# Decision theory for agents with incomplete preferences

Adam Bales, Daniel Cohen, and Toby Handfield

*9 September 2013*

**Abstract**

Orthodox decision theory gives no advice to agents who hold two goods to be incommensurate in value because such agents will have incomplete preferences. According to standard treatments, rationality requires complete preferences, so such agents are irrational. Experience shows, however, that incomplete preferences are ubiquitous in ordinary life. In this paper, we aim to do two things: (1) show that there is a good case for revising decision theory so as to allow it to apply non-vacuously to agents with incomplete preferences, and (2) to identify one substantive criterion that any such non-standard decision theory must obey. Our criterion, Competitiveness, is a weaker version of a dominance principle. Despite its modesty, Competitiveness is incompatible with prospectism, a recently developed decision theory for agents with incomplete preferences. We spend the final part of the paper showing why Competitiveness should be retained, and prospectism rejected.

**Keywords:** Decision theory, incommensurate value, practical reason, incomplete preferences, dominance.

Thanks to Wylie Breckenridge, Caspar Hare, Jakob Hohwy, Morgan Luck, Miriam Schoenfield, and Alastair Wilson for helpful discussions regarding this paper. We also benefited from sharing earlier versions of this paper with audiences at MIT, the University of Melbourne, and the Frontiers of Rationality Conference, held at the University of Groningen in 2012.

This is a preprint of an article whose final and definitive form will be published in the *Australasian Journal of Philosophy*; available online at http://www.tandf.co.uk/journals.

# 1 The demands of decision theory

There is a growing popular literature in psychology and behavioural economics examining the nature of choice.[1] In such literature, one often reads claims to the effect that humans do not conform to the axioms of decision theory, and thus that we are irrational. Consequently, it is claimed, our economic theories should be revised so as to better predict and explain the behaviour of genuinely human, irrational agents, rather than designed to predict behaviour of the fictional *homo economicus*.[2]

No doubt, humans are frequently irrational, and no doubt there are serious problems with the use of decision theory as a tool to predict our behaviour. But the line of thought glossed above is too quick to concede that decision theory provides an adequate normative account of rationality itself. Orthodox versions of decision theory, founded upon the idea of maximizing expected utility, are extremely demanding, or contain significant idealizations. For these sorts of reasons, we might well suspect that decision theory is not an adequate theory of rationality. There are a number of ways in which this complaint could be made, and in this paper we focus on only one of them: the demand that rational agents have a *complete and transitive* preference ordering over possible actions. This premiss is in tension with the thought that a rational agent might regard two states of affairs as *incommensurate* in value. Agents who hold states of affairs to be incommensurate will have incomplete preferences. That is, they will lack an all things considered preference for one state over the other, but they will also fail to hold the two states to be equally preferable. It has long been known that decision theory's premisses are implausible for this reason, but little has been done to address the concern.[3]

To briefly illustrate what it is to hold two things to be incommensurate in value, suppose you are confronted with a terrible dilemma. Both your mother and father are in mortal danger. Your mother is attending to her garden, located in a valley, and is endangered by an imminent flood. Your father is out gathering mushrooms in the foothills of nearby mountains, and is threatened by an impending avalanche. You have time to save one or the other, but not both. Whomever you do not save will surely die. You are very close to both your parents, both of whom are of outstanding character, and both are in good health. You have

[1] E.g. Ariely 2008; Brafman and Brafman 2009; Schwartz 2003; Thaler and Sunstein 2008.

[2] Ariely is particularly explicit in this regard, see Ariely 2008: xxix–xxx.

[3] E.g. Broome 1991: 92–3; Savage 1954: 21.

no preference that one rather than the other live. You know that you will feel the burden of this decision for the rest of your life.

But there is a complication that you have just remembered: the government has recently advertised that it will provide free grief counselling for the bereaved relatives of flood victims, but no such scheme exists for the bereaved of avalanche victims. Though you would never express such a callous-sounding preference out loud, if your mother died, you would prefer to receive the free grief counselling than not. But even so: this consideration is not enough to help you *knowingly to choose between* your mother and your father. The thought that, by saving your father, you will secure free counselling for your mother's death in no way convinces you that saving your father is the right thing to do. You would still have no preference to save one or the other.

This complication is important, because it suggests that you don't take your father's survival and your mother's survival to be *equally desirable*. To make this point clear, we need to give a more general characterisation of the value relations that two things, $A$ and $B$, can possibly stand in. The obvious possibilities are:

- $A > B$ ($A$ is better than $B$)
- $B > A$ ($B$ is better than $A$)
- $A \equiv B$ ($A$ and $B$ are equal in value)

The relation of being equal in value can be defined in terms of the better than relation.[4] Two things, $A$ and $B$ are equal in value if, and only if:

> it is not the case that $A > B$; it is not the case that $B > A$; and for all things $C$, ($C > A$ iff $C > B$), and ($A > C$ iff $B > C$).

Is it possible for two things to stand in any other value relation than these? Certainly: two things may be *incommensurate in value* if none of the other comparative value relations hold between them. The above example appears to be one such case. You do not value saving your father more than saving your mother (Save Father $\not>$ Save Mother). You do not value saving your mother more than saving your father (Save Mother $\not>$ Save Father). But do you value these things equally? If so, then a very small improvement (a mild 'sweetening') in one option should render it superior to the other. Saving your father, while receiving grief

[4] Here we follow, and revise slightly, Broome 2004: 21.

counselling for the death of your mother (call this 'Save Father+'), *is* a better outcome than saving your father, without receiving grief counselling (Save Father+ > Save Father). So if the two original options were equally good, we would find that the grief counselling is a 'tie-breaker'. We would be able to reason as follows:

$$\begin{aligned} \text{Save Father} &\equiv \text{Save Mother} \\ \text{Save Father+} &> \text{Save Father} \\ \therefore \text{Save Father+} &> \text{Save Mother} \end{aligned}$$

That is, from the assumption that you value the initial alternatives equally, it follows that you are rationally required to save your father, in order to obtain the free grief counselling. This is – for at least some agents – an implausible conclusion. Rather, it appears that we have good reason to reject the first premiss: Saving your father and saving your mother are not equal in value, but are *incommensurate*.[5]

Standard decision theory uses the preferences of individuals to determine the relevant value of outcomes. An assumption of standard decision theory is that rational agents have preferences over all outcomes that stand in a complete and transitive total ordering. But agents who take two things to be of incommensurate value violate these conditions. More precisely, agents who regard two states as incommensurate in value have incomplete preferences. For such agents, for some three outcomes $x, y, z$, the agent has no preference between $x$ and $y$, has no preference between $y$ and $z$, and yet prefers $x$ to $z$. In this case: $x$ = Save Father+; $y$ = Save Mother; $z$ = Save Father.

When an agent has incomplete preferences, standard decision theory is silent – it gives no advice. This is because, to handle decisions under uncertainty, decision theory uses expected utility calculations. Constructing a utility function on which to base such calculations requires a complete preference ordering. Incomplete preferences block construction of the utility function, and thereby prevent us from deriving any recommendations from the theory. (Without going into the technicalities of how utility functions are constructed (e.g. Debreu 1954), it is easy enough to see why a utility function is incompatible with incommensurate value. A utility function assigns to every outcome a real number. The real

[5] Some prefer to use the term 'incomparability' to describe this value relation. See, e.g. Hsieh 2008. Ruth Chang (2002) claims that, in addition to the possibility of incomparability – understood as the absence of any comparative relation – there is also the possibility of the relation of 'parity', which is a fourth comparative relation that might be described as 'rough equality'. In this paper, we ignore the alleged distinction between parity and incomparability, and assume that they can be treated identically.

numbers stand in a total ordering – there are no incommensurabilities between real numbers! Consequently, utilities cannot represent incommensurabilities.)

To illustrate the plausibility of there being incommensurate value, we have appealed to a decision that involved a very weighty, terrible decision. But in trivial choices, also, it seems that we can have incomplete preferences without being irrational. Shall we go bowling or have a picnic? It's a toss-up – they both have their attractions. Now you discover that the tram fare to the picnic has increased by ten cents. Paying ten cents extra for a picnic is worse than not, though only marginally. But does this constitute a tie-breaker? Are you rationally required to go bowling? For some agents, it is plausible that they hold the attitudes that the cheaper picnic is preferable to the dearer one, but they have no preference between the picnic and bowling, whether or not the picnic is 10 cents more or less costly. This does not strike us as strong grounds to deem such agents irrational.[6]

By appeal to the plausibility of such examples as this, we believe we can mount an argument that decision theory ought to be reformed. It can be rational to hold two goods to be incommensurate in value. Incommensurability would require us to have incomplete preferences. Therefore it can be rational to have incomplete preferences.

We also endorse a second argument, which relies less upon intuitions about incommensurate value, and instead on a plausible claim about our psychology: viz, that we are in some sense finite agents. For instance, we can only perform a finite number of computations in our lifetimes; and we can only hold a finite number of things in our mind at the one time. There are infinitely many possible outcomes of our actions, however, and so the requirement of completeness appears to require that we have preferences over infinitely many states of affairs. Maintaining coherence between all these preferences appears to require enormous – perhaps infinite – computational power. Thus decision theory demands the impossible of us. Because decision theory is supposed to have normative force for finite agents, and ought implies can, decision theory must be reformed.

Much more could be done to elaborate and defend this second argument, but we don't expect that doing so will persuade the advocate of orthodox decision theory.[7] Rather, more

[6] A terminological note: one might want to refer to an agent who lacks a preference between choices in trivial cases, such as the example above, as 'indifferent'. Standardly, however, decision theorists use the term 'indifference' to refer to the state of holding two things to be equally preferable. As we hope the above cases illustrate, these two varieties of 'indifference' are crucially different, and we will – with one exception – avoid the term to minimize confusion.

[7] For those seeking further persuasion, Gigerenzer, Todd, and the ABC Research Group (1999: 76) argue that

is likely to be achieved by attempting to establish that an alternative decision theory can be developed, and thereby showing what advantages and disadvantages accrue to each method of formally stating the normative requirements of decision-making. So we propose to set aside the explicit advocacy of non-standard decision theory, and instead spend the rest of this paper attempting to show what an adequate decision theory might look like, for agents with incomplete preferences.

## 2 Towards a decision theory for agents with incomplete preferences

There are two principal problems that will confront anyone attempting to devise a theory of decision for agents with incomplete preferences.

**Decision under uncertainty.** For any given action that an agent may perform, the agent is typically uncertain what the outcome will be. Given this, we cannot simply say that it is rational to act in the way that will bring about the best outcome, for agents typically cannot know which action that is. The standard way to explain rational decision-making under uncertainty is to use *expected utility*. We associate with each possible outcome a cardinal measure of its desirability: its utility. Also associated with each outcome is the agent's subjective probability that the outcome will occur, conditional on having performed a given action. Rational action then, involves taking the action whose outcomes have the greatest probability-weighted utility.

As mentioned above, however, in order to ascribe to an agent a utility function, standard techniques require a complete preference ranking. Without a utility function, we cannot associate with each decision an expected utility. So if we hope to give an adequate theory of rational choice under uncertainty, we will need to find some novel way of ranking available actions.

There are some decision rules which do not require *probabilities* in order to be applied. Maximin, for instance, advises the agent to take the action whose worst outcome is best, relative to the worst outcome of all other actions. This extremely pessimistic decision rule

Bayesian conditionalisation is computationally intractable, and therefore not possible for humans. Simon (1955: 101) suggests a similar connection as we do between human cognitive limitations and the normativity of rational choice theory.

is mirrored by maximax, which decides based upon the most optimistic assessment of all available actions. Both rules are absurd in their extremism. Moreover, for many realistic decisions, it is plausible that the worst possible outcome (or best possible outcome) is *identical* for all available actions. Whatever your greatest fear (or wish), it is *possible* that it will come about, no matter what you do. This suggests that maximin and maximax will end up saying that all available actions are permissible. Hardly a satisfying result.

Consequently, it would be very desirable to develop a method that retains some of the virtues of expected utility theory, but which does not require utilities.

**Serial decision making** The second problem that we will encounter concerns diachronic rationality – achieving rationally acceptable outcomes over a series of decisions. This problem arises even in cases of decision under certainty.

Suppose an agent has the following preferences:

$A \not\sim B$ (no preference)

$A \not\sim B^+$

$B^+ > B$ (strict preference)

It is very plausible to think that, in decisions under certainty, an agent is rationally required to select a maximal option. That is: the agent must choose an option that is not worse than any other option. Suppose there is no other rule governing rational choice. The following series of choices, then, is permissible. Suppose the agent starts with $A$, and is offered the option to swap to $B$:

1. $\{A, B\} \mapsto B$
2. $\{B, B^+\} \mapsto B^+$
3. $\{A, B^+\} \mapsto A$
4. $\{A, B\} \mapsto B$
5. …

The agent starts by choosing $B$, trades up to $B^+$, swaps to $A$, and then swaps back to $B$. Note that on the second trade, the agent is rationally *required* to accept $B^+$ rather than $B$. If we make some mild assumptions about the psychology of our agent, we can suppose that the agent would also be willing to pay a very small sum for this trade – say ten cents.[8] But now, if we repeat the cycle of trades, including this small payment, the agent will keep paying out more and more money, while merely cycling through the options *B*, $B^+$, *A*.

In short, agents with incomplete preferences are vulnerable to money-pumps. That said, diachronic rationality is a challenge for many theories of rational choice, and it might be that the sorts of measures proposed to assist agents with complete preferences to avoid money-pumps can be adapted to agents with incomplete preferences.[9] Partly for this reason, and partly because it is a large topic in its own right, we propose to ignore the second problem in this paper.

Finally, our project is modest in a further respect: we do not go so far as to establish a particular decision theory that we believe will satisfactorily address the problem of decision under uncertainty. Rather: we try to establish a criterion that any adequate decision theory must satisfy. Our criterion is a close cousin of a dominance principle, and we call it Competitiveness.

## 3 Competitiveness

One of the least controversial principles of rational choice is Dominance. Here is how it is introduced in a classic paper by Nozick (1969).

> *Dominance Principle:* If there is a partition of states of the world such that relative to it, action alpha weakly dominates action beta, then alpha should be performed rather than beta.
>
> Action alpha weakly dominates action beta for person *P* iff, for each state of the world, *P* either prefers the consequence of alpha to the consequence of beta, or

[8] See Mandler 2005: 261 for formal characterisation of the necessary assumptions.

[9] In particular, it has been suggested that agents who experience preference reversals will need to be capable of making and committing to plans in order to avoid money pumps. Perhaps such resoluteness will also allow agents with incomplete preferences to avoid money pumps. See McClennen (1990) or Bratman (1998) for a discussion of such views of rational agency. Alternatively, perhaps backward-looking agents that consider their past decisions when making their current ones will be able to avoid money pumps. See Mandler (2005) for a discussion.

A referee has pointed out to us that a parallel money-pump concern has been raised for defenders of imprecise *credences* (Elga 2010; White 2010). To that end, the defender of incomplete preferences may be able to take inspiration from James Joyce's (2010) responses to those objections.

> is indifferent between the two consequences, and for some state of the world, *P* prefers the consequence of alpha to the consequence of beta.

As Nozick goes on to show, this principle is too permissive. There are some partitions of states of the world such that a manifestly irrrational action will be required by the Dominance Principle. Suppose you hear an air-raid siren. You can shelter or ignore it. Fatalistically, you think: either I'm going to die in the raid, or I won't. Whichever way things turn out, I would prefer to spend less time in an uncomfortable shelter, so I will ignore the warning. This reasoning appeals to dominance, but ignores the fact that the probability of dying is much lower if you shelter. There is a probabilistic dependency between the action you choose and the state of the world, at least on the death–survival partition.

For many decision problems, however, it is possible to identify a partition such that the probability of a given world-state is independent of what action you take. For those decision problems, relative to those partitions, the Dominance Principle above gives seemingly impeccable advice. Henceforth, we will be discussing only cases where this probabilistic independence obtains, so will omit the explicit relativisation to a partition.

Suppose an agent faces a decision problem with the structure illustrated in Table 1, where $A^+ > A$, but the agent holds both $A^+$ and $A$ to be incommensurate with $B$. According to Nozick's Dominance principle, an action alpha is rationally obligatory if, for every way the world could be, the consequences of alpha are either better than or equal in value to the consequences of all alternative actions, and for some way the world could be the consequences of alpha are better than the consequences of all alternative actions. Because, in State Y, the outcome of the two possible actions is incommensurate, Dominance is silent in this case.

| | State X | State Y |
|---|---|---|
| Option 1 | A+ | A+ |
| Option 2 | A | B |

Table 1: A scenario in which it is tempting to apply Dominance-like reasoning, but strictly speaking Dominance is silent.

Now consider a different principle, which we take to be at least as plausible as Nozick's original: an action alpha is rationally obligatory if, for every way the world could be, the consequences of alpha are *not worse* than the consequences of all alternative actions, and for some way the world could be the consequences of alpha are better than the consequences of all alternative actions. We suggest that our intuition that the first option in Table 1 is rationally obligatory is best explained by this principle.

Now this principle appears to build on a simpler, and even more compelling, principle: that an action is rationally permissible if, for every way the world could be, its consequences are no worse than the consequences of all alternative actions. In such a case, let's say that the action is 'competitive' and call the principle that competitive actions are rationally permissible 'Competitiveness'.

> *Competitiveness:* If an action alpha is competitive, it is rationally permissible to perform alpha.

Our variant on Nozick's dominance principle can now be reformulated as follows:

> *Strong Competitiveness:* If one or more actions are competitive, and other actions are not competitive, it is rationally required to perform a competitive action.

Strong Competitiveness explains why option 1 is the only permissible action in the above example: it is the only competitive action. But for dialectical purposes in this paper, it will suffice to defend the weaker Competitiveness principle. Any adequate decision theory for agents with incomplete preferences will respect Competitiveness, we claim. It might be thought that this is too modest a claim to be of interest. But one of the more promising recent proposals to reform decision theory, intended precisely to give advice to agents with incomplete preferences, violates this requirement.

## 4 Hare's prospectism

Caspar Hare has recently developed two novel extensions to decision theory that promise to give advice to agents with incomplete preferences (Hare 2010). Hare argues that both his proposals have genuine appeal, and he leans towards one called prospectism.[10]

The key idea behind prospectism is to make up for the difficulty caused by incomplete preferences by appealing to *possible coherent completions* of the preferences with which the

[10] Prospectism is not entirely new. An extremely similar idea has been defended by Paul Weirich (2004: 70), and Weirich cites a predecessor of the idea due to I. J. Good (1952). We take our criticisms of Hare's proposal to generalize to Weirich's account also.

agent begins. For our agent with the preferences: $B \not\sim A^{+} > A \not\sim B$, the following are some coherent completions:

- $B \equiv A^{+} > A$
- $A^{+} > A > B$
- $A^{+} > A \equiv B$
- $A^{+} > B > A$
- …

Using these possible completions of an agent's preferences, we can define utility functions. Hare's prospectism says that an action is rationally permissible if (and only if), according to *some* coherent completion of the agent's preferences, that action has maximal expected utility (Hare 2010: 242–3).

Note that, in all of these coherent completions, we have to respect the original preference for $A^{+}$ over $A$, so $A$ will never have a higher utility than $A^{+}$. In some completions, however, $A^{+}$ will have maximal utility, and in others, $B$ will. Consequently, applying prospectism to a choice under certainty between the above three outcomes, it will be permissible to choose $A^{+}$ or $B$, but impermissible to choose $A$, because there is no completion of preferences according to which $A$'s utility is maximal.

## 4.1 Prospectism and decision under uncertainty

Consider the following case of decision under uncertainty, given by Hare.

> Suppose I lack preferences between my getting item A and my getting item B. Suppose this attitude is insensitive to mild sweetening [hence my preferences manifest incompleteness]. And suppose we play a kind of game:
>
> > *Two opaque boxes*
> > You show me items A and B, a dollar, a coin, and two opaque boxes. You toss the coin and, governed by the toss, place item A in one box and item B in the other. I don't see which item went where. You toss the coin again and, governed by the toss, place the dollar inside the right box. I see that – which leaves me with credence 0.5 that things are like so:

> | **Left box** | **Right box** |
> |---|---|
> | A | B + \$1 |
>
> and credence 0.5 that things are like so:
>
> | **Left box** | **Right box** |
> |---|---|
> | B | A + \$1 |
>
> Then you invite me to walk away with one of the boxes.
>
> (Hare 2010: 239–40)

It is obvious that we may take the right box. The important question is: *must* you take the right box? Is it rationally permissible to take the left box? Because your preferences are incomplete, standard decision theory is silent on the matter. Hare claims that there are prima facie 'powerful' arguments both for and against the rational permissibility of taking the left box. In favour of it being permissible to take the left box, Hare argues, first: if I were fully informed – if, e.g. I could look inside the boxes – I would have no preference between them, and I would think it rationally permissible to take the left box. Knowing that my fully informed, rational self would have this view, it must be at least permissible to defer to my more informed self.[11] Second, I know that I have no preference for the contents of either box. Where I have no preference between two choices, surely I am rationally permitted to take either.

Drawing on considerations like these, Hare develops a version of decision theory which he calls deferentialism, and which delivers the verdict, for cases like this, that it is indeed permissible to take either box. We find many aspects of deferentialism intuitively appealing, but do not discuss it further in this paper.[12]

[11] This sort of reasoning has clear echoes of the Reflection principle – though it seems less vulnerable to the sorts of problem cases that have been raised against Reflection, given those cases tend to exploit future states in which the agent is not rational (Christensen 1991), or is in receipt of very special information that is inaccessible to others (Elga 2000). Neither of those sorts of concerns seems relevant here.

A further, indirect, consideration in favour of deferring to the preferences of my future self is that this idea features heavily in Krister Bykvist's (2006) elegant account of how to make prudent decisions in anticipation that one's future preferences will depend upon one's present choices.

[12] Although deferentialism respects Competitiveness, one reservation we have is that it violates Strong Competitiveness. Discussing why, and how this feature might best be redressed, would take us too far afield, however.

For the contrary view, that it is *impermissible* to take the left box, Hare offers two further arguments. First, 'there is a consideration of which I am aware that counts in favour of my taking the right, rather than the left box: I will get a dollar if I take the right box, no dollar if I take the left box. But there is no consideration of which I am aware that counts in favour of my taking the left, rather than the right box' (240).

Hare's second argument for this conclusion is more complicated, and appeals to an analogous case, which we simplify as follows:

> *One opaque box*
> Based on the toss of a coin, you put either item A or B in a box, without showing me which. You then offer me the box and a dollar, or just the box without the dollar.

In this game, it is rationally required that I take the box plus the dollar. But in its decision-theoretic structure, this game is extremely similar to Two Opaque Boxes. The decision tables of the games are given in Tables 2 and 3.

| | A is in right | B is in right |
|---|---:|---:|
| Take right | A+ | B+ |
| Take left | B | A |

Table 2: Decision Table for Two Opaque Boxes

| | A is in box | B is in box |
|---|---:|---:|
| Take box + dollar | A+ | B+ |
| Take box | A | B |

Table 3: Decision Table for One Opaque Box

Take the *prospect* associated with an action to be the possible outcomes that might come about if I take that action, paired with my credences in those outcomes.[13] In One Opaque Box,

- The prospect of taking the box and the dollar is: $\{\langle A^+, 0.5\rangle, \langle B^+, 0.5\rangle\}$.

[13] Hare allows that the relevant credences will be different, depending upon whether one favours causal decision theory or evidential decision theory. Thus his defence of prospectism remains neutral on that dispute. For all the examples discussed in this paper, the dispute between evidentialists and causalists is irrelevant.

- The prospect of taking the box alone is: $\{\langle A, 0.5\rangle, \langle B, 0.5\rangle\}$.

The evaluation of an action, in standard decision theory, is a function of the prospect of that action. The utilities for the various outcomes are multiplied by the probability of those outcomes to give an expected utility associated with that prospect. In this case, there is no utility function that can be assigned to the four possible outcomes because of the incompleteness of my preferences. Nonetheless, even in cases like these, we might think that the prospects of one's alternatives determine what one may and may not do. Hare articulates this claim in a principle:

> *Prospects Determine Permissibility*: Facts about what it is rationally permissible for me to do are determined by facts about the prospects associated with the options available to me. (p. 240)

If we accept this principle, however, then we can argue that we must take the right box in Two Opaque Boxes, as follows:

(1) Taking the box and the dollar is rationally required in One Opaque Box.

(2) The prospects of this game are identical to the prospects of Two Opaque Boxes. (And the prospect of taking the right box, in particular, is identical to the prospect of taking the box and the dollar.)

(3) Prospects Determine Permissibility.

Therefore:

(4) Taking the right box is rationally required in Two Opaque Boxes.

Drawing inspiration from this argument, Hare develops the theory we have explained above: prospectism. Prospectism entails the principle Prospects Determine Permissibility.

Returning to Two Opaque Boxes, it is easy to see why prospectism entails that it is rationally required to take the right box. In all coherent completions of your preferences, the utility of A+ will be greater than the utility of A, and the utility of B+ will be greater than B. The expected utility of each action is given by the following equations.

$$EU(\text{Right}) = 0.5(A^{+}) + 0.5(B^{+})$$

$$EU(\text{Left}) = 0.5(A) + 0.5(B)$$

Accordingly, the expected utility of taking the right box must be greater than that of taking the left box, for all possible completions. So prospectism entails that it is impermissible to take the left box.

## 4.2 Dominance reasoning with incommensurate values

Given the strong structural similarity between Two Opaque Boxes and One Opaque Box, why do we claim that what is rationally permissible is different? Consider the following sort of rationale that an agent could employ.

> In the case of the Two Opaque Boxes, either item A or item B is in the right box. If item A is in the right box, my two possible choices lead to the outcomes: A+ and B. I have no preference between these. If item B is in the right box, my two possible choices lead to the outcomes: B+ and A. I have no preference between these. So however the world turns out, I have no preference between the outcome of either action.
>
> In the case of One Opaque Box, the situation is different. If item A is in the box, then there are two possible outcomes (A+ and A) and I *prefer* one of these. The preferred outcome will come about if I take the box and the dollar. If item B is in the box, then again there are two possible outcomes (B+ and B) and I *prefer* one of these. The preferred outcome will come about if I take the box and the dollar. So *however the world turns out*, I prefer the outcome of taking the box and dollar over taking the box alone.

In the second case, the sort of reasoning being used is clearly *dominance* reasoning, as already explained above.

In Two Opaque boxes, dominance reasoning – strictly understood – gives no advice because, for each way the world might be, it is not the case that one option is at least as good as the other. However, as we also argued above, we can weaken the notion of dominance to give us the principle of Competitiveness. In the case of Two Opaque Boxes, both taking the right box and taking the left box are competitive. Competitiveness entails, then, that taking either box is permisible.[14]

[14] The idea of Competitiveness is essentially the thought captured by Hare's principle of 'Recognition' (p. 241), though Hare does not identify how closely it follows the idea of dominance reasoning, extended into the domain of incomplete preferences.

So prospectism entails that Competitiveness is false. We think this is a considerable mark against it. Moreover, prospectism implausibly entails that the two decisions in One Opaque Box and Two Opaque Boxes are, for the purposes of decision theory, identical! That is, although the decision matrices of these decisions differ, they do not differ in their prospects. Consequently, prospectism implies that agents who treat these decisions differently are irrational. Of course, *some* rational agents might think these two decisions are identical insofar as rational choice is concerned, but it is very plausible to think that it is rationally permissible, for an agent with preferences like those described, to treat the decisions differently.

The committed prospectivist, however, may insist that we are begging the question. If prospectism is false, then what is wrong with the apparently tempting arguments that Hare presents for the view?

## 4.3 The considerations argument

The first of Hare's arguments, in full, is as follows:

> There is a consideration of which I am aware that counts in favour of my taking the right, rather than the left box: I will get a dollar if I take the right box, no dollar if I take the left box. But there is no consideration of which I am aware that counts in favour of my taking the left, rather than the right box. So, it is rationally impermissible for me to take the left box. It is rationally impermissible to do something when I have no reason to do it, and a reason to do something else. (Hare 2010: 240)

Setting this out even more explicitly, Hare seems to have the following argument in mind:

In a different context, Amartya Sen (1997; 2000) has argued that consequentialists do better to adopt a 'maximizing' form of consequentialism, rather than an 'optimizing' form. The maximizing consequentialist merely seeks to bring about outcomes that are *no worse* than any alternative. The optimizing consequentialist seeks to bring about an outcome that is *at least as good* as all alternatives. If there are incommensurate goods, or there is some other failure of the assumptions of completeness and transitivity in the betterness relation, then the optimizing goal may be impossible – there may be no outcome that is at least as good as all others – but the maximizing goal remains viable. The adoption of a rule permitting 'competitive' actions as opposed to a rule permitting only dominant actions complements, in decision theory, Sen's proposal for ethics more generally.

(1) If you have a reason to make decision X rather than decision Y and no reason to make decision Y rather than decision X then you ought to make decision X.

(2) In Two Opaque Boxes you have a reason to take the right box rather than the left box (i.e. your getting a dollar).

(3) In Two Opaque Boxes you do not have a reason to take the left box rather than the right box.

Therefore,

(4) in Two Opaque Boxes you ought to take the right box.

Premiss (1) does look very appealing. We will call it the *Reasons principle*, and will not dispute it. Premises (2) and (3) rely upon the notion of having a reason to do one action rather than another. What exactly does it mean to say, in general, that something is a reason in favour of one action, over another? We claim that, when this concept is examined carefully, premiss (2) turns out to be unsupportable.

As a preliminary, let us say that *p* is a reason in favour of taking action A over action B just in case (*i*) for some way the world might be, *p* will be true if I take A, but false if I take B, and (*ii*) *p*'s being true is in some yet to be defined way better than *p*'s being false.

Plausibly, *p*'s being true is better than *p*'s being false, in the relevant sense, if and only if I prefer *p* being true to *p* being false. However, there are at least two ways of cashing out this preference:

One obvious hypothesis is that I prefer *p* being true to *p* being false if and only if I prefer any world in which *p* is the case to any world in which *p* is not the case. But on that reading, premise 2 is false. I do not prefer that I have an additional dollar and A rather than B alone, nor do I prefer that I have an additional dollar and B rather than A alone. Thus, either way, it would not be better if I had an additional dollar, and the additional dollar does not thus provide me with a reason to take the right box rather than the left box.

A second hypothesis is that I prefer *p* being true to *p* being false if and only if, *other things being equal*, I prefer *p* being the case to *p* not being the case. That does seem to correctly describe the phenomenon in Two Opaque Boxes. Other things being equal, I do prefer it to be the case that I have an additional dollar than otherwise. But this account of preference is so weak that it seems inappropriate to use it in this dialectical context, where we are trying to use it in conjunction with Hare's Reasons principle to derive a conclusion about what we ought to do. On this understanding of preference, it follows that I have a reason to take the right box because, other things being equal, I prefer to have an additional dollar rather than not. But it is also the case that *I know that other things are not equal*. Indeed, I know that I will only obtain the additional dollar at the cost of forgoing a good of great value to me. So to think that preferences, in this sense, can generate 'reasons' that can reliably be plugged into the Reasons principle is implausible – or at best question-begging.

In general, it is worth reflecting that the reason why decision theory adopts *maximal* states of affairs as the objects of evaluation is precisely so as to avoid complicated interaction-effects in the ranking of outcomes. By appealing to the thought that a sub-maximal state of affairs – your getting a dollar – is a reason, Hare is undermining this sensible methodological move.

To illustrate the difficulties that can arise if we follow Hare in his use of sub-maximal states of affairs to constitute reasons in this way, consider the following case.

> *Two Charitable Boxes:*
>
> Your house has recently burnt down, containing all your possessions. A charity has come to your rescue – offering you a $2000 gift voucher for one of two clothes stores. The first, store A, sells suits and the second, store B, sells casual clothing. You find these two options incommensurate. After all, your work doesn't require that you wear a suit but you think that if you wore one to the monthly management meeting it might increase your chances of getting a promotion. On the other hand, if you get the casual store voucher then you can be comfortable at home and don't have to go to the beach in a suit. Rather than having you simply choose one or the other of the vouchers, the charity has placed them in two boxes according to a coin toss that you didn't see. The boxes have been filled as follows:

| | Heads | Tails |
|---|---|---|
| Right box | Store A | Store B |
| Left box | Store B | Store A |

The charity now tells you two extra things: First, they have added a dollar (D) to the right box. Second, if the coin toss came up heads they added a suit (S) to each box. Your preferences are as follows:

$$(A + S + D) > (A + S) > (A + D) > A$$

$$(B + S + D) > (B + S) > (B + D) > B$$

You find all outcomes involving A incommensurable with respect to outcomes involving B, *except* you very strongly prefer (B + S) to any outcome involving A. After all, this way you can be comfortable most of the time but still wow your bosses at the monthly meeting so this outcome is more desirable than any other. So the decision table is now:

| | Heads | Tails |
|---|---|---|
| Right box | A + S + D | B + D |
| Left box | B + S | A |

It seems that in this case you should take the left box. After all, the left box gives you have a 50% chance of an outcome that you vastly prefer to any other outcome.

However, if you allow yourself to consider the reasons as decomposed into: a chance of getting A; a chance of getting B; a chance of getting S; and the certainty of getting D, then all the reasons to take the left box (chances of A, B, and S) are also reasons to take the right box. But there is one reason – the certainty of the dollar – for taking the right box rather than the left box. Hare's Reasons principle implies that it is rationally impermissible to take the left box. But taking the left box is the only rationally permissible decision and so we have been led astray. Thus, we must reject either Hare's Reasons principle or the assumption that the reasons, in Two Charitable Boxes, may be decomposed into chances of getting A, B, S and D. We recommend the second option.

Similarly, we should reject the assumption that one's reasons, in Two Opaque Boxes, are decomposable into A, B, and a dollar. Rather, we only have reason to think one's reasons can reliably be analysed in terms of the four possible outcomes in the case: A, A+, B, and B+.

## 4.4 The argument from analogy

The second argument Hare gives relies upon the analogy between Two Opaque Boxes and One Opaque Box. In the two cases, the decisions are very similar. In particular, the corresponding prospects are identical. In One Opaque Box it is rationally required that we take the box and the dollar. If Prospects Determine Permissibility is true, then it is rationally required to take the right hand box in Two Opaque Boxes.

As we have indicated above, we are dubious about Prospects Determine Permissibility. We claim that that premiss is false. We have a further complaint, however, which is that the intuition regarding One Opaque Box does not straightforwardly support prospectism. Prospectism requires that we 'take the sugar' *only where our credences are perfectly balanced* between the relevant alternatives. But the intuitive appeal of taking the box and the dollar in One Opaque Box is surely driven by dominance considerations. As such, the intuition is insensitive to the credences we have in the different possibilities. Consider a variant on One Opaque Box where we are told that the contents of the box was determined by the roll of a die. If the die landed 6, A was placed inside the box. Otherwise, B was placed inside the box. Now your choice is whether to take the box alone, or the box with a dollar.

Obviously, it is still rationally required that you take the box with the dollar. One might have thought, then, that prospectism would entail that you must take the right box in a similarly modified version of Two Opaque Boxes, where if the die landed six, A was placed in the right hand box, and otherwise B was placed in the right hand box. But this is not so. The prospects of these two games are not identical in their unsweetened actions:

- One opaque box, using die, take box alone: $\{\langle A, \frac{1}{6}\rangle, \langle B, \frac{5}{6}\rangle\}$.
- Two opaque boxes, using die, take left: $\{\langle A, \frac{5}{6}\rangle, \langle B, \frac{1}{6}\rangle\}$.

Consequently, prospectism allows that it is permissible to take either the right or the left box. This is a surprising mismatch between the intuitive justification of prospectism and the details of the theory.[15]

We conclude that prospectism is an implausibly radical departure from intuition, and Hare's affection for it is insufficiently motivated, given it violates a natural extension of

[15] For those who lack our anti-prospectivist intuitions, but would like to take our advice in distancing themselves from Hare's highly sensitive form of prospectism, the following, non-probabilistic principles might seem to be a better way to formalise the intuitive idea behind the second argument:

P1. If none of the possible outcomes associated with a given action $A_1$ are worse than any of the possible outcomes associated with any alternative action $A_i$, then $A_1$ is rationally permissible.

P2. If one of the possible outcomes associated with a given action $A_1$ is worse than one of the possible outcomes associated with an alternative action $A_i$, and no possible outcomes associated with $A_1$ are better than the possible outcomes associated with $A_i$, then $A_1$ is rationally prohibited.

These principles would entail that taking the right box is rationally required in variants of Two Opaque Boxes where the credences are not equal. However, these principles would also implausibly entail that no decision is rationally permissible in some cases (as there will be scenarios where P2 will imply that all decisions are rationally prohibited). As such, this alternative theory is also unsatisfactory.

dominance reasoning.

## 5 Conclusion

We conclude that a satisfactory decision theory will give advice to agents with incomplete preferences. Such a theory should satisfy Competitiveness. This rules out proposals such as prospectism. There is also significant tension between Competitiveness and any theory that accepts that two decisions are equally desirable if they have the same prospects.[16] While Competitiveness is modest, then, it nevertheless rules out a broad class of possible theories.

Having met the demands of Competitiveness, there is still more to do. A satisfactory decision theory for agents with incomplete preferences must also address the problem of serial decision making so as to avoid money-pumps. Whether such a theory can be given remains to be seen.

*Monash University* (Bales, Handfield)
*Charles Sturt University* (Cohen)

## References

Ariely, Dan. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. Harper Collins.

Brafman, Ori, and Rom Brafman. 2009. *Sway: The irresistible pull of irrational behaviour*. Broadway Books.

Bratman, Michael E. 1998. "Toxin, Temptation, and the Stability of Intention". In *Rational Commitment and Social Justice*, edited by Jules L. Coleman and Christopher W. Morris. Cambridge: Cambridge University Press.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

———. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Bykvist, Krister. 2006. "Prudence for Changing Selves". *Utilitas* 18: 264–283.

Chang, Ruth. 2002. "The Possibility of Parity". *Ethics* 112: 659–88.

[16] Using terminology due to Larry Temkin (2012: pp. 238–44), we can say that such theories satisfy the first principle of equivalence. The impossibility proofs that Temkin considers following his introduction of this principle could be used, with slight changes, to illustrate the tension between Competitiveness and the first principle of equivalence.

Christensen, David. 1991. “Clever Bookies and Coherent Beliefs”. *The Philosophical Review* 100: 229–47.

Debreu, Gérard. 1954. “Representation of a preference ordering by a numerical function”. In *Decision Processes*, edited by R. M. Thrall, C. H. Coombs, and R. L. Davis. New York: Wiley, 159–165.

Elga, A. 2000. “Self-locating belief and the Sleeping Beauty problem”. *Analysis* 60: 143.

Elga, Adam. 2010. “Subjective Probabilities Should Be Sharp”. *Philosophers' Imprint* 10.

Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Good, I. J. 1952. “Rational Decisions”. *Journal of the Royal Statistical Society* 14: 107–14.

Hare, Caspar. 2010. “Take the sugar”. *Analysis* 70: 237–47.

Hsieh, Nien-hê. 2008. “Incommensurable Values”. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, fall 2008 edn.

Joyce, James M. 2010. “A Defense of Imprecise Credences in Inference and Decision Making1”. *Philosophical Perspectives* 24: 281–323.

Mandler, Michael. 2005. “Incomplete Preferences and Rational Intransitivity of Choice”. *Games and Economic Behaviour* 50: 255–77.

McClennen, Edward F. 1990. *Rationality and dynamic choice: Foundational explorations*. Cambridge: Cambridge University Press.

Nozick, Robert. 1969. “Newcomb's Problem and Two Principles of Choice”. In *Essays in Honor of Carl Hempel*, edited by Nicholas Rescher. Dordrecht: D. Reidel.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley & Sons.

Schwartz, Barry. 2003. *The Paradox of Choice: Why more is less*. New York: Harper Collins.

Sen, Amartya. 1997. “Maximization and the Act of Choice”. *Econometrica* 65: 745–79.

———. 2000. “Consequential Evaluation and Practical Reason”. *Journal of Philosophy* 98: 477–502.

Simon, Herbert A. 1955. “A Behavioral Model of Rational Choice”. *The Quarterly Journal of Economics* 69: 99–118.

Temkin, Larry S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Weirich, Paul. 2004. *Realistic Decision Theory*. Oxford: Oxford University Press.

White, Roger. 2010. “Evidential Symmetry and Mushy Credence”. *Oxford Studies in Epistemology* 3: 161–86.