the one to the other, it must), then $eu(Toss) = eu(A) = eu(B)$. Despite initial appearances, it seems you can't prefer the coin-toss to both of the other options—you should be indifferent between them all. Expected utility theory and **Conjecture 1** cannot both be true.

2.1 Response 1: Redescription

But not so fast! Expected utility theory and **Conjecture 1** can be reconciled by redescrbing the outcomes of your options—individuating them more finely—so that the former can accommodate the attitudes ascribed by the latter. This strategy—let's call it the *Re-individuation Strategy*—has become a popular way for proponents of expected utility theory to diffuse putative counterexamples to the view's axioms.[5] One such counterexample—first advanced by Diamond (1967)—is particularly germane: it provides an example of a gamble that appears to be more valuable than any of its outcomes. Perhaps, then, the classic way of responding to this counterexample—by making use of the *Re-individuation Strategy* (see, for example, Broome, 1984b; Dreier, 1996)—can be adapted to our purposes?

First, let's look at a version of Diamond's example. Suppose there is an indivisible good (e.g., a puppy) that you'd like to distribute, and two people (e.g., Agnes and Boris) who are equally deserving of it. Ideally, you'd be able to equally divide the good between the two—but that's not possible. What is possible, though, is to toss a fair coin, and give the good to Agnes if it lands heads and to Boris if it lands tails. You have three options: give the good to Agnes (A), give the good to Boris (B), or use the coin to decide in the manner just described (*Toss*). See Table 1 for a representation of your Solomonic predicament (where a '1' represents receiving the reward, and 'o' represents not receiving it.)

What should you do? The answer seems fairly clear: you should *Toss*! The resulting *ex post* distribution will be unequal no matter what you do, so, if you care about fairness and equality, the best you can do is to distribute the good with the toss of a fair coin. Doing so, at the very least, equalizes each person's *ex ante*

[4] Preferring the coin-toss appears to violate (what is sometimes called) *Betweenness*: if X is preferred to Y, then, for any probabilistic mixture of the two, X must be preferred to it and it must be preferred to Y. (If you also accept *Continuity*, this entails that, if you're indifferent between X and Y, you must likewise be indifferent between them and any of their probabilistic mixtures.) The central axiom underlying expected utility theory—*Independence*—entails *Betweenness*, but not *vice versa*. A number of decision theories that reject *Independence* accept *Betweenness* (See, Camerer and Ho, 1994, for further discussion).

[5] The strategy is deployed to diffuse counterexamples to *transitivity* (Cf. Broome, 1991, 1993); counterexamples to *Basic Contraction Consistency*, a.k.a the α-property, a.k.a. the Independence of Irrelevant Alternatives (Cf. Neumann, 2007; Rulli and Worsnip, 2016; Sen, 1993, 1997); Allais' counterexample to the *Sure-Thing Principle* (Cf. Bermudez, 2009; Broome, 1991; Buchak, 2013; Jeffrey, 1987, 1988; Raiffa, 1968); and so forth.

Table 1: Diamond's Counterexample

	Heads		**Tails**	
	Agnes	Boris	Agnes	Boris
A	1	0	1	0
B	0	1	0	1
Toss	1	0	0	1

chance of receiving the good—it gives each person a fair shake.

But, as Diamond demonstrates, preferring *Toss* to *A* and to *B* is inconsistent with expected utility theory.[6] If the coin lands heads, the results of *A* and *Toss* are the same (Agnes gets the puppy and Boris doesn't). So, given the *Sure-Thing Principle*, *Toss* is better than *A* only if the former is better than the latter when the coin lands tails—in other words, only if it's better for Boris to get the puppy than it is for Agnes to get it. But if it's better for Boris to get the puppy than Agnes, then *B*—which guarantees that outcome—should be better than *Toss*. So, *Toss* can't be better than both.Something has to go.

Or does it? The *Re-individuation Strategy* to the rescue! It's reasonable to prefer *Toss* because distributing the good by coin-toss is *fair* in a way that directly giving it to Agnes (or Boris) is not. There are, of course, various ways of spelling out what *fairness* consists in.[7] But whatever account is given, it's not implausible to think that a world in which Agnes receives the reward as the result of a coin-toss is one in which Boris is treated more fairly than one in which Agnes receives the reward by being given it directly. But this means that the outcomes as depicted in Table 1 are misleading: by only including information about who received what and not information about the *process*, Table 1 suggests that certain outcomes are the same, when they in fact aren't. The outcome of *Toss*, if the coin lands heads, is *Agnes-gets-the-reward-as-the-result-of-a-fair-process*, whereas the outcome of *A* is *Agnes-gets-the-reward-as-a-result-of-my-decision-to-give-it-to-her*. And while these two might be the same with respect to what Agnes gets, they differ in a

[6] More carefully, Diamond (1967), while willing to accept expected utility theory (and, in particular, the *Sure-Thing Principle*) for individual choice, is hesitant to do so for social choice "since it seems reasonable for the individual to be concerned solely with final states while society is also interested in the process of choice," (p. 766).

[7] For example, it might be argued that *Toss* is fairer than *A* and *B* for, roughly, *egalitarian* reasons: the former affords Agnes and Boris *equal chances* of receiving the reward. Alternatively, it might be argued that *Toss* is fairest because it embodies a procedure that gives each person their due—it constitutes a way for the distributor to *treat* each of them fairly. Relatedly, *Toss* might be considered fairest because it ensures *impartiality* on part of the distributor, and such impartiality is called for in such a case. There are other views as well. (For greater discussion of these issues, see, Broome, 1984a,b, 1990; Hooker, 2005; Sher, 1980; Voorhoeve and Fleurbaey, 2012; Wasserman, 1996).

respect that makes it not unreasonable to have a preference between them. Table 2, then, offers a more faithful representation of the case.

Table 2: Diamond's Counterexample, Re-individuated

	Heads		**Tails**	
	Agnes	Boris	Agnes	Boris
A	1 as the result of my decision	0	1 as the result of my decision	0
B	0 as the result of my decision	1	0 as the result of my decision	1
Toss	1 as the result of a fair process	0	0 as the result of a fair process	1

If we re-individuate the outcomes in this way, it's no longer inconsistent with expected utility theory to prefer *Toss* to both *A* and *B*. It's true that a 50/50 lottery cannot have a higher expected utility than both of its outcomes. But we now see that, despite initial appearances, *Toss* is *not* a 50/50 lottery between *A* and *B*.[8]

The *Re-individuation Strategy* is not uncontroversial. A common complaint is that, because it is always possible to re-individuate outcomes in the face of purported counterexample, the strategy renders expected utility theory vacuous.[9] The complaint can be answered by providing a principled account of when it is, and is not, appropriate to re-individuate. Broome (1991, 1993), for example, thinks that it's kosher to distinguish outcomes only if they differ in a way that it would be rational to care about. Dreier (1996), on the other hand, re-individuation is kosher just so long as one actually does have a preference between the two. There are surely other views as well. Whichever view one holds, whether a particular instance of the *Re-individuation Strategy* is justified will depend on the details of the case. The details in Diamond's example justify it—the outcomes differ with respect to *fairness*, which is something reasonable people care about.

Although Diamond's example clearly shares certain similarities to our own, it's less clear what could play the role *fairness* plays in justifying the re-individuation of the outcomes. It doesn't seem possible—or, at least, it's quite a stretch—to treat a color swatch or a meal or a vacation destination *unfairly*. And, even if it were, it's not clear why this would be something worth caring about. It makes sense to prefer outcomes in which Agnes and Boris are treated fairly over outcomes in

[8] Broome (1984b) responds to Diamond's counterexample along these lines. (For further discussion, see Broome, 1991; Buchak, 2013; Dreier, 1996)

[9] See, for example, Sen (1997, p. 754), who is concerned with a similar problem regarding rationality constraints on choice-functions; Tversky (1975, p. 171), who worries that, without constraints, the strategy makes one's decision theory "empty from both descriptive and normative standpoints"; and Baccelli and Mongin (2021) for more details.

which they are not, because Agnes and Boris are the kinds of things—people—that can be, and should be, treated fairly. Acadia White, Drunken Noodles, etc. are *not* the kinds of things—except, perhaps, in a metaphorical sense—that can be, or should be, treated fairly.

That leaves open, of course, that there might be some other quality—not fairness, but something—that justifies the appropriate re-individuation of outcomes in our *dice ex machina* cases. But, of course, the real work is in saying specifically what that other quality might be, and why it might be a quality worth caring about. After surveying some other possibilities, we shall return to this thought.

2.2 Response 2: Salience

Let's set the *Re-individuation Strategy* aside. Recall that expected utility theory conflicts with **Conjecture 1** (the claim that you prefer the coin-toss to both A and B). One way to resolve the tension is to reject **Conjecture 1**, and to follow expected utility theory where it appears to lead: if you're indifferent between A and B, you should be indifferent between them and the coin-toss as well.

Conjecture 2: You're indifferent between the coin-toss and A and between the coin-toss and B.

But this is puzzling. If your indifference between A and B made it hard to decide between the two, how could introducing the coin-toss help? If anything, shouldn't the coin-toss make things even worse? Before, you were indifferent between only two options; now, you're indifferent between *three*! What, then, explains why you are nevertheless disposed to choose the coin-toss?

Here's a potential answer: although you don't *prefer* the coin-toss to the other options on the menu, the coin-toss is particularly *salient* to you. Its salience—rather than a preference—underlies your disposition to choose it. On this view, it's not the case that, if you're disposed to choose one thing over another, you prefer it.

How plausible is this? Here's an analogy (adapted from Schelling, 1960): Suppose we are to meet at noon in Paris tomorrow, but we have no way of contacting each other beforehand. I am completely indifferent about where we meet—I just want us to go to the *same* place, wherever that happens to be. You feel the same way. And our predicament is common knowledge. Where should I go? The Eiffel Tower. Why? I'm disposed to go to the Eiffel Tower rather than, say, the Louvre, not because I prefer meeting there (I don't), but because the Eiffel Tower is a particularly salient landmark. I must meet you somewhere, and the Eiffel Tower's salience acts as a focal point. Analogously, the coin-toss' salience singles it out as the thing to do—again, not because I *prefer* it to my other options, but because I have to choose something and its salience acts as a focal point.

I find this account unsatisfying for three reasons. First, the analogy doesn't hold. While it's true that I don't prefer meeting at the Eiffel Tower to meeting at the Louvre, I *do* prefer meeting you somewhere. Furthermore, because the Eiffel Tower's salience acts as a "focal point for each person's expectation of what the other expects him to expect to be expected to do" (Schelling, 1960, p. 57), I think it's more likely that I will find you at the Eiffel Tower than at the Louvre. And so I do prefer *going to the Eiffel Tower* to *going to the Louvre*—even though, conditional on successfully meeting you somewhere, I'm indifferent about where. Your disposition to go to the Eiffel Tower is based on a preference after all. As far as I can tell, there's no analogous story—involving, say, the expectations of others—that can be told about the coin-toss.

Second, it's not clear that the coin-toss option will always be the salient option. And, in those case in which it is, why? Here are a couple possibilities. In the stories that opened the paper, the coin-toss was introduced last—after your indecision between the two options had made itself known. It's *newness*—in virtue of being a property that distinguishes it from the other options—might, then, make it the most salient option. But that doesn't seem to hold in general. Suppose, for example, that after discovering the coin in your pocket, you also discover a color shade or Thai dish or vacation destination that you hadn't considered before. Even though these options are now even newer than the coin-toss, I'm skeptical that one would be disposed to choose them on those grounds. Instead, the coin-toss' salience surely has something to do with the fact that it's a "mixed" option: it probabilistically combines the options already on the table. But this raises further questions because there are many—infinitely many, in fact—probabilistic mixtures of the other options. Would a (known by you to be) biased coin work just as well?

Finally, notice that there are actually two different ways of deciding by coin-toss.You can toss the coin with the intention to opt for A if the coin lands heads and to opt for B if the coin lands tails, or you can toss the coin with the intention to opt for B if the coin lands heads and to opt for A if it lands tails. Call the former option '$Toss_{(A,B)}$' and the latter '$Toss_{(B,A)}$'. So, in fact, when you discover the coin in your pocket, you introduce *two* new options—resulting in four options you're indifferent between.

Whatever story gets told about why flipping the coin is more salient than A and B, it's hard to see how either $Toss_{(A,B)}$ or $Toss_{(B,A)}$ could be more salient than the other. (Or, if one is, why the features that make it more salient wouldn't also thereby conspire to make A or B more salient than the other.) And yet, as much as we might struggle with our indecision between A and B, we seem to have no problem whatsoever deciding which option to associate with which side of the coin.

Table 3: *Dice Ex Machina*

	Heads	**Tails**
A	A	A
B	B	B
$Toss_{(A,B)}$	A	B
$Toss_{(B,A)}$	B	A

But, if that is so, our *dice ex machina* is even more puzzling still. Here's why. Following Ullmann-Margalit and Morgenbesser (1977, p. 757), let's distinguish between *picking* and *choosing*. You *choose* when your selection of an option reflects your preference for it. You *pick* when you select an option in spite of strict indifference between it and its alternatives. In our examples, we've been assuming that you're *indifferent* between A and B, and thus must *pick*; that this indifference has hindered your ability to make a selection—because, after all, what could justify you in selecting one rather than the other?; and that you can "call upon chance to extricate [yourself] …" (Ullmann-Margalit and Morgenbesser, 1977, p. 769) with the toss of a fair coin. But there is a tension here because,

> [T]he very use of a random device is premised on the possibility of picking, that is, on our capacity to extricate ourselves from a picking situation: the matching of each of the alternatives up for selection with some one of the possible outcomes of the device is, inherently, a matter of picking, (Ullmann-Margalit and Morgenbesser, 1977, p. 770).

So, it would seem, the device needed to rescue us from a picking situation presupposes our ability to do without it. In order to use a coin-toss to decide between A and B, we must be able to pick between the two different ways of using the coin to decide between A and B (i.e., we must be able to pick between $Toss_{(A,B)}$ and $Toss_{(B,A)}$). But if we have such an ability, why not exercise it in the picking situation between A and B? Why toss the coin at all?

Working toward answering these questions requires a greater understanding of what it is to *pick* among options you're indifferent between.

2.3 Interlude: Picking and Indifference

Ullmann-Margalit and Morgenbesser (1977) argue that we have the ability to *pick*—to select among equally preferred alternatives. Unlike Buridan's ass, we won't starve between equidistant bales of hay. Their argument appeals to, what I'll call, *supermarket scenarios*: the quotidian selection of a consumer good from among

a shelf of nearly-identical alternatives (e.g., Campbell tomato soup cans, boxes of General Mills' Lucky Charms). Because we routinely *do* select between nearly-identical items, we clearly have the *ability* to. And, they argue, this selection-process doesn't require us to somehow extricate ourselves from the picking situation by, e.g., coming to prefer one of the items to the others.[10]

Notice that the cases that opened the paper—deciding between two different paint colors, deciding between different entrees, deciding between two different destinations—are very much not *supermarket scenarios*; they aren't cases of selecting between nearly-identical items. Although both kinds of cases might involve options you're indifferent between, the nature of your indifference is different. In fact, there are three categories of indifference worth distinguishing:

I: *Lucky Charms.* You are indifferent between A and B because, at least with respect to the features that matter to you, A and B are (nearly) qualitatively identical. (To the extent that the two differ, either that difference doesn't matter to you or it's so slight that you're unlikely to ever notice.)

EXAMPLE: Selecting a box of Lucky Charms from a shelf of nearly-identical ones at the supermarket.

II: *Door #1 vs Door #2.* You are indifferent between A and B—even though you're certain that one is, as a matter of fact, better than the other—because you're evidence regarding which is which supports thinking it equally likely to be either.

EXAMPLE: You know the grand prize is either behind Door #1 or Door #2, but you know not which. You have no reason to think it's more likely to be behind one as it is to be behind the other.

III: *Amazon vs Starbucks.* You are indifferent between A and B—even though they differ from each other along various dimensions—because those various differences perfectly balance each other.

EXAMPLE: You can receive either an Amazon or a Starbucks gift card.[11] If you accept the former, you'll be able to partially pay for a book you've been meaning to read. If you accept the latter, you'll spend it on several sugary lattes. There are pros and cons to each, but when you think about it, they seem equally good to you.

[10] In particular, Ullmann-Margalit and Morgenbesser (1977, p. 769-772) survey a number of ways to extricate oneself from a picking situation—including, the use of a randomizing device—and find them unsatisfying, incomplete, or unrealistic. In addition to a random device, they consider the adoption of a random tie-breaking *policy* (which is the solution recommend by Rescher, 1960), the use of *convenience* to break the tie, selecting out of *habit*, and the operation of the "will".

[11] This example comes from Dwenger et al. (ms), who offered a similar choice to participants in a study. They found that more than half opted for a lottery to select between the two.

It's not the case that all indifference is the same. Likewise, we might think, for picking situations. And so, while Ullmann-Margalit and Morgenbesser (1977) might be right that we have the ability pick, it may be easier to exercise that ability in some situations than in others. We'll come back to this suggestion in the end (§4).

What does this ability to pick consist in? Ullmann-Margalit and Morgenbesser (1977, p. 773) say,

> [W]hen we are in a genuine picking situation we are in a sense transformed into a chance device that functions at random and effects arbitrary selections (our misgivings about "arbitrary" notwithstanding; still other can be added about "random".)

They go on to clarify that 'random' can (and likely should) be understood epistemically rather than metaphysically—it's not that the selection is *undetermined* or *unexplainable*, but rather that (at least given our typical level of epistemic access) we have no reason to think it any more likely that you'll select any one of the options than any of the others.[12] And this is true of you—the decision-maker—too. If you opt to pick between A and B, you should be 50% confident that you'll select A and 50% confident that you'll select B. What need, then, is there to toss a coin if you can effectively transform yourself into one? Furthermore, because there is always *some* cost to tossing a coin—e.g., you might drop it!—it's hard to see how, if facing a picking situation, it could ever be rational to do so.

Perhaps it is true that, on the orthodox picture, it's irrational to toss a coin to decide between options you're indifferent between. But perhaps the orthodox picture is *incorrect*. And perhaps your indecisiveness isn't a product of *indifference* between your options, but something else.

3 Unorthodoxy: Impediments and Information

The orthodox picture abstracts away from our computational limitations. We have preferences, and we have credences—both are well-defined and readily accessible for the purposes of decision-making—and they combine to issue recommendations about what to do. The picture leaves little room for genuine *deliberation*, which is often costly (in terms of both time and mental energy).

A *dice ex machina* might be better rationalized as a response to some of the real-life limitations being abstracted away from on the orthodox picture. Perhaps we can find a place for them in an unorthodox picture that is sensitive to the

[12] Although far from settled, there's empirical evidence that, in some circumstances, people make random selections. (For discussion, see Agranov and Ortoleva, 2017; Glimcher, 2005; Sippel, 1997)

ways in which we are *boundedly rational*.[13] In particular, let's jettison two of the assumptions made in the previous section: first, that deliberating isn't costly; and second, that your preferences are transparent to you.

I will argue that these approaches don't fair much better than the previous ones.

3.1 Costly Deliberation

Here's an initially plausible suggestion: we toss a coin to decide when we surmise that the costs of further deliberation would be higher than the risks of selecting incorrectly. On this picture, you're *not* indifferent between A and B—you have a preference for one over the other, but you're not entirely sure which. If you knew, you would choose it. You're fairly confident that if you deliberated long and hard enough, you'd come to have a better idea about what you'd prefer to do. But deliberation is costly.[14] If the difference between getting what you prefer and missing out on it isn't too great, it might make more sense to decide by coin-toss than to continue on deliberating over the matter.

When should you continue to deliberate? And when should you toss the coin? Making some simplifying assumptions (e.g., that it's equally bad to select B when A is what's preferred as it is to select A when B is what's preferred), you should favor deliberation when the following inequality holds:

$$r > 1/2 + \frac{costs}{stakes} \tag{1}$$

Where r is how reliable you think deliberating will be—that is, it's your credence that deliberation will result in a particular verdict conditional on that verdict being correct.And where the stakes are the difference between choosing the option you prefer and choosing the option you disprefer—in other words, it measures how bad it would be to select the "wrong" option. The higher the stakes, the lower the costs, and the more confident you are that deliberation would be successful, the more likely it is that you should continue to deliberate than to toss the coin; on the other hand, the lower the stakes, the higher the costs, and the less confident you are in your deliberations, the more likely it is that you should cut your deliberations short and just toss a coin.

Although there is something compelling about the model, it's not entirely

[13] Once we move away from the orthodox picture, there are many different ways of modeling different aspects of our limitations. (See, for example, Conlisk, 1996, for a survey of some of the positions on offer.)

[14] The picture sketched here shares certain similarities with the model put forth by Manski (2017). For different approaches to modeling similar phenomena, see Diecidue and Ven (2008); Tyson (2008).

satisfying. First—and this is a general worry for all proposals that concern different methods for how to decide how to decide—the model invites a regress. We start with a first-order decision between A and B, and end up considering a higher-order decision about how to select between A and B—by deliberating further or by coin-toss? Which of *those* we should select depends on features that—just like with A and B—might be somewhat opaque to us. If so, it looks like we must ascend further to determine whether we should continue to deliberate about whether to continue to deliberate about A and B or if we should stop deliberating about whether to continue to deliberate about A and B. How should we determine that? Do we keep climbing? The regress problem is, I think, a real one—but it'll be a problem for any view that attempts to take our computational limitations into account.[15]

The more serious problem is that, although the model specifies conditions under which it would be better to toss a coin than to deliberate, it doesn't rationalize tossing the coin over simply *picking* one of the options—and, if there's a cost to tossing the coin, it recommends picking over tossing. In other words, so long as *pick either A or B* is an available option, this simplified model will: (i) in some cases, recommend deliberating; (ii) in other cases, recommend picking; (iii) but, in no cases, recommend tossing the coin. After all, if picking is just like tossing a coin but without the coin, why toss a coin?

3.2 Tossing for Truth

The previous model was unable to rationalize selecting the *dice ex machina* because the model treats the coin-toss primarily as a (slightly costly) *picking procedure*, but there already is a (less costly) picking procedure available—namely, to simply pick one of the first-order options (A or B) without using a coin. On the picture this model depicts, deliberation is a (somewhat costly) way of gathering pertinent information about your first-order preferences, and tossing a coin is a way of selecting a first-order option. But perhaps this under-appreciates the role tossing a coin may play: in particular, like deliberation, perhaps tossing a coin can help us *learn* which of A or B we prefer.

How so? As the Danish poet (and scientist), Piet Hein, explains:

A PSYCHOLOGICAL TIP

[15] See, for example, Conlisk (1996, p. 687). The regress problem is, arguably, a problem even for those views that don't. For a particularly lucid articulation of the issue, see Resnik (1987, p. 11), who motivates (but doesn't endorse) the worry that, because all decisions ultimately depend on choices that aren't themselves the product of rational decision making (on pains of an infinite regress), no decisions are rational (see, also, Joyce, 1999; Smith, 1991, for further discussion of these issues.)

> Whenever you're called on to make up your mind,
> and you're hampered by not having any,
> the best way to solve the dilemma, you'll find,
> is simply by spinning a penny.
>
> No - not so that chance shall decide the affair
> while you're passively standing there moping;
> but the moment the penny is up in the air,
> you suddenly know what you're hoping.
>
> Piet Hein (1905–1996)

If Hein is correct that tossing the coin is also a method for gathering information about your preferences—and if you expect it to be a reliable and less costly method than continuing to deliberate about the matter—then the coin-toss very well might come out on top! Simply *picking* might be an effective way to select one of the two options, but—on this proposal—*tossing the coin* affords you a better shot at selecting the option you most prefer.

I agree that Hein's "psychological tip" depicts a familiar phenomenon—we sometimes do come to have a much clearer idea about which of our options we'd like to have, as the result of tossing a coin—but I don't think this suggestion is entirely successful either. For, while it is surely sometimes true that a coin-toss can help elicit relevant feelings about our options, this needn't be true of every case in which we resolve our indecision with the toss of a coin. In other words, sometimes, like when we're decidedly indifferent between our options, we toss the coin to *decide* (rather than help *determine*) what to do—and this paper is concerned with the former.

It's also unclear how and why such a method works (assuming it does). How does tossing a coin help me better know my own preferences? Jaffe et al. (2019), who demonstrate in a series of experiments that a coin-toss can act as a "catalyst" strengthening one's "affective reactions" to their choices, sketch the following speculative story:

> We assume when using a decision-making aid such as a coin-flip, the suggestion of one option over the other is not binding, but it may feel very real. Therefore, using a decision-making aid may result in a more vivid representation of the suggested option, which may be linked to stronger feelings [...]
>
> Strengthened feelings can then serve as pieces of information in their own right. [...] If the feelings that were strengthened by vividly imagining obtaining the suggested option are positive, individuals can use this information as an indicator for their actual preference. If, however, the feelings that were strengthened by vividly imagining

> obtaining the suggested option are negative, individuals can use this information to conclude that they might prefer the alternative option. Therefore, using a catalyst may result in an immediate feeling of which option individuals prefer, (Jaffe et al., 2019, p. 3)

It's unclear why the coin-toss (which one knows is *not* binding) would result in a more vivid representation of the suggested option than simulating the choice in some other way (e.g., by picking one at random). Perhaps the thought is that, by tossing the coin, you are tricking yourself into simulating a scenario in which the decision has been made—it's now out of your hands. But, even still, what does the coin have that pretending-to-make-up-your-mind lacks?

Furthermore, taking for granted that the coin-toss results in stronger feelings, it's not obvious that these feelings are a reliable guide to which of the options you prefer. I might have positive feelings when I vividly imagine receiving A—but that's consistent with it being the case that, were I to vividly imagine receiving B instead, I would have even stronger positive feelings. Same with feelings of regret. The fact that I feel pangs of regret when I vividly imagine selecting A doesn't entail that A isn't preferable to B. I might feel regret no matter which option is selected.

Finally, there's evidence that people are more willing to toss a coin when facing "low importance" decisions (e.g., choosing between a concert or the theater) than when facing "high importance" decisions (e.g., deciding who will be first author on a paper, custody disputes, matters of life-and-death).[16] When matters of importance are at stake, people prefer deliberating to coin-tossing. But, if tossing the coin is a means of gathering information about which option is best, this is puzzling. For there are, then, two methods for gathering this kind of evidence: using the coin-toss to elicit your feelings, and deliberation. The cost of tossing the coin (roughly, the risk of dropping it) is constant, whereas the cost of deliberation is presumably sensitive to the importance of the decision. The higher and more important the stakes, the more stressful the deliberation. So, offhand, we should expect the opposite: as the stakes increase in importance, should people be *more* inclined to toss the coin?

Tossing a coin can, in some circumstances, help you sort out what you want. That's true (even if it's mysterious exactly how it helps). Nevertheless, that can't be the whole story.

---

[16] See, in particular, Keren and Teigen (2010). These studies are also discussed in Jaffé et al. (2020). In addition to importance, the propensity to toss a coin is sensitive to several other dimensions as well. For example, people are more likely to toss a coin when the decision is easy than when difficult (although it's not entirely clear how the participants understood this distinction), people are less likely to toss a coin when the decision affects others than they are when it primarily only affects themselves, and people are more likely to toss a coin later, rather than earlier, in the decision-making process. (For discussion of using coin-flips in custody disputes (as well as many other interesting topics), see Elster (1989).)

4 Expressing Your Attitude

Here's a different story. Sometimes we might toss a coin to help us determine what we want, but not always. Sometimes, we toss a coin to *express* something about our attitudes toward our options: namely, that we don't prefer one to the other.

In sketching the proposal, I want to do two things. First, I want to demonstrate its *extensional adequacy*: it correctly distinguishes between the cases in which we're disposed to flip a coin and those in which we aren't. Second, I want to make plausible the idea that we often reasonably care about what our choices say about us—and we should care about this.

To the end of demonstrating the proposal's implications, consider the following simplified agent. In addition to this agent's other goals (whatever they may be), she has a slight preference for onlookers (real or hypothetical) to come to have accurate beliefs about her motivational-state. Our agent longs to be (correctly) understood. The onlookers form their beliefs about the agent's motivational-state in roughly the same way that we all typically form such beliefs about each other: largely on the basis of observations of choice-behavior. If we observe that the agent selected A from the menu $\{A, B\}$ we (provisionally) conclude that she doesn't disprefer A to B. (I say 'provisionally' because agents can suffer temporary bouts of irrationality, alternatives can be selected accidentally, etc. Really, our conclusions should reflect a more holistic picture of the agent's choice-behavior, by ascribing to her those attitudes that make the most sense of her choices modulo a principle of charity, and the like.) Importantly, us onlookers cannot peer directly into the agent's soul, determine her intentions are good, and ensure she won't be misunderstood. We don't have access to facts about her *dispositions* to choose, we can't control the menus of options she encounters, and she can't simply tell us (and, because actions speak louder than words, we wouldn't necessarily believe her even if she could). Our evidence is, the grand scheme of things, meager.

Knowing that her actual choice-behavior is all we have to go on, if the agent wants us to form accurate beliefs about her attitudes, she should (all else equal) try her best not to confuse us. Suppose she prefers A to B. If she encounters the menu $\{A, B\}$, what should she do? She should choose A. It would be rash of us, the onlookers, to infer from her selection that she prefers A to B—it is consistent with her choice-behavior that she is *indifferent* between the two. But, so it seems, selecting A is the best she can do to give us onlookers the material to come to the right conclusions. (She could *routinely* select A from the menu $\{A, B\}$—that would be helpful. But I'm assuming she typically doesn't have the power to control which menus she encounters.) *Mutatis mutandis* if she prefers B to A. But what if she's indifferent? What then should she do?

If she is indifferent between A and B, and she encounters the menu $\{A, B\}$, it's permissible for her to select either. If she would like us to know that she is

indifferent, what can she do? Selecting one over the other is, as we just saw, compatible with having a preference. Here's a suggestion. The agent could, in cases in which she has a preference, *choose*; and, when indifferent, *pick*. (And, if Ullmann-Margalit and Morgenbesser (1977) are right, that is in fact what she will do.) Unfortunately, in a one-off decision, this is no help to us onlookers because, from the outside, we cannot discern the difference between a *choice* and a *pick*; all we observe is which option is selected. The agent could, if she encounters the menu again and again, sometimes select A and sometimes select B. But, again, the agent doesn't have control over which menus she happens by.[17]

Here's a final suggestion. If she has a coin, she can toss it to decide which to select. Because tossing the coin isn't dispreferred to A or B, it would be permissible for her to do it. Additionally, selecting the coin-toss *expresses* her indifference between A and B—it provides strong evidence to us onlookers that she lacks a preference between A and B. How so? If she didn't—that is, if she preferred one to the other—it *wouldn't* be permissible for her to toss the coin. Tossing the coin would have lower expected value than the preferred option. Because electing to toss the coin is incompatible with having a preference between A and B, if she tosses the coin, she must not have one. The onlookers can, thus, infer her lack of a preference from her choice to toss the coin. On the other hand, the other permissible options—A, B—are, if performed, much less informative. So, if the agent longs to be understood, she has a reason to toss the coin rather than to pick A or B.

That's the rough idea, but let me spell out one of its implications. The reason the agent had to toss the coin, rather than select A or B, was that tossing the coin *expressed* something about her motivational-state that wouldn't have been expressed had she selected otherwise. In general, though, if some fact is in the *common ground*—we all accept it, and accept that we all accept it, and so on—it will no longer be necessary (or maybe even possible) to assert it. There's no reason to tell me my pants are on fire if it's already common ground that they are. Likewise, if it's *clear and obvious* that the agent is indifferent between A and B, tossing the coin losses its unique appeal. We all already know. So, here's a prediction: if the agent wants to be understood, we should expect her to toss the coin when it's less than obvious that she lacks a preference between A and B, but not expect a coin-toss when it *is* obvious that she's indifferent.

This point dovetails with the discussion of the three types of *indifference* in §2.3. The first one—Type I—concerned cases in which you're indifferent between two items because they are (nearly) qualitatively identical. These were the examples that characterized the supermarket scenarios in Ullmann-Margalit and Mor-

[17] Also, that behavior—sometimes selecting A, sometimes B—is compatible with having a preference that routinely toggles between the two.

genbesser (1977). If A and B are qualitatively identical—like two different boxes of Lucky Charms sitting on the supermarket shelf—it should be clear and obvious to all onlookers that the agent is indifferent. It simply wouldn't be reasonable—not for creatures like us—to have a preference for one over the other. It would also be quite odd, I think, for one to select a box of Lucky Charms from the supermarket with the toss of a coin. That would be pathological. Ullmann-Margalit and Morgenbesser (1977) are correct about the supermarket—one should just *pick*.

Type II cases concerned indifference that resulted from symmetrical knowledge—e.g., you know that A and B are, as a matter of fact, different from each other, but your evidence about which one is which is symmetrical. This is a more complicated case because it involves both preference and belief. If our agent wants onlookers to have accurate beliefs, not just about her motivational-state, but about her beliefs as well, then whether she should toss the coin to decide depends on how private and equivocal she thinks her evidence is. If she thinks onlookers think it's possible for her to have evidence that the prize is behind Door #1 rather than Door #2, when she doesn't, she might want to toss a coin to make clear that she doesn't.

Type III cases—the ones in which your options differ from each other along various dimensions—are the ones where we should expect the agent to toss the most coins. If she is indifferent between A and B, but it wouldn't be unreasonable for someone to have a preference between the two, tossing the coin would be a way to make herself understood.

We can now offer an explanation for why someone might toss a coin to select between A and B, and yet forgo tossing the coin to decide between the two different ways of using that coin to select between A and B (i.e., $Toss_{(A,B)}$ and $Toss_{(B,A)}$). If the agent's indifference is of Type III, tossing the coin to select between them is a way of communicating to the onlookers that you are indifferent. However, your indifference concerning the two different ways of using that coin to select between A and B—$Toss_{(A,B)}$ and $Toss_{(B,A)}$—is not of Type III. Because the coin is fair, even if you had a preference between A and B (which you don't), you shouldn't have one between $Toss_{(A,B)}$ and $Toss_{(B,A)}$. No reasonable person could have a preference between the two (even if a reasonable person could have a preference between A and B). Therefore, there's no need to express your indifference with a coin toss—doing so would be redundant.

There's more that can be said about what this simplified picture entails about particular cases, but let's set that aside and turn now to the question of how to interpret the simplified picture. Why should it matter what an agent in a highly idiosyncratic situation—where she both longs to be known and is surrounded by attentive onlookers—would do with a coin? While the picture is, of course, simplified in some ways and exaggerated in others, it isn't as idiosyncratic as it

might seem on the surface. The picture has two features that require attention. First, there is the agent's desire for understanding—for her attitudes to be known by others. Second, there are the others—the throng of omnipresent onlookers.

Although oversimplified, the desire to be understood by one's community is not terribly unusual. Is it plausible to think that people have a desire like this? And, whether or not they do, is it a reasonable desire to have? The first of these is an empirical question, and while my evidence is largely anecdotal, I suspect that many people do have a drive to be seen accurately by others. It is, I also think, a very reasonable desire to have. Our success often depends on our ability to coordinate, and understanding each other's motivations is instrumental to our ability to do so successfully. In order to coordinate with you, I need to be able to make reasonably accurate predictions about what you would do in various circumstances. It's hard to make accurate predictions when ignorant or incorrect about someone's motivations. So, whatever other goals you have, making yourself easy to understand—particularly for those with whom you must coordinate—is likely a wise idea.

While the desire to be understood turns out to be less unusual than it may have first seemed, the model's other assumptions—regarding the vigilance of the onlookers, in particular—aren't terribly realistic. It's not true that your every choice is subject to the scrutiny of attentive onlookers. We sometimes select in private. Even in public, no one is paying very much attention to you.

That's true, but let me make two points in response. First, because it isn't always easy to tell when we're being watched, and because it is easier for creatures like us to internalize and follow coarser- rather than finer-grained rules, it might turn out that acting *as if* your choices are apparent is the best of the feasible alternatives. Second, your choices are all apparent to someone—yourself. This might seem utterly irrelevant—after all, *you* presumably already know whether your selection of A was the result of a preference or a picking. And while that is, to some extent, undeniably true, it overlooks the various ways in which we—parts of us, at least—can be opaque to ourselves. This is especially so (and less Freudian) in *diachronic* contexts. I'll have an easier time tomorrow remembering what I chose to do today than how I might have felt while doing it. And so a desire to better understand myself tomorrow can underwrite what I choose to do today.

Then, again, the best laid schemes of dice and men gang aft agley.

References

Agranov, Marina and Pietro Ortoleva. 2017. Stochastic Choice and Preferences for Randomization. *Journal of Political Economy*, 125(1): 40–68

Baccelli, Jean and Philippe Mongin. 2021. Can Redescriptions of Out-

comes Salvage the Axioms of Decision Theory? *Philosophical Studies*, https://doi.org/10.1007/s11098-021-01723-z

Bermudez, Jose Luis. 2009. *Decision Theory and Rationality* (Oxford: Oxford University Press)

Broome, John. 1984a. Selecting People Randomly. *Ethics*, 95(1): 38–55

——. 1984b. Uncertainty and Fairness. *Economic Journal*, 94(375): 624–632

——. 1990. Fairness. *Proceedings of the Aristotelian Society*, 91: 87–101

——. 1991. *Weighing Goods* (Oxford: Blackwell Publisher)

——. 1993. Can a Humean be Moderate? In *Value, Welfare and Morality*, edited by R. G. Frey and Christopher Morris (Cambridge, UK: Cambridge University Press), pp. 279–286

Buchak, Lara. 2013. *Risk and Rationality* (Oxford: Oxford University Press)

Camerer, Colin and Teck-Hua Ho. 1994. Violations of the Betweenness Axiom and Nonlinearity in Probability. *Journal of Risk and Uncertainty*, 8(2): 167–196

Conlisk, John. 1996. Why Bounded Rationality? *Journal of Economic Literature*, 34(2): 669–700

Diamond, Peter A. 1967. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment. *Journal of Political Economy*, 75(5): 765–766

Diecidue, Enrico and Jeroen Van De Ven. 2008. Aspiration Level, Probability of Success and Failure, and Expected Utility. *International Economic Review*, 49(2): 683–700

Dreier, James. 1996. Rational Preference: Decision Theory as a Theory of Practical Rationality. *Theory and Decision*, 40: 249–276

Dwenger, Nadja, Dorothea Kübler, and Georg Weizsäcker. ms. Flipping a Coin: Theory and Evidence. Ms

Elster, Jon. 1989. Taming Chance: Randomization in Individual and Social Decisions. In *Solomonic Judgements: Studies in the Limitations of Rationality*, chap. II (Cambridge, UK: Cambridge University Press)

Glimcher, Paul W. 2005. Indeterminacy in Brain and Behavior. *Annual Review of Psychology*, 56: 25–56

Hausman, Daniel. 2012. *Preference, Value, Choice, and Welfare* (Cambridge, UK: Cambridge University Press)

Hooker, Brad. 2005. Fairness. *Ethical Theory and Moral Practice*, 8(4): 329 – 352

Jaffé, Mariela E., Maria Douneva, and Rainer Greifeneder. 2020. Solve the dilemma by spinning a penny? On using random decision-making aids. *Judgment and Decision Making*, 15(4): 561–571

Jaffe, Mariela E., Leonie Reutner, and Rainer Greifeneder. 2019. Catalyzing decisions: How a coin flip strengthens affective reactions. *PLoS ONE*, 14(8): 1–14

Jeffrey, Richard. 1987. Risk and Human Rationality. *The Monist*, 70(2): 222–236

——. 1988. Biting the Bayesian Bullet: Zeckhauser's Problem. *Theory and Decision*, 25: 117–122

Joyce, James. 1999. *Foundations of Causal Decision Theory* (Cambridge, UK: Cambridge University Press)

Keren, Gideon and Karl H. Teigen. 2010. Decisions by coin toss: Inappropriate but fair. *Judgment and Decision Making*, 5(2): 83–101

Little, I.M.D. 1950. *A Critique of Welfare Economics* (Oxford: Oxford University Press)

Manski, Charles F. 2017. Optimize, Satisfice, or Choose without Deliberation? A Simple Minimax-regret Assessment. *Theory and Decision*, 83: 155–173

Neumann, M. 2007. Choosing and Describing: Sen and the Independence of Irrelevant Alternatives. *Theory and Decision*, 63: 79–94

Raiffa, Howard. 1968. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty* (Reading, MA: Addison-Wesley)

Rescher, Nicholas. 1960. Choice Without Preference: A Study of the History and of the Logic of the Problem of "Buridan's Ass". *Kant-Studien*, 51(1-4): 142–175

Resnik, Michael D. 1987. *Choices: An Introduction to Decision Theory* (University of Minnesota Press)

Rulli, Tina and Alex Worsnip. 2016. IIA, Rationality, and the Individuation of Options. *Philosophical Studies*, 173: 205–221

Samuelson, Paul A. 1938. A Note on the Pure Theory of Consumer's Behaviour. *Economica*, New Series, 5(17): 61–71

Schelling, Thomas C. 1960. *The Strategy of Conflict* (Cambridge: Harvard University Press)

Sen, Amartya. 1993. Internal Consistency of Choice. *Econometrica*, 61(3): 495–521

——. 1997. Maximization and the Act of Choice. *Econometrica*, 65(4): 745–780

Sher, George. 1980. What makes a Lottery Fair? *Nous*, 14(2): 203–216

Sippel, R. 1997. An Experiment on the Pure Theory of Consumer's Behaviour. *The Economic Journal*, 107: 1431–1444

Smith, Holly. 1991. Deciding How to Decide: Is There a Regress Problem? In *Foundations of Decision Theory*, edited by M. Bacharach and S. Hurley (Oxford: Basil Blackwell)

Tversky, Amos. 1975. A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, 9: 163–173

Tyson, Christopher J. 2008. Cognitive Constraints, Contraction Consistency, and the Satisficing Criterion. *Journal of Economic Theory*, 138: 51–70

Ullmann-Margalit, Edna and Sidney Morgenbesser. 1977. Picking and Choosing. *Social Research*, 44(4): 757–785

Voorhoeve, Alex and Marc Fleurbaey. 2012. Egalitarianism and the Separateness of Persons. *Utilitas*, 24(3): 381–398

Wasserman, David. 1996. Let them Eat Chances: Probability and Distributive Justice. *Economics & Philosophy*, 12(1): 29–49