

Opaque Sweetening & Transitivity

Abstract

I argue that any plausible decision theory for agents with incomplete preferences which obeys the *Never Worse Principle* will violate Transitivity. The *Never Worse Principle* says that if one option never does worse than another, you shouldn't disprefer it. Transitivity says that if you prefer X to Y and you prefer Y to Z , then you should prefer X to Z . Violating Transitivity allows one to be money pumped. Although agents with incomplete preferences are already, in virtue of having incomplete preferences, vulnerable to being money pumped, I argue that the money pump argument for Transitivity is more serious than the one for Completeness.

1 Introduction

A preference ordering is *complete* just in case, for any two alternatives X and Y , you either prefer X to Y , prefer Y to X , or are indifferent between the two. Are you rationally required to have complete preferences? Many, myself included, think you are not. One reason your preferences might fail to be complete is if there are alternatives whose values you take to be incommensurable. You might think, for example, that the pros and cons of X fail to weigh-up against the pros and cons of Y in a manner that's precise enough to deliver an all-things-considered preference one way or the other. Let's say that, in such cases, you are *ambivalent between X and Y* .¹

¹ The phenomenon I'm gesturing toward is related to what Ruth Chang calls "parity" ([Chang, 2002, 2005]), James Griffin called "rough equality" ([Griffin, 1986, p. 80-98], [Parfit, 1984, p. 431]), what

One notable feature of ambivalence is that, unlike indifference, it is *insensitive to mild sweetening*: if you're ambivalent between X and Y , you will sometimes also be ambivalent between a slightly improved version of X and Y . Here's an example. Suppose you are deciding between two vacations: you can go on an adventurous skiing trip in the Alps (A) or you can take a relaxing trip to the beach (B). There are various things to be said in favor of each. You don't prefer either one to the other. Moreover, you don't prefer the beach vacation *plus a couple dollars* ($B^{+\$2}$) to the ski vacation, and you don't prefer the ski vacation *plus a couple dollars* ($A^{+\$2}$) to the beach vacation. Small improvements aren't enough to resolve your ambivalence between the alternatives.

How should you rationally evaluate your options when you're ambivalent between at least some of their potential outcomes? Traditional decision theories enjoin you to *maximize expected utility*. But if your preferences are incomplete, they cannot be represented with a utility-function; and, thus, there is no well-defined quantity you can maximize the expectation of. Hare [2010] develops two conflicting ways of generalizing traditional decision theory when your preferences fail to be complete. The proposals disagree about cases like the following.

Opaque Sweetening: Vacation Boxes. There are two opaque boxes: a Larger box (L) and a Regular box (R). A fair coin was tossed. If the coin landed heads, then a voucher for an all-expenses-paid Alpine ski vacation (A) was placed in the Larger box and a voucher for an all-expenses-paid beach vacation (B) was placed in the Regular box; if the coin landed tails, then B was placed in the Larger box and A was placed in the Regular box. In either case, you don't know which prize is in which box.

some economists call "indecisiveness" ([Ok et al., 2012]), and what's sometimes characterized as evaluative indeterminacy ([Broome, 2007]). I take no stand on which, if any, of these views is correct. Furthermore, I intend 'ambivalence' to refer to a subjective, psychological state: in addition to *preference* and *indifference*, you can be in the psychological state of being *ambivalent between* two alternatives. The psychological state is conative, not cognitive: ambivalence might correspond to, but is not identical with, a belief about the objective value-relations that hold between the items.

$$\text{Larger box} = \begin{cases} A & \text{if Heads} \\ B & \text{if Tails.} \end{cases} \quad \text{Regular box} = \begin{cases} B & \text{if Heads} \\ A & \text{if Tails.} \end{cases}$$

Suppose \$2 is added to the Larger box. If you choose the Larger box, you will win whichever prize it contains plus \$2. Nothing is added to the Regular box. You are asked to choose one of the two boxes, taking home whichever prize it contains.

| | HEADS | TAILS |
|---|-------|-------|
| L | A+\$2 | B+\$2 |
| R | B | A |

According to *Prospectism* [Hare, 2010, 2013], you should evaluate your options solely in terms of their corresponding prospects, which are the probability-distributions over their potential outcomes.² L’s corresponding prospects (a 50% chance of A+\$2 and a 50% chance of B+\$2) are better than R’s (a 50% chance of B and a 50% chance of A). Therefore, according to *Prospectism*, you should prefer L to R.

On the other hand, you know, no matter how the coin landed, you don’t prefer the outcome that would result from choosing L to the outcome that would result from choosing R. If the coin landed heads, L contains A+\$2 and R contains B. You don’t prefer A+\$2 to B. If the coin landed tails, L contains B+\$2 and R contains A. You don’t prefer B+\$2 to A. So, in either case, you don’t prefer the contents of L to the contents of R. You’re in a position to know which value-relation holds between your options: L isn’t better (or worse) than R. Several philosophers, like Bales et al. [2014] and Schoenfield [2014], have argued that, for this reason, you’re not rationally required to prefer L to R.³

² More carefully, *Prospectism* represents your preferences with the set of utility-functions corresponding to all the coherent extensions of your incomplete preferences. If each utility-function in the set ranks one option ahead of another, then you’re rationally required to prefer it. In addition to Hare’s *Prospectism*, there are a number of other views with a similar structure: e.g., [Good, 1952], [Levi, 1986], [Sen, 2004], and [Weirich, 2004].

³ Bales et al. [2014] appeal to a principle — *Competitiveness* — that says if an option never does worse

According to the former view, you're rationally required to prefer L to R ; according to the latter, you're not. Which view is correct? It's not the aim of this paper to address that question (at least not directly). Instead, I'm interested in a related question: How could this latter view — the one defended by [Bales et al. \[2014\]](#) and [Schoenfield \[2014\]](#) — be generalized into a full-fledged decision theory? And what would such a decision theory look like? Relatively little attention has been devoted to this question.⁴

I will argue that things don't look great. Any plausible decision theory that recommends taking either box comes with considerable costs: it saddles you with non-transitive preferences. There is, then, a tension between two attractive ideas. The first is that you shouldn't prefer one box to the other if you know that you're ambivalent between their contents. The second is that your preferences should be transitive. One of the two must go. Which one? That isn't something that I will settle here. My aim is merely to show that we can't have both.

2 Two Constraints

The argument proceeds in two steps. First, I will set forth two constraints that must be obeyed by any plausible decision theory that recommends taking either box. Then, I will show that any decision theory meeting those constraints violates Transitivity: it will say of some options, X , Y , and Z , that you should prefer X to Y , that you should prefer Y to Z , but that you shouldn't prefer X to Z .

The first constraint — which is analogous to those of [Bales et al. \[2014\]](#) and [Schoenfield \[2014\]](#) — ensures that the resulting decision theory recommends taking

than the alternatives, it's permissible. R can never do worse than L ; so, according to their principle, it's permissible. [Schoenfield \[2014\]](#) appeals to a principle — *Link* — that says you shouldn't prefer one option to another if you're certain it's actually not worse. You're certain R isn't actually worse than L ; so, according to her principle, you shouldn't prefer L . [Hare \[2010\]](#), too, considers a principle — *Deference* — that says it's permissible to adopt the attitude toward your options that you know any fully-informed, rational person would have on your behalf. You know that any fully-informed, rational person with your best interests at heart wouldn't prefer L to R (she would be ambivalent); so, according to his principle, it's permissible to not prefer L to R .

⁴ Although neither [Bales et al. \[2014\]](#) nor [Schoenfield \[2014\]](#) develop a full-fledged decision theory of this kind, [Hare \[2010\]](#) does: *Deferentialism*. But — especially in comparison to *Prospectism* — it remains fairly underexamined. Assessing it (and its close cousins) is the aim of this paper.

either option in cases, like Vacation Boxes, of opaque sweetening.

[THE NEVER WORSE PRINCIPLE]

Let $S = \{S_1, S_2, \dots, S_n\}$ be a partition of the ways the world might be. If, for all $S \in S$, you don't prefer $(\psi \wedge S)$ to $(\varphi \wedge S)$, then you shouldn't prefer ψ to φ .

Decision problems can be characterized by three sorts of entities: the *options* that are available to you, the *states* that the world might be in, and the *outcomes* that might result from taking an option depending on which state obtains. Your options (e.g., φ , ψ) are the things you can do, and are the objects of your instrumental preferences. For example, in Vacation Boxes, you have two available options: take the Larger box (L), and take the Regular box (R). The states (e.g., S_1 , S_2 , etc.) are ways the world might be which are outside your control. In this paper, we will confine our attention to cases in which the relevant states are both evidentially and causally *independent* of your options. In other words, we'll assume, for all options φ and all states S , that φ ing neither provides you evidence for, nor causally influences, S being the case. In Vacation Boxes, there are two relevant states: the coin landing heads (HEADS), and the coin landing tails (TAILS). The outcomes, which are the objects of your non-instrumental preferences, are determined by which option you take and which state obtains (e.g., $(\varphi \wedge S)$, $(\psi \wedge S)$). In Vacation Boxes, because (let us assume) you care about the prizes themselves and not the ways in which you might receive them, there are four possible outcomes: receiving $A^{+\$2}$, receiving $B^{+\$2}$, receiving B , and receiving A .⁵

Say that φ *never does worse* than ψ if, for each of the ways you think the world might be, you don't prefer the outcome that would result from ψ ing to the outcome that would result from φ ing. The *Never Worse Principle* says, then, that if φ never does worse than ψ , you shouldn't prefer ψ to φ . Because R never does worse than L , according to the *Never Worse Principle*, you shouldn't prefer L .

⁵ In other words, I will assume that you're indifferent between $(L \wedge \text{HEADS})$ and $A^{+\$2}$, that you're indifferent between $(R \wedge \text{TAILS})$ and A , and so on. If you, e.g., value the coin landing heads for its own sake or, e.g., value possessing the Regular box for some reason other than what it might contain, this assumption won't hold. For simplicity, though, let's assume that it does. Additionally, let's assume that you value the prizes independently of each other: the value of having one doesn't affect the value of having another.

The second constraint is one, I argue, any full-fledged decision theory must satisfy in order to be plausible.⁶

[INTER-STATE TRADE-OFFS]

Suppose you prefer X^+ to X and prefer Y^+ to Y . Let φ be an option that, were you to take it, would result in X^+ if E is the case, and in Y if it isn't. Let ψ be an option that, were you to take it, would result in X if E is the case, and in Y^+ if it isn't.

| | | |
|-----------|-------|----------|
| | E | $\neg E$ |
| φ | X^+ | Y |
| ψ | X | Y^+ |

If $Cr(E) \geq Cr(\neg E)$ and your preference for X^+ over X is stronger than your preference for Y^+ over Y , then you are rationally required to prefer φ to ψ .

The rough idea underlying the principle is the following: if your preference for φ over ψ , conditional on E being the case, is stronger than your preference for ψ over φ , conditional on E not being the case, then, so long as you don't think $\neg E$ is more likely than E , you should prefer φ to ψ .

The principle makes reference to an as-of-yet unexplained, but fairly intuitive, notion: the relative strength of a preference. What is it for your preference for X^+ over X to be *stronger* than your preference for Y^+ over Y ? If your preferences were complete (and could thus be represented with a utility-function), the notion is clear enough: your preference for X^+ over X is stronger than your preference for Y^+ over Y just in case $u(X^+) - u(X) > u(Y^+) - u(Y) > 0$. But what about when your preferences are incomplete (and thus cannot be represented with a utility-function)? We can make use of the same idea by generalizing it in the following way. First, although your preferences, if they are incomplete, cannot be represented with a *single* utility-function, they can be represented with a *set* of utility-functions. Consider all of the *coherent extensions* of your preferences: that is, all of the ways of rendering

⁶ The constraint is only meant to hold when the states and your options are independent. If, for example, choosing φ will make it very likely that E is false, then it's not true that you should choose it.

your preferences complete while holding fixed the preferences you do have.⁷ Then we represent each of these coherent extensions with a utility-function. Let \mathcal{U} be the set of all these utility-functions. It represents your incomplete preferences. Finally, let's say that your preference for X^+ over X is *stronger* than your preference for Y^+ over Y just in case, for every u in \mathcal{U} , $u(X^+) - u(X) > u(Y^+) - u(Y) > 0$.

With that notion now made precise, let's turn our attention to defending *Inter-State Trade-Offs*. If your preferences could be represented with a single utility-function, the principle is clearly correct; it follows straightforwardly from traditional decision theory. Here's why. If your preference for X^+ over X is stronger than your preference for Y^+ over Y , then $u(X^+) - u(X) > u(Y^+) - u(Y)$. So, if $Cr(E) \geq Cr(\neg E)$, then $Cr(E)(u(X^+) - u(X)) > Cr(\neg E)(u(Y^+) - u(Y))$. And so, $EU(\varphi) = Cr(E)u(X^+) + Cr(\neg E)u(Y) > Cr(E)u(X) + Cr(\neg E)u(Y^+) = EU(\psi)$. That is to say: φ 's expected utility is higher than ψ 's. Therefore, according to traditional decision theory, you should prefer φ to ψ . But what if your preferences cannot be represented with a single utility-function? Is this constraint one that it's reasonable to demand *any* full-fledged decision theory to satisfy? I think so.

Allow me to motivate the constraint with an example. Suppose you are given the Regular box as a gift. You know the box contains B if the coin landed heads and A if it landed tails. But, because you don't know how the coin landed and because you've yet to open the box, you don't know which prize you have. Before you get the chance to find out, you are offered the following deal. You can accept a bet that pays \$10 if the coin landed heads. Or you can accept a bet that pays \$1 if the coin landed tails.⁸

| | HEADS | TAILS |
|--------------|-------------|------------|
| Bet on Heads | $B^{+\$10}$ | A |
| Bet on Tails | B | $A^{+\$1}$ |

⁷ More formally, let's say that \succeq^+ is a coherent extension of your (potentially incomplete) preferences just in case (1) for all X and Y , $X \succeq^+ Y$ if you weakly prefer X to Y , and (2) \succeq^+ is complete.

⁸ In this example, the bets turn on the same coin toss that determined which prize is in the box. This is an inessential feature of the example. An example without this feature — one in which the bets turn on, say, the flip of some other coin — would work nearly as well, but would complicate the presentation somewhat.

Which bet should you prefer? The answer, so it seems to me, is obvious: you should prefer to bet on heads. It's clear, setting aside your possession of the Regular box, that betting on heads is better than betting on tails: both have an equal chance of winning, but the former promises a larger win than the latter. Because which bet you take has no effect on which prize is in the box, the fact that the box is in your possession should have no effect on which you prefer. *Inter-State Trade-Offs* is what ensures this result. Because your preference for $B^{+\$10}$ over B is stronger than your preference for $A^{+\$1}$ over A , and because you don't think it more likely that the coin landed tails than that it landed heads, *Inter-State Trade-Offs*, correctly, entails that you should prefer to bet on heads. Any decision theory that violates this constraint will be implausibly permissive.

Moreover, the constraint is consistent with the presumable reason for rejecting views, like *Prospectism*, that evaluate options in terms of their corresponding prospects. The prospects of an option abstract away from the facts about which outcomes "reside in" which states. The 50% chance of getting $A^{+\$2}$ ($B^{+\$2}$) if you take L outweighs the 50% chance of getting A (B) if you take R . And that's that. But, proponents of the *Never Worse Principle* will object that this involves illegitimately comparing outcomes that reside in one state to ones that reside in others. In Vacation Boxes, the relevant comparison should be between, e.g., $A^{+\$2}$ and B , not $A^{+\$2}$ and A . Here's why this might be so. We can think of the states (e.g., E , $\neg E$) (so long as they are independent of your options) as corresponding to hypotheses about how valuable your options, in comparison to each other, *actually* are. For example: if, for all states S , $(\psi \wedge S) \not\prec (\varphi \wedge S)$, you can be certain that ψ isn't actually more valuable to you than φ . And what you should ultimately care about is how valuable your options *actually* are.⁹

Inter-State Trade-Offs, however, doesn't involve making such illegitimate comparisons. Rather, it involves comparing *differences* in value between outcomes *in the same state*. Again, if we think of states as corresponding to hypotheses about how valuable your options actually are, the constraint can be put as follows: Sup-

⁹ Schoenfield [2014], in particular, appeals to something like this idea in defending a principle that's analogous to the *Never Worse Principle*: e.g., "[i]f expected value theory required us to make choices that we are certain would lead to no improvement in value, then expected value theory is imposing requirements that transcend what we actually care about: the achievement of *value*" [Schoenfield, 2014, p. 268]. See [Bader, 2017] and [Doody, 2018], however, for criticisms of this defense.

pose you know that the extent to which φ might be better than ψ is greater than the extent to which ψ might be better than φ ; then the only reason you could have to not prefer φ to ψ is if you think it's likely that ψ is actually better than φ . The constraint doesn't presume that you'll always be able to make such comparisons. All it says is that, when you can make such comparisons, you should take them into account in the manner described. I can see no way of developing a plausible full-fledged decision theory that denies this.

3 Failure of Transitivity

I'll now show that any decision theory meeting these two constraints — *The Never Worse Principle* and *Inter-State Trade-Offs* — is problematic: it violates Transitivity.

Consider the following cases:

The Ski Button. In front of you is a big button, which reads “A.” A fair coin has been tossed. If you press the button, you win the Alpine ski vacation if and only if the coin landed heads. If you don't press the button, you win the Alpine ski vacation if and only if the coin landed tails. Also, if you press the button, I'll give you a dollar.

| | HEADS | TAILS |
|-------------|------------|-------|
| Press “A” | $A^{+\$1}$ | $\$1$ |
| Don't Press | o | A |

You ought to press the button. You know the coin is fair, so $Cr(\text{HEADS}) = Cr(\text{TAILS})$. Moreover, you prefer $A^{+\$1}$ to A , A to $\$1$, and $\$1$ to o (the status quo); so, your preference for $A^{+\$1}$ over o is stronger than your preference for A over $\$1$.¹⁰ So, according to *Inter-State Trade-Offs*, you should prefer pressing the button to not pressing it. And, intuitively, that seems right.

¹⁰ More carefully: because you prefer $A^{+\$1}$ to A , A to $\$1$, and $\$1$ to o , every utility-function u in \mathcal{U} is such that $u(A^{+\$1}) > u(A) > u(\$1) > u(o)$; and so, for all u in \mathcal{U} , $u(A^{+\$1}) - u(o) > u(A) - u(\$1)$.

The Beach Button. In front of you is another big button. This one reads “*B*.” The same coin toss has occurred. If you press the button, you win the beach vacation if and only if the coin landed tails. If you don’t press the button, you win the beach vacation if and only if the coin landed heads. Also, if you press the button, I’ll give you a dollar.

| | HEADS | TAILS |
|--------------------|-------|-------------|
| Press “ <i>B</i> ” | \$1 | $B^{+\$1}$ |
| Don’t Press | B | \emptyset |

Again, and for analogous reasons, you ought to press the button.

Imagine you face these decisions in sequence: you must decide whether or not to press “*A*” and then whether or not to press “*B*”. As we’ve seen, in each case you prefer pressing to not. Imagine that, before you can carry out your plan to push both, you are told that there’s been a minor malfunction: the buttons are stuck together — by pressing one, you’ll also thereby press the other. “No matter,” you might say, “I was going to push both anyway.” But not so fast! According to the *Never Worse Principle*, it’s not the case that you should press both rather than press neither.

| | HEADS | TAILS |
|---------------|------------|------------|
| Press Both | $A^{+\$2}$ | $B^{+\$2}$ |
| Press Neither | B | A |

Pressing “*A*” is like taking a gamble that pays out $A^{+\$1}$ if heads and a dollar if tails. Pressing “*B*” is like taking a gamble that pays out $B^{+\$1}$ if tails and a dollar if heads. So, pressing both is like a gamble that pays out $A^{+\$2}$ if heads and $B^{+\$2}$ if tails. It’s structurally identical to *L*. Not pressing “*A*” is like taking a gamble that pays out A if tails and nothing otherwise. Not pressing “*B*” is like taking a gamble that pays out B if heads and nothing otherwise. So, pressing neither is like a gamble that pays out B if heads and A if tails. It’s structurally identical to *R*. Thus, according to the *Never Worse Principle*, you shouldn’t prefer pressing both to pressing neither.

This is puzzling. You prefer pressing “A” to not, and you prefer pressing “B” to not, but you don’t prefer pressing both to pressing neither.¹¹ In fact, as we’ll see, your preferences violate Transitivity.

[TRANSITIVITY]

If you prefer X to Y and you prefer Y to Z , then you should prefer X to Z .

When facing both decisions, there are four options: you can press “A” and then press “B” (Press Both); you can press “A” and then not press “B” (Press Only “A”); you can not press “A” and then press “B” (Press Only “B”); or you can not press “A” and then not press “B” (Press Neither). *Inter-State Trade-Offs* entails that you should prefer Press Both to Press Only “A” and that you should prefer Press Only “A” to Press Neither.¹² But, as we’ve seen, the *Never Worse Principle* entails that you shouldn’t prefer Press Both to Press Neither. Transitivity is violated.

Press Both \succ Press Only “A”
 Press Only “A” \succ Press Neither
 Press Both $\not\succeq$ Press Neither

Why should you prefer Press Only “A” to Press Neither? Recall that by pressing “A” you’ll get $A^{+\$1}$ if heads and a dollar if tails, and that by not pressing “B” you’ll get B if heads and nothing if tails. So, Press Only “A” is equivalent to a gamble that pays out both prizes plus a dollar ($A^{+\$1} \wedge B$) if heads and a dollar if tails. And, as we’ve seen, Press Neither is equivalent to a gamble that pays out B if heads and A if tails.

¹¹ The example shows that decision theories satisfying the *Never Worse Principle* violate an agglomeration principle. Agglomeration says: if you ought to φ and you ought to ψ , then you ought to φ and ψ . In Ski Button, you ought to press “A”; in Beach Button, you ought to press “B”; but it’s not the case that you ought to press “A” and “B”. One might think this is reason enough to reject such theories. (Hare [2016], for example, argues against a class of moral theories on the grounds that they violate agglomeration.) I think there are independent reasons to reject agglomeration in these contexts, though, and so will not pursue such an argument here.

¹² It also entails that you should prefer Press Both to Press Only “B” and that you should prefer Press Only “B” to Press Neither. In other words, *Inter-State Trade-Offs* says that you should prefer pressing “A” irrespective of whether you press “B” and that you should prefer pressing “B” irrespective of whether you press “A”. Either route is sufficient for demonstrating the violation of Transitivity.

| | HEADS | TAILS |
|----------------|---------------------|-------|
| Press Only “A” | $A^{+\$1} \wedge B$ | $\$1$ |
| Press Neither | B | A |

You know the coin is fair, so $Cr(\text{HEADS}) = Cr(\text{TAILS})$. Moreover, because you’d rather have A *plus* a dollar than A *minus* a dollar, your preference for $A^{+\$1} \wedge B$ over B is stronger than your preference for A over $\$1$.¹³ (Or at least it is assuming, as we have been, that you value the two prizes *independently* of each other: that having the one does not add to or subtract from the value of having the other. If this assumption doesn’t hold, change the example by picking prizes for which it does.) Therefore, according to *Inter-State Trade-Offs*, you should prefer Press Only “A” to Press Neither.

For analogous reasons, you should also prefer Press Both to Press Only “A”.

| | HEADS | TAILS |
|----------------|---------------------|------------|
| Press Both | $A^{+\$2}$ | $B^{+\$2}$ |
| Press Only “A” | $A^{+\$1} \wedge B$ | $\$1$ |

Because you’d rather have B *plus* a dollar than B *minus* a dollar, your preference for $B^{+\$2}$ over $\$1$ is stronger than your preference for $A^{+\$1} \wedge B$ over $A^{+\$2}$.¹⁴ And so, again, according to *Inter-State Trade-Offs*, you should prefer Press Both to Press Only “A”.

¹³ Because you prefer $A^{+\$1}$ to A , according to every $u \in \mathcal{U}$, $u(A^{+\$1}) > u(A)$. Furthermore, $u(A^{+\$1}) + u(B) - u(B) = u(A^{+\$1})$. So, for every $u \in \mathcal{U}$, $u(A^{+\$1}) + u(B) - u(B) > u(A) > u(A) - u(\$1)$. Let’s say that you value two or more goods, X, Y etc., *independently* just in case, for all $u \in \mathcal{U}$, $u(X \wedge Y \wedge \dots) = u(X) + u(Y) + \dots$. Because, as we’ve been assuming, you value the prizes independently, $u(A^{+\$1} \wedge B) = u(A^{+\$1}) + u(B)$. And thus, for all $u \in \mathcal{U}$, $u(A^{+\$1} \wedge B) - u(B) > u(A) - u(\$1)$. In other words, your preference for $A^{+\$1} \wedge B$ over B is stronger than your preference for A over $\$1$.

¹⁴ Because you prefer $B^{+\$1}$ to B , according to every $u \in \mathcal{U}$, $u(B^{+\$1}) > u(B) > u(B) - u(\$1)$. Furthermore, $u(B^{+\$1}) = u(B^{+\$1}) + u(\$1) - u(\$1)$, and $u(B) - u(\$1) = u(A^{+\$1}) + u(B) - (u(A^{+\$1}) + u(\$1))$. And, because you value the prizes independently, $u(B^{+\$1}) + u(\$1) - u(\$1) = u(B^{+\$2}) - u(\$1)$ and $u(A^{+\$1}) + u(B) - (u(A^{+\$1}) + u(\$1)) = u(A^{+\$1} \wedge B) - u(A^{+\$2})$. And, thus, for all $u \in \mathcal{U}$, $u(B^{+\$2}) - u(\$1) > u(A^{+\$1} \wedge B) - u(A^{+\$2})$. Your preference for $B^{+\$2}$ over $\$1$ is stronger than your preference for $A^{+\$1} \wedge B$ over $A^{+\$2}$.

Because the *Never Worse Principle* says that you shouldn't prefer Press Both to Press Neither, Transitivity is violated. In the previous section, I argued that a full-fledged decision theory is plausible only if it satisfies *Inter-State Trade-Offs*. Therefore, any plausible decision theory that obeys the *Never Worse Principle* will violate Transitivity.

4 Money Pump Arguments

If your preferences are non-transitive, you're at risk of being turned into a *money pump*: you'll accept a series of trades that collectively leave you worse off.¹⁵ Imagine that you must outsource the execution of your decision to your trusted butler, Jeeves: you'll instruct Jeeves which combination of buttons, if any, you'd like pressed, and he'll carry out your wishes. Pushing buttons is work, though, so if you'd like Jeeves to press a button, you must pay him a small fee (twenty-five cents, let's say). You'd prefer Jeeves to press button "A" and not "B" to him pressing neither, so you pay Jeeves the twenty-five cents. But, because you prefer Jeeves to press both over pressing "A" alone, you pay Jeeves an additional twenty-five cents to push both. Before Jeeves leaves to carry out your wishes (but after pocketing the fifty cents you've paid him), he graciously offers to, instead, push neither button. You don't prefer for Jeeves to push both rather than neither, so it's rationally permissible for you to accept his gracious offer. But if you do, you've paid him fifty cents for nothing. Jeeves has turned you into a money pump.

This, at least offhand, seems bad. But there's a response. While it's true that you are moneypumpable in virtue of having non-transitive preferences (as the case above illustrates), you are only *weakly* moneypumpable: there are some decisions you ought to make, and a decision it's *permissible* (but not required) for you to make such that, if you make all of them, you're guaranteed to be worse off than you would be otherwise. You could avoid the money pump by turning down Jeeves' last offer, and that's something it would be rationally permissible for you to do. In other words, rationality wouldn't *compel* you to make yourself worse off; it merely fails at

¹⁵ The earliest version of this kind of pragmatic argument can be found in [Ramsey, 1928, p. 182], which attempts to show that one's subjective degrees of belief should conform to the probability axioms. It was further developed into an argument for the transitivity of preferences by Davidson et al. [1955].

protecting you from doing so. And insofar as you have incomplete preferences (no matter how we develop a decision theory under uncertainty), rationality will fail to offer this kind of protection anyway. It's permissible for you to trade in A^+ for B , and it's permissible for you to trade in B for A , but, by doing both, you make yourself worse off than you would be otherwise. If you're already vulnerable to being weakly moneypumped (in virtue of having incomplete preferences), the fact that decision theories obeying the *Never Worse Principle* fail to protect you from being weakly moneypumped is not a sufficient reason to reject them. Or so the response goes.

But is the response convincing? I'm not so sure. The argument that you're weakly moneypumpable if you have incomplete preferences implicitly presupposes a particular decision-rule:

It's rationally permissible to do something so long as there is no available option you prefer to it.

And, while that decision-rule is fairly attractive, it could be denied. In particular, you might think that what it's rationally permissible to do, when choosing between options you're ambivalent between, depends on what other choices you've made.¹⁶ It's permissible to trade B for A but not if you've just traded A^+ for B .

In order to see why this might be so, consider Chang [2009, 2017]'s view about the source of practical normativity: *Hybrid Voluntarism*. On her view, when your given reasons run out (e.g., your options are, in her words, "on a par"), "you have the normative power to *create* new will-based reasons for one option over another by putting your agency behind some feature of one of the options," ([Chang, 2017, p. 16]). In choosing B over A^+ , you focus on one of the distinctly valuable features of B and, by "putting your agency" behind that feature, create for yourself a decisive reason to choose B . If you are then offered to exchange B for A , it's no longer obvious that it's permissible to do so.¹⁷

¹⁶ In fact, this is how Chang [2005] responds to the money pump. She says, "If one chose B when offered a choice between A^+ and B , one is thereby rationally prohibited from choosing A when offered a choice between B and A . This is true even though there is a sense in which because B and A are on a par, it is rationally permissible to choose either. This is the sense in which if one had not already chosen B over A^+ , it would have been rationally permissible to choose A over B ." ([Chang, 2005, p. 347])

¹⁷ Although Chang [2005] holds that it's not permissible to choose A over B after having chosen B over

There's a story here about how the money pump can be avoided when choosing among options you're ambivalent between. But the analogous story cannot be told in Jeeves' case. Because you prefer Jeeves to push both buttons rather than just one and because you prefer Jeeves to push one of the buttons rather than none, your given reasons haven't run out. In making the first two choices, you are not choosing between options you're ambivalent between. And so, by choosing one option over the other, you are not willing any new reasons into existence which could constrain what it's rationally permissible for you to do in the final round. The money pump argument for Transitivity is more serious than the one for completeness.

Furthermore, there might be other reasons — aside from money pumps — to worry about non-transitive preferences. For example, you might think that there's a tight connection between *preferring* one thing to another and *judging* that the former is personally better than the latter; and that (perhaps as a matter of conceptual necessity) the relation *better than* is transitive.¹⁸ If so, non-transitive preferences beget inconsistent judgements. Or you might, like Davidson [1976, p. 273], think that transitivity is constitutive of what it *means* to prefer one thing to another. If your “preferences” aren't transitive, then you can't rightly be said to have them at all. Or you might find it simply *obvious*, and thus in need of no further justification, that one's preferences should be transitive.

5 Conclusion

Decision theories that obey the *Never Worse Principle* come with a cost that must be answered for: they saddle you with non-transitive preferences. One way to answer

A⁺, the discussion in [Chang, 2017] suggests that she might not be entirely happy with the explanation I've given for why. She doesn't think that by committing yourself to a particular course of action (in the manner described) you thereby normatively bind yourself in the future. Your commitments can be undone. Furthermore, she doesn't think that, when choosing between options on a par, you are *required* to resolve the choice by creating a new will-based reason. You are free to do so, but it's also permissible to *drift*: to intentionally choose an option without putting yourself behind some distinctly valuable feature of it. Drifters, so it seems, *are* vulnerable to being money pumped.

¹⁸ Broome [1991, 2004] is a proponent of this latter claim: he thinks it's a logical truth that *better than* is transitive. He thinks that all comparative relations (“more ... than”) are necessarily transitive. Others (e.g., [Temkin, 1996]) disagree. Broome [2006] is possibly a proponent of the former claim too. Hausman [2011] argues for transitivity along similar lines.

for it is to embrace it: accept that it's not always irrational to have non-transitive preferences. But, for some, this might be too high a price to pay, no matter how initially plausible the *Never Worse Principle* might've seemed. The argument of this paper, then, could be seen as providing some indirect support for views, like *Prospectism*, which say you're required to take the sweetened option in cases of Opaque Sweetening. There are no plausible full-fledged decision theories that say otherwise and obey Transitivity.

On the other hand, rejecting the *Never Worse Principle* also comes at a price. If an option never does worse than another, you're in a position to know that, no matter how the world turns out to be, you won't disprefer it. If you know that you won't disprefer it, why disprefer it now? Hare [2013, p. 50–52] offers an explanation. Being rational, he claims, involves being guided by the reasons you have. In cases like Vacation Boxes, you have a reason to prefer *L* to *R* (namely, that if you take *L* you'll get an extra two dollars, but if you take *R* you won't) and you have *no* reason to prefer *R* to *L* (anything that can be said in favor of the former can equally well be said in favor of the latter). The reasons you have, then, support preferring *L* over *R*. The *Never Worse Principle* fails because the fact that *R* never does worse than *L* doesn't provide you with a reason to prefer it. The principle might nevertheless *seem* true because, in many cases, the fact that an option never does worse than another *does* provide you with a reason to prefer it. Only when there's symmetry — when both options never do worse than the other — will these facts fail to be reasons. And so, only in special cases — cases, like Vacation Boxes, in which your reasons pull you in only one direction — will the *Never Worse Principle* lead you astray.

I don't think this is the final word on the matter, however. Hare's argument appeals to a particular way of thinking about rationality — that you should be guided by only those reasons you have — and a particular conception of reasons. Both of which can be questioned.¹⁹ And it's not obviously mistaken to think that, if you know you're ambivalent between the contents of the boxes, rationality shouldn't require you to prefer one to the other. My aim isn't to convince you otherwise. It's more modest. I've argued that there's a cost to accepting that thought: there's no

¹⁹ See the criticisms of [Bales et al., 2014] and [Schoenfield, 2014], for example. The argument is also discussed in [Doody, 2018].

way to develop it into an adequate full-fledged decision theory without violating Transitivity.

References

- Ralf Bader. Stochastic dominance and opaque sweetening. *Australasian Journal of Philosophy*, 96(3):498–507, 2017.
- A. Bales, D. Cohen, and Toby Handfield. Decision theory for agents with incomplete preferences. *Australasian Journal of Philosophy*, 92(3):453–470, 2014.
- John Broome. *Weighing Goods: Equality, Uncertainty and Time*. Basil Blackwell, 1991.
- John Broome. *Weighing Lives*. Oxford University Press, 2004.
- John Broome. Reasoning with preferences? *Royal Institute of Philosophy Supplements*, 59:183–208, 2006.
- John Broome. Does rationality consist in responding correctly to reasons? *Journal of Moral Philosophy*, 4(3):349–374, 2007.
- Ruth Chang. The possibility of parity. *Ethics*, 112(4):659–688, 2002.
- Ruth Chang. Parity, interval value, and choice. *Ethics*, 115(2):331–350, 2005.
- Ruth Chang. Voluntarist reasons and the sources of normativity. In David Sobel and Steven Wall, editors, *Reasons for Action*, pages 243–271. Cambridge University Press, Cambridge, 2009.
- Ruth Chang. Hard choices. *APA Journal of Philosophy*, 3(1):1–21, 2017.
- Donald Davidson. Hempel on explaining action. In *Essays on Actions and Events*, pages 261–275. Oxford University Press, 1976.
- Donald Davidson, John McKinsey, and Patrick Suppes. Outlines of a formal theory of value, i. *Philosophy of Science*, 22:140–160, 1955.

- Ryan Doody. Parity, prospects, and predominance. *Philosophical Studies*, <https://doi.org/10.1007/s11098-018-1048-0>, 2018.
- I.J. Good. Rational decisions. *Journal of the Royal Statistical Society, Series B (Methodological)*, 14(1):107–114, 1952.
- James Griffin. *Well-Being: Its Meaning, Measurement and Importance*. Clarendon Press, Oxford, 1986.
- Caspar Hare. Take the sugar. *Analysis*, 70(2):237–247, 2010.
- Caspar Hare. *The Limits of Kindness*. Oxford University Press, 2013.
- Caspar Hare. Should we wish well to all? *The Philosophical Review*, 125(4):451–472, 2016.
- Daniel Hausman. *Preference, Value, Choice, and Welfare*. Cambridge University Press, 2011.
- Isaac Levi. *Hard Choices*. Cambridge University Press, Cambridge, 1986.
- E. Ok, P. Ortoleva, and Gil Riella. Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica*, 80(4):1791–1808, 2012.
- Derek Parfit. *Reasons and Persons*. Clarendon Press, Oxford, 1984.
- Frank Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, chapter VII, pages 156–198. Kegan, Paul, Trench, Trubner and Co., London, 1926.
- Miriam Schoenfield. Decision making in the face of parity. *Philosophical Perspectives*, 28(1):263–277, 2014.
- Amartya Sen. Incompleteness and reasoned choice. *Synthese*, 140(1/2):43–59, May 2004.
- Larry Temkin. A continuum argument for intransitivity. *Philosophy and Public Affairs*, 25(3):175–210, 1996.

Paul Weirich. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford University Press, 2004.