Economics and Philosophy, **31** (2015) 371–395 © Cambridge University Press
doi:10.1017/S0266267115000152 First published online 28 May 2015
journals.cambridge.org/eap

FAIR CHANCE AND MODAL CONSEQUENTIALISM

H. ORRI STEFÁNSSON*

Abstract: This paper develops a *Multidimensional Decision Theory* and argues that it better captures ordinary intuitions about fair distribution of chances than classical decision theory. The theory is an extension of Richard Jeffrey's decision theory to counterfactual prospect and is a form of *Modal Consequentialism*, according to which the value of actual outcomes often depends on what could have been. Unlike existing versions of modal consequentialism, the multidimensional decision theory allows us to *explicitly* model the desirabilistic dependencies between actual and counterfactual outcomes that, I contend, are at the heart of common intuitions about fair distribution of chances.

Keywords: fairness, counterfactuals, decision theory, consequentialism

1. INTRODUCTION

Consequentialists hold that the moral value of an alternative is determined by its consequences. This position however allows for a variety of different views, for instance depending on how narrowly we define *consequences*, and the way in which the values of different consequences are combined when evaluating the overall value of an alternative. This paper explores two views within this broad consequentialist school. One view, which I call *non-modal consequentialism* (NMC), holds that the moral value of an alternative is determined by its *non-modal consequences* and that there should be no interaction between consequences in different states of the world. The second view, which I call *modal consequentialism* (MC), states that the moral value of an alternative is determined by both its *modal and non-modal* consequences and that

* Fondation Maison des Sciences de l'Homme, Collge d'études mondiales, 190 avenue de France, 75013 Paris, France. Email: hlynur.orri@gmail.com. URL: www.orristefansson.is.

consequences in different states of the world can interact (in a sense explained below).[1]

I will use the following example (inspired by Diamond 1967) to explore the difference between modal and non-modal consequentialism:

Example. A hospital has only a single kidney but two patients, Ann and Bob, who are in equal need of the kidney, have equal rights to treatment, etc. Assume that in every respect that you find relevant for the decision of who should receive the kidney, Ann and Bob's situation is exactly symmetric. Moreover, Ann and Bob do not know that there is only one kidney but two patients in need of it, nor will they know why they got the kidney if they do, or why they didn't if they don't.

According to what I call the *Fair Chance View* (FCV), we should toss a fair coin, or hold some other lottery that gives each patient a 0.5 chance of winning, to decide whether Ann or Bob receives the kidney. Below I show that unlike modal consequentialism, non-modal consequentialism is inconsistent with the FCV. But existing versions of modal consequentialism do not, I contend, respect the intuition behind the FCV either. The main aim of this paper is to formulate a new modal-consequentialist theory that is consistent with the FCV.

The next section defines the Fair Chance View and the two versions of consequentialism more precisely. In section 2, I use Leonard Savage's (1972) classical decision theory to show the contradiction between the FCV and non-modal consequentialism, and explain why some have seen a property called *separability* as the main culprit. Section 3 briefly discusses a modal-consequentialist theory that does not satisfy separability. Although this theory can be made consistent with a preference for tossing a coin in situations like the one described above, it does not, I argue, satisfy the intuition behind the FCV. John Broome (1991) has famously shown that we can, given the right description of consequences, make separability compatible with the preference for tossing a coin. As I explain in section 4, the resulting theory is modal, but nevertheless violates the intuition behind the FCV. Finally, in section 5 I come to the main aim of the paper. Based on Richard Bradley's (2012) recent *Multidimensional Possible World Semantics for Conditionals*, I formulate a *Multidimensional Decision*

[1] This distinction has, to my knowledge, not yet received any attention. Phillip Pettit has recently argued that some goods, such as love and friendship, make *modal demands*, in that they should persist through changes or after events that will (in all likelihood) never actualize. (Pettit discussed this in his 2011 Uehiro Lecture at the University of Oxford.) Similarly, my discussion establishes that fairness often makes modal demands, in that its requirements concern what happens not only in the actual world but also in merely possible worlds. But as far as I am aware, Pettit does not discuss the aforementioned distinction within consequentialism.

Theory, which in effect extends Richard Jeffrey's (1983) decision theory to counterfactual prospects (and does not satisfy separability). I show that the version of modal consequentialism that emerges is compatible with the FCV. The new framework allows us to explicitly model the desirabilistic relationship between actual and counterfactual outcomes, which, I suggest in section 6, is among the advantages it has over other modal versions of consequentialism.

2. TWO FORMS OF CONSEQUENTIALISM

Most people seem to have the intuition that in circumstances like those described in the above example, we should hold a lottery to decide how to distribute the good in question. To justify this intuition from a consequentialist point of view, we need to show that the *consequences*[2] of holding the lottery are better than the consequences of giving the kidney to either Ann or Bob without holding a lottery. There may be many different consequentialist justifications of the discussed intuition. But the one I will focus on is the following: A consequence (or situation) where Ann has received the kidney as a result of a lottery is (strictly) morally better than a consequence where Ann has received the kidney without 'winning' it in a lottery, *because* in the former case Bob *had a chance*. I take this (commonly heard) justification to follow from the more general *Fair Chance View*:

Fair Chance View (FCV). Suppose n individuals are in equal need of an indivisible good $m < n$ of which we are about to distribute, and that the individuals are identical in every other respect that is morally relevant to the decision of who should receive a good. Then a situation (or consequence) where m of these individuals receive the good but all n individuals *had an equal chance* of receiving the good is (strictly) morally better than a situation where m of these individuals receive the good and it is not true that all n individuals had an equal chance of receiving the good.

Giving people a (or an equal) chance of getting a good, in situations like the one under discussion, is valuable in and of itself, according to the FCV as I understand it, rather than merely instrumentally valuable. (Moreover, I will assume that on this view, a situation where any of the individuals receive the good is fair just in case a lottery was used to determine who was to receive a good.) I will not attempt to make a normative assessment of the view, nor address the many and deep philosophical issues surrounding it. For instance, I will set aside questions about

[2] In what follows, I will talk of 'consequences', 'outcomes' and 'situations' interchangeably.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

whether the view requires that we distribute equally *objective* chances of receiving the good, or whether an equal distribution of *subjective* probabilities suffices (and if so, the subjective probabilities of *whom*). Instead, I will try to capture this common view a bit more formally.

The main thing to notice, for the present purposes, is the relation between chance and *counterfactuals*. What does it mean to say that even though Ann actually got the kidney, Bob had *a chance* of receiving it? It means that things *could* (in some meaningful sense) have turned out differently, and if they had, Bob would have received the kidney.[3] In particular, had some random event turned out differently than the way it actually did, Bob would have received the kidney. Using the possible world framework for counterfactuals, we can express this by saying that a situation in world w where Ann has received the kidney is made morally better by the 'existence' of a possible world w' that *only* differs from w in that, firstly, a random event turns out differently from the way it does in w, and, secondly, Bob receives the kidney.

As already indicated, one claim to be defended in this paper is that unlike modal consequentialism, non-modal consequentialism is incompatible with the Fair Chance View. Before defending this claim, let me define the two views a bit more precisely. Above I said that according to consequentialism, the moral value of an alternative is determined by its consequences. But this description may be somewhat misleading. To be more precise, let us say, using the terminology developed in Broome (1991), that according to consequentialism in its most general form, the value of an alternative is determined by how it distributes consequences across *locations*. In Broome's view, there are three *dimensions* of these locations to consider: different states of the world, different people and different times. To keep things simple, I will assume that each location in both the dimension of time and people is valued equally (as is the case according to most forms of *utilitarianism*). Hence, for the present purposes, I define consequentialism, in its most general form, as the claim that the value of an alternative is determined by how it distributes consequences across different states of the world.

This characterization is compatible with different forms of consequentialism, for instance depending on how consequences in different states of the world are weighted in the calculation of the overall value of an alternative.[4] But more importantly for the present discussion, consequentialism, thus characterized, comes in different forms depending

[3] The claim that Bob had an *equal* chance can just as naturally be captured in terms of counterfactuals. For it simply means that the counterfactual outcome where Bob receives the kidney was just as likely (when the lottery took place) as the actual outcome of the lottery.

[4] According to *ex ante* consequentialists, for instance, the consequence in each state of the world is weighted by the probability of that state being actual when the alternative is

on, first, whether we allow that modal properties matter for the moral value of the consequence in each (or some) state; and, second, whether we allow for the possibility that, when evaluating the overall value of an alternative, the contributions that consequences in different states make depend on consequences in other states.

What I call 'non-modal consequentialism' is a strict version of consequentialism, in that it neither allows that modal properties can be of moral importance nor for value interactions between consequences in different states of the world:

Non-Modal Consequentialism (NMC). The moral value of an alternative is determined by how it distributes non-modal consequences across different states of the world, and consequences in different mutually incompatible states of the world make independent contributions to the overall value of an alternative.

From a general consequentialist theory (as described above) we get non-modal consequentialism when we add the following principles:

First Principle of NMC. The moral value of a consequence in a particular state of the world is fully determined by the *non-modal* properties of that consequence.

Second Principle of NMC. If alternative A has different consequences depending on whether state of the world s_1, s_2, s_3, etc., turns out to be actual (where s_1, s_2, s_3, etc., are mutually incompatible), then for any of these s_i, the moral value that the consequence in s_i contributes to the overall moral value of A is independent of the consequence in any $s_j \neq s_i$.

I will call a consequentialist theory *modal* if it violates *either* the first or the second principle of NMC. Here is what the two principles have in common which justifies calling a theory that violates either of these modal. If a theory does not satisfy the first principle, then the *value* of a consequence in one state of the world may depend on what occurs in other states of the world, but if a theory does not satisfy the second principle, then the *contribution* that a consequence in some state of the world makes towards the overall value of an alternative may depend what occurs in other states of the world. So violations of the two principles have in common that there is some sort of value dependency between what occurs in different, mutually incompatible states of the world.

The second principle of NMC is related to a property called *separability* that has been much discussed in decision theory. Separability is usually

being considered. According to *ex post* consequentialists, however, consequences in all states except the actual one get weighted by 0.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

discussed as a property of *preferences* – or, in a moral context, as a property of what we might call 'betterness judgements'.[5] To explain separability, let us represent each alternative A by an n-tuple, e.g. $A = \langle a_1, a_2, ..., a_n \rangle$, where the a_is are the possible consequences of A. Now take the alternative A and create two new alternatives: A_b created by replacing a_i in the original alternative with b and A_c created by replacing a_i with c. Do the same for alternative D: create D_b by replacing d_i in the original alternative with b and D_c by replacing d_i with c. A betterness judgement (or preference) is separable just in case for any manoeuvre like the one just described, A_b is better than (or preferred to) A_c if and only if D_b is better than (or preferred to) D_c.

Decision theorists typically start with an ordering of alternatives, or a set of properties of orderings, and then show what kind of value functions can represent such an ordering (or an ordering with those properties). But it can be useful to start with a property of a valuation and see what ordering properties it implies. Let us suppose that when a non-modal consequentialist orders a set of alternatives according to 'betterness', she first finds out the moral value of each alternative, in accordance with the two principles of NMC, and then orders the alternatives according to moral value. (Moreover, let us restrict our attention to alternatives where both probabilistic and causal independence holds between the alternatives and the states of the world.) Then since moral value, according to her, satisfies the second principle of NMC, her betterness judgement satisfies separability. In the next section I will explain why some have thought that separability is inconsistent with a view like the Fair Chance View. In section 4 we will see that that thought is mistaken.

But first, let me briefly explain why I think the distinction I have just made is important. Surely *everyone* accepts that modal properties are morally important, someone might say. So what is the point of discussing this distinction between modal and non-modal consequentialism?[6] Well, the best known version of consequentialism, namely *classical utilitarianism*, is non-modal. According to classical utilitarianism, one should always choose the act that maximizes the total amount of pleasure over pain: 'the greatest happiness for the greatest number', as it is often put. Of course, people might feel (psychological) pleasure and pain due to what could have been. However, after we have accounted for such attitudes, an

[5] A 'betterness judgement', as I am using the term, is an overall comparative judgement. Hence, it is (formally) very much like preference. But to emphasize that the judgement in question may be objective (e.g. if some version of moral realism is true) I will often talk about betterness judgements rather than preferences. When talking about the requirements of decision in general, I will however talk about preferences.

[6] I thank a referee for *Economics and Philosophy* for pressing me on this issue.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

evaluation of alternatives should, according to classical utilitarians, satisfy both the first and the second principle of NMC.[7]

Economists and decision theorists have traditionally also been very reluctant to accept that what could have been matters for the (rational) evaluation of actual outcomes. In a classical defence of the 'Independence Axiom', found in some form or other in most decision theories (and implied by separability as previously defined), Nobel laureate Paul Samuelson for instance argues for formal separability as captured by the second principle of NMC, based on an intuition like the one expressed by the first principle of NMC. A simple version of the axiom Samuelson was defending states that if $(A)_1$ is at least as good as $(B)_1$ and $(A)_2$ is at least as good as $(B)_2$, then an alternative that results in either $(A)_1$ or $(A)_2$ depending on whether a coin comes up heads or tails, is at least as good as an alternative that results in either $(B)_1$ or $(B)_2$ depending on how the coin lands.[8]

Here is Samuelson's informal justification of the axiom:

> [E]ither heads or tails must come up: if one comes up, the other cannot; so there is no reason why the choice between $(A)_1$ and $(B)_1$ should be 'contaminated' by the choice between $(A)_2$ and $(B)_2$. (Samuelson 1952: 672–673)

In other words, the *reason* an evaluation or ordering of alternatives should satisfy separability (as the second principle of NMC states), is that there should be no desirabilistic dependencies between mutually incompatible outcomes (as the first principle of NMC states).

Finally, it might be worth mentioning that certain consequentialist theories should be classified as 'modal' for reasons that have nothing to do with what they say about the fairness of lotteries. These include John C. Harsanyi's (1977) and Richard Arneson's (1990) views. Both authors claim (roughly) that, for example, social policies should maximally satisfy people's *hypothetical* preferences; i.e. those preferences that people *would* have in ideal circumstances. What is true in a counterfactual world therefore makes a difference to the moral value of outcomes and

[7] Peter Hammond also famously defines consequentialism in a way that, in effect, makes what I am calling 'non-modal consequentialism' a non-consequentialist theory (see e.g. Hammond 1987, 1988).

[8] Notice that if $(A)_1 = (B)_1 = (A)_2$ then the axiom still says that if $(A)_2$ is at least as good as $(B)_2$, then an even chance of getting $(A)_1$ or $(A)_2$ is at least as good as an even chance of getting $(B)_1$ or $(B)_2$. Now interpret $(A)_1 = (B)_1 = (A)_2$ as *Ann received the kidney* and B_2 as *Bob receives the kidney*. Then the axioms says that since Ann receiving the kidney is considered as good as Bob receiving the kidney, an alternative that is guaranteed to give Ann the kidney is as good as a lottery that results with equal chance in either Ann or Bob receiving the kidney. In the next section we will see how a separable value function (the existence of which requires the preference that is being represented to satisfy some version of the Independence axiom) leads to a similar conclusion.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

alternatives in the actual world. Although I will focus on the fairness of lotteries in this paper, it should become clear that the framework developed in section 5 also provides a formal model in which to state and explore claims made by theories like Harsanyi's and Arneson's.

3. NON-MODAL CONSEQUENTIALISM VS. FAIR CHANCE

I will focus on Leonard Savage's version of decision theory to show why a non-modal consequentialist theory is incompatible with the Fair Chance View (Savage 1972). According to Savage's theory, the value of an alternative A, denoted by $U(A)$, is given by:

Savage's equation $U(A) = \sum_{s_i \in \mathcal{S}} u(s_i(A)) \cdot Pr(s_i)$

where $\mathcal{S}$ is a set of mutually exclusive and collectively exhaustive states of the world that determine the consequence of A, Pr a probability measure on $\mathcal{S}$, $s_i(A)$ represents the consequence of A when s_i happens to be the actual state of the world, and u is a value measure on (maximally specific) consequences.

Savage's equation satisfies the second principle of non-modal consequentialism; i.e. the separability property. In decision theoretic jargon, Savage's utility function is *additively separable*: the value of each alternative is a weighted *sum* of the values of each of its possible consequences. Hence, the value that $u(s_i(A))$ contributes towards the overall value of A is independent any $s_j(A)$. But, crucially, for Savage's theory to be an appropriate formalization of non-modal consequentialism, we need to assume that each $s_i(A)$ is a *non-modal* consequence.

To relate the above characterization of non-modal consequentialism back to the example of Ann, Bob and the kidney, let L be a lottery that gives Ann and Bob an equal chance of receiving the kidney (depending for instance on whether a fair coin lands heads up or tails up), and A (B) the alternative of giving the kidney to Ann (Bob) without holding a lottery. Then the Fair Chance View implies the following betterness judgement, which I will refer to as the *Fair Chance Judgement* (FCJ): $A \prec L$, $B \prec L$ (where '.. $\prec$ ...' denotes '... is worse than ...'). Then given that Savage's theory is an appropriate formalization of NMC, the latter is only compatible with the FCV if the following inequalities can simultaneously be satisfied:

$$(1) \qquad \sum_{s_i \in \mathcal{S}} u(s_i(A)) \cdot Pr(s_i) < \sum_{s_i \in \mathcal{S}} u(s_i(L)) \cdot Pr(s_i)$$

$$(2) \qquad \sum_{s_i \in \mathcal{S}} u(s_i(B)) \cdot Pr(s_i) < \sum_{s_i \in \mathcal{S}} u(s_i(L)) \cdot Pr(s_i)$$

where each $s_i(\alpha)$ is a *non-modal* consequence.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

Let ANN (BOB) represent the *consequence* where Ann (Bob) receives the kidney. Then A (B) is certain to have ANN (BOB) as consequence, but L will either result in ANN or BOB. Thus we can represent A (B) by ANN (BOB), and L by the n-tuple $\langle ANN, BOB\rangle$. There do not seem to be any modal properties built into the description of these consequences, so non-modal consequentialists should be happy with this representation of the three alternatives.

But now we run into trouble. For according to NMC, the value of the three alternatives is then given by: $U(A) = u(ANN)$, $U(B) = u(BOB)$ and $U(L) = u(ANN) \cdot 0.5 + u(BOB) \cdot 0.5$. But obviously, $u(ANN) \cdot 0.5 + u(BOB) \cdot 0.5$ can never be greater than $u(ANN)$ and *also* greater than $u(BOB)$. If one of $u(ANN)$ and $u(BOB)$ is greater than the other, then the value of the lottery L falls strictly between the values of the 'risk-free' alternatives A and B, but if the value of ANN and BOB is the same, then the value of the lottery will be the same as that of the risk-free alternatives. Hence, it seems, NMC is not compatible with the Fair Chance Judgement (and since the latter is implied by the FCV, NMC is incompatible with the FCV).

The above tension clearly has something to do with separability; in particular, the additively separable form of Savage's utility function. If the above is the right description of the alternatives, then if the value of an alternative is a probability weighted sum of the values of its possible consequences, then the value of L can never be greater than the value of both A and B. In the next section I discuss an attempt to make the FCV compatible with consequentialism by dropping separability, thus violating the second principle of NMC. But the above tension can also be put down to the way in which the alternatives have been described. In sections 4 and 5 I discuss two attempts, one old and one novel, to make the FCV compatible with consequentialism by describing the alternatives (and their consequences) in a way that violates the first principle of NMC. (The novel attempt also violates the second principle of NMC.) Perhaps unsurprisingly, I will argue that only my new solution succeeds in making consequentialism compatible with the intuition behind the FCV.

4. CONSEQUENTIALISM WITHOUT SEPARABILITY

Suppose we calculate the values of the three alternatives as follows:

$$U(A) = u(ANN) \cdot r(P(\top)),$$
$$U(B) = u(BOB) \cdot r(P(\top)),$$
$$U(L) = u(ANN) \cdot r(P(S)) + u(BOB) \cdot r(P(\neg S))$$

where $\top$ is a tautology and r a *risk function* of a *risk seeking* agent – i.e. an agent who prefers a gamble with an expected value of x to a risk-free alternative the value of whose consequence is x – and $\{S, \neg S\}$

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

is a partition of the possible states of the world into two equiprobable events.[9] Then we can represent the judgement that L is better than both A and B as maximizing risk-weighted utility; if, for instance, we allow for the possibility that $u(ANN) \cdot r(P(S)) + u(BOB) \cdot r(P(\neg S))$ is greater than *both* $u(ANN) \cdot r(P(\top))$ and $u(BOB) \cdot r(P(\top))$. And given how I have characterized consequentialism – i.e. as the view that the value of an alternative is determined by how it distributes consequences across states of the world – risk-weighted utility theory is a consequentialist theory.

This solution satisfies the first principle of NMC. For the consequences of the lottery, thus described (ANN and BOB), do not contain modal properties. But this solution is still incompatible with non-modal consequentialism, since it violates the second principle of NMC. Formally, the risk-weighed utility theory does not satisfy separability. To see this, notice that we can partition the tautology into S and $\neg S$, and then alternative A can be reformulated as the 'lottery' that has ANN as consequence in both the S-states and $\neg S$-states. Hence, this solution violates the second principle of NMC.

Should *modal* consequentialists be happy with this solution? There are, in my view, two related reasons for seeking an alternative way of making consequentialism compatible with the FCV. First, it seems to me that the Fair Chance View is an example of a more general phenomenon where what could have been is important for the evaluation of the desirability of what actually occurs. Unlike the above solution, the one I develop in section 5 explicitly models this relationship between what is and what could have been. Second, the above solution suggests that accepting the FCV has something to do with being risk seeking. More precisely, this way of making consequentialism compatible with the FCV suggests that the reason consequentialism as formalized by Savage seems incompatible with the FCV, is that that formalization places certain restrictions on attitudes towards risk. But those who accept my first objection will agree that the problem with Savage's framework, from the perspective of the FCV, is not so much the theory's restriction on risk attitudes, but rather its insensitivity to the desirabilistic relationships between what is and what could have been.

5. BROOME'S REDESCRIPTION STRATEGY

Contrary to my suggestion in section 2, many people will undoubtedly have the intuition that the consequence where Ann (Bob) receives the

[9] This could be seen as a variant of Lara Buchak's (2013) *Risk-Weighted Expected Utility Theory*, albeit with some important differences. (For instance, the way in which she defines r means that it is always the case that $r(P(\top)) = P(\top) = 1$, and that the expected utility of a lottery can never exceed the utilities of all of its possible prizes, contrary to what I am assuming here.)

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

kidney as a result of the lottery L is *not* the same consequence as Ann (Bob) receiving the kidney as a result of the risk-free alternative $A\,(B)$. The former consequence is *fair* whereas the latter is not, which must mean that these are not the same consequence. Hence, it seems, the consequences of A, B and L were not properly described in last section. Similarly to what Broome (1991: ch. 5) suggests, we should perhaps write the fairness directly into the description of the outcomes of the lottery; such that, for instance, L has $ANN\&Fair$, $BOB\&Fair$ as its two possible consequences, but $A\,(B)$ has $ANN\,(BOB)$ as the only possible consequence. And then the trouble we saw in section 2 disappears, since an (additively separable) EU function can, of course, simultaneously satisfy:

$$EU(A) = u(ANN) < EU(L)$$
$$= u(ANN\&Fair) \cdot 0.5 + u(BOB\&Fair) \cdot 0.5 \quad (3)$$

$$EU(B) = u(BOB) < EU(L)$$
$$= u(ANN\&Fair) \cdot 0.5 + u(BOB\&Fair) \cdot 0.5 \quad (4)$$

Broome's solution thus makes a preference for tossing the coin in the example discussed at that start of the paper compatible with consequentialism without giving up separability (i.e. without violating the second principle of NMC). But the solution clearly violates the first principle of NMC. Assuming that the FCV is part of our conception of fairness, then when we start refining our consequence set to include consequences that have fairness written into their description (to deal with examples like the one under discussion), we are in effect creating dependencies between consequences in different mutually incompatible states of the world. The consequence $ANN\&Fair$ for instance implicitly refers to what could have been, in the sense that a *necessary* condition for $ANN\&Fair$ to be a possible consequence of some alternative C, is that $BOB\&Fair$ is *also* (at least considered to be) a possible consequence of C.[10] For given the FCV, C can only result in $ANN\&Fair$ if C is some sort of lottery or random choice mechanism that has $BOB\&Fair$ as consequence in some state of the world. Hence, given that the moral value of the consequence of the lottery is, on Broome's suggestion, partly a function of this modal property, his suggestion violates the first principle of NMC.

The implicit reference to what could have been is precisely the reason why Broome's preferred description of the consequences runs into troubles with (what he calls) the *Rectangular Field Assumption* (RFA). RFA is a technical assumption of many of the traditional decision theoretic representation theorems, such as Savage's, needed (given the other

[10] That is, assuming that C is the initial choice of who should receive the kidney, rather than for instance the act of giving the kidney to Ann after she has won it in a lottery.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

assumptions of these theorems) to construct from an agent's preferences a value function that is unique up to positive affine transformation. Recall that an alternative can be represented by an n-tuple of consequences, e.g. $A = \langle a_1, a_2, ..., a_n \rangle$, where a_i is the consequence of alternative A if state of the world s_i happens to be actual. Given any set of alternatives that an agent's preferences are defined over, each state has associated with it a set of possible outcomes. Call that set for the i-th state C_i. Then take the product of all of these sets: $C_1 \times C_2 \times ... \times C_n$. Now the RFA states that *any* n-tuple created by picking consequences from some or all of the sets in the product is an alternative in the agent's preference ordering. That is, if we go through the sets, from C_1 to C_n, and pick arbitrary consequences from some or all of these sets, then the resulting n-tuple of consequences is an alternative in the agent's preference ordering.

Going back to our example, we could for instance pick two consequences: ANN and $ANN\&Fair$. The resulting alternative would then be the ordered pair $\langle ANN,\ ANN\&Fair \rangle$. This is an alternative where Ann gets the kidney in any state of the world; but, in addition, if some state turns out to be actual, then Ann gets the kidney fairly! But that is, of course, conceptually impossible on the Fair Chance View. For given this conception of fairness, either Ann or Bob, who have an equal claim to the good, can only receive the kidney fairly if some random mechanism has been used to determine who is to receive the good. But whenever such a random choice mechanism is used, it will *not* be the case that the same patient receives the kidney in all states of the world. In other words, an outcome is fair only if it is not true that that same patient receives the kidney in all possible states of the world.[11] Another way to put this, is that given Broome's description of consequences, the RFA requires that it be possible that a lottery that is unfair results in a consequence that is fair. It seems clear that this requirement goes against the intuition behind the FCV. Hence, it is not at all clear that Broome's 're-description strategy' makes the FCV compatible with Savage's consequentialist framework.[12]

[11] It could be objected that the alternative where Ann gets the kidney in any state of the world and receives it fairly if the coin comes up heads, is the alternative that results if the decision maker decides to let a coin toss determine who receives the kidney, but is so biased towards Ann that he is only able to stick to his decision if a side favourable to Ann (heads) comes up. Hence, when heads comes up, Ann does receive the kidney as a result of winning it in a lottery, and thus the outcome is fair. (I thank Weyma Lübbe for bringing this objection to my attention.) I do not find this objection convincing, however, since we are evaluating a prospect where we know, before learning the result of the coin toss, that Ann will receive the kidney no matter what side comes up. So although there is some sense in which Ann receives the kidney as a result of a lottery if the coin comes heads up, it is also true that she will receive the kidney irrespective of the result of the coin toss. Hence, I contend, all outcomes of the 'lottery' are unfair.

[12] It may be worth stating the RFA in the terminology of Savage's framework. Here, the assumption is that any function from the state space $\mathcal{S}$ we are working with to the

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

I should emphasize that the tension with the Rectangular Field Assumption is not the main reason I think we should seek alternative ways of making the Fair Chance View compatible with consequentialism. Not all decision theoretic representation theorems require the RFA,[13] and there are well known problems with the assumption that have nothing to do with fairness. The main problem I have with Broome's solutions, is rather that unlike the solution suggested in next section, his fails to make explicit the desirabilistic dependency between actual and counterfactual outcomes that seems at the heart of the FCV. Examining the tension between Broome's solution and the RFA nevertheless does serve an important role in the present argument, since it illustrates the difficulty in dropping the first principle of non-modal consequentialism while holding on to Savage's framework. The reason the RFA creates trouble for Broome is that some consequences as Broome describes them refer, as we have seen, to what occurs in states of the world in which they themselves do not occur, and thus cannot be combined with any arbitrary consequence as the RFA requires. In other words, the tension with the RFA stems from Broome's violation of the first principle of NMC.

6. MULTIDIMENSIONAL DECISION THEORY

I now finally turn to formulating a version of modal consequentialism that better satisfies the intuition behind the Fair Chance View than the theories already discussed. The framework I develop is a multidimensional extension of Richard Jeffrey's (1983) decision theoretic framework to counterfactuals. The resulting theory allows us to represent various different judgements according to which counterfactual outcomes influence the desirability of actual outcomes. A general discussion of such judgements would take us too far from the topic of this paper, so I will

consequence set $\mathcal{C}$, is an *act* in the agent's preference ordering. Assume that H and T (standing for e.g. *coin comes up heads* and *coin comes up tails*) are two events that partition the state space. Then if $\mathcal{C}$ contains both ANN and $ANN\&Fair$, the function

$$f^*(s_i) = \begin{cases} ANN & \text{if } s_i \in T, \\ ANN\&Fair & \text{if } s_i \in H. \end{cases}$$

should be an act in the agent's preference ordering. This is an act that has the consequence that Ann receives the kidney in any state of the world, and moreover receives it fairly if a state if H happens to be actual.

[13] In particular, the Bolker-Jeffrey theorem for Jeffrey's decision theory does not contain the assumption (see e.g. Bolker 1966). Broome's solution can easily be reformulated for Jeffrey's framework.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

focus on showing that the multidimensional theory can represent the Fair Chance Judgement as maximizing desirability.[14]

According to Jeffrey's theory, both probabilities and desirabilities are measures on a Boolean algebra Ω of propositions (that is, a set of propositions closed under negation and conjunction) from which impossible (zero probability) propositions have been removed.[15] The desirability of any particular proposition, p, is according to Jeffrey's measure a weighted average of the desirabilities of the different (mutually exclusive) ways in which the proposition can be true, where the weight on each way p_i that proposition p can be true is given by the conditional probability of p_i given p. More formally:

Jeffrey's equation (a) $Des(p) = \sum_{p_i \in p} Pr(p_i \mid p) \cdot Des(p_i)$

A proposition, in Jeffrey's framework, is a set of possible worlds. So probabilities and desirabilities are measures on a set of sets of worlds. We can thus interpret the p_is as different *worlds* compatible with the proposition p; i.e. think of $\{p_1, p_2 \ldots\}$ as a world partition that is equivalent to p. So, if we like, we can formulate Jeffrey's equation as:

Jeffrey's equation (b) $Des(p) = \sum_{w_i \in p} Pr(w_i \mid p) \cdot Des(w_i)$

Jeffrey's evaluation of propositions is clearly consequentialist, given how I have characterized consequentialism. The possible ways in which a proposition can come true can be understood as the possible *consequences* of the proposition coming true, and we can interpret that which determines the way in which a proposition comes true (if it comes true) as a 'state' of our world. Representing the Fair Chance Judgement as maximizing the value of a Jeffrey-desirability function is therefore one possible way of making the Fair Chance View compatible with consequentialism. However, someone who evaluates propositions according to Jeffrey's equation will not always satisfy separability, since the contingencies that determine how a proposition becomes true are not, in Jeffrey's theory, probabilistically independent of the proposition that is being evaluated.[16]

[14] A more general and technical discussion of the theory is forthcoming in Bradley and Stefánsson (2015), where we also show what needs to be added to the framework developed here to get an expected utility representation.

[15] For the quasi-uniqueness part of the Bolker-Jeffrey representation theorem, Ω has to be atomless.

[16] Suppose for instance we create from p two new propositions, p_a by replacing $w_i \in p$ with w_a and p_b by replacing $w_i \in p$ with w_b; and we do the same for q. Then it will

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

Although Jeffrey's theory violates separability, the theory as Jeffrey himself interpreted it – i.e. as a version of *non-modal* consequentialism (where Ω contains only factual propositions) – nevertheless runs into the same problem we have seen with Savage's: there is no pair of desirability and probability functions relative to which the Fair Chance Judgement can be represented as maximizing desirability. To see this, now let ANN (BOB) be the *proposition* that Ann (Bob) receives the kidney, and L a proposition that can come true in one of two ways, ANN or BOB, and does so with equal probability. Then for the Fair Chance View to be compatible with Jeffrey's original theory, there has to be a function Des such that:

(5) $Des(ANN) < Des(ANN) \cdot Pr(ANN \mid L) + Des(BOB) \cdot Pr(BOB \mid L)$

(6) $Des(BOB) < Des(ANN) \cdot Pr(ANN \mid L) + Des(BOB) \cdot Pr(BOB \mid L)$

which implies that:

(7) $$Des(ANN) < 0.5Des(ANN) + 0.5Des(BOB)$$

(8) $$Des(BOB) < 0.5Des(ANN) + 0.5Des(BOB)$$

But again, a probability mixture of the desirabilities of ANN and BOB can of course never exceed the desirability of both ANN and BOB.[17]

It could be objected that the consequences of the lottery should be formulated as $ANN \wedge L$, in which case there will be pairs of desirability and probability functions relative to which the Fair Chance Judgement can be represented as maximizing desirability. But $ANN \wedge L$ is not a non-modal consequence: in effect, this description of the consequence has built into it that the consequence in question had 0.5 chance of occurring (since L is the proposition that either ANN or BOB will occur with equal chance). Making Jeffrey's theory compatible with the Fair Chance Judgement by including the lottery in the description of the consequences moreover suffers from the same problem as Broome's suggestion, namely that it does not make explicit the importance of counterfactuals for the Fair Chance View.

not necessarily be the case that $Des(p_a) < Des(p_q)$ if and only if $Des(q_a) < Des(q_b)$, since the conditional probabilities of the worlds, that are used to calculate the desirabilities of propositions, may differ.

Another way to see that Jeffrey's theory is not separable, is to notice that the representation theorem for Jeffrey's theory does not contain a strong separability axiom, such as Savage's Sure Thing Principle, but instead the considerably weaker Averaging axiom (see Appendix 2).

[17] A more detailed discussion the clash between Jeffrey's original theory and the FCV is forthcoming in Bradley and Stefánsson (2015).

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

	Q	$\neg Q$
P	w_1	w_2
$\neg P$	w_3	w_4

TABLE 1. Worlds-sentences

Actuality	Possible Situations	
w_1	$\langle w_1, w_1\rangle$	$\langle w_1, w_2\rangle$
w_2	$\langle w_2, w_1\rangle$	$\langle w_2, w_2\rangle$
w_3	$\langle w_3, w_1\rangle$	$\langle w_3, w_2\rangle$
w_4	$\langle w_4, w_1\rangle$	$\langle w_4, w_2\rangle$

TABLE 2. Space of possibilities

I will base the extension of Jeffrey's theory to counterfactuals on Richard Bradley's (2012) *Multidimensional Possible World Semantics for Conditionals*. The basic ingredients in the multidimensional semantics are *n-tuples of worlds*, $\langle w_1, w_2, w_3, w_4, ...\rangle$ (i.e. *ordered* sets of worlds), where the first world representa a potential actual world and the rest represent potential *counter-actual* worlds under different suppositions. To keep things simple, let's assume that what is being supposed true are *sentences*. Then w_2 might e.g. represent a counter-actual world under the supposition that sentence P is true, w_3 a counter-actual world under the supposition that $\neg P$ is true, w_4 a counter-actual world under the supposition that Q, etc. In what follows I will refer to such an n-tuple as a possible *situation*: $\langle w_1, w_2, w_3, ...\rangle$ is the *situation* that w_1 is actual, w_2 would be actual if P, w_3 would be actual if $\neg P$, etc.

Let us focus on a model with only four worlds, w_1 to w_4, and suppose Table 1 represents how worlds and sentence match up. In other words, the sentence P is true at worlds w_1 and w_2, Q at worlds w_1 and w_3. To explain the semantics it is useful to focus on only one supposition: the supposition that P. Table 2 represents the eight possible situations; for each $\langle w_i, w_j\rangle$ the possibility that w_i is the case and w_j would be if P.[18]

Now suppose the sentence P expresses the *proposition* p. Thus the sentence P is true if and only if p. So p is true in the situations in the first and second row of the table, but false in the situations in the third and fourth row. And given that we identify a proposition with the situations where it holds true, we have $p = \{\langle w_1, w_1\rangle, \langle w_1, w_2\rangle, \langle w_2, w_1\rangle\langle w_2, w_2\rangle\}$. So

[18] Assuming Centring, which states that if P is true at w_i then the counter-actual world to w_i under the supposition that P is w_i itself, eliminates the situations $\langle w_1, w_2\rangle$ and $\langle w_2, w_1\rangle$. Nothing of substance depends on whether we make this assumption or not.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

a proposition, on the multidimensional semantics, is a set of n-tuples of worlds.

Now let $\Box\!\!\rightarrow$ stand for the *counterfactual conditional connective*.[19] The conditional sentence $P\Box\!\!\rightarrow Q$ is true, on the multidimensional semantics, if and only if Q is true in the world that is counter-actual under the supposition that P. In other words, $P\Box\!\!\rightarrow Q$ is true if Q is true in the second world in the n-tuple that correctly represents what is the case and what would be the case under the supposition that P. So $P\Box\!\!\rightarrow Q$ is made true by the situations in the first column of the above table, but made false by situations in the second column. And the *proposition* $p\Box\!\!\rightarrow q$ is thus the set $\{\langle w_1, w_1\rangle, \langle w_2, w_1\rangle, \langle w_3, w_1\rangle, \langle w_4, w_1\rangle\}$.

Let us then return to the project of representing the Fair Chance Judgement.[20] Now let P express the proposition (p) that the coin comes up heads and $\neg P$ the proposition that the coin comes up tails. Let Q express the proposition (q) that Ann receives the kidney and $\neg Q$ the proposition that Bob receives the kidney. We have thus made two simplifying assumptions already. Firstly, it might seem more natural to let P ($\neg P$) express the *conditional* that the coin comes heads (tails) up *if* tossed. But nothing is lost if we think of this just as a factual sentence. Secondly, we have limited our attentions to situations where either Ann or Bob receives the kidney. But what is distinctive about the FCV is what it has to say about situations where a number of individuals have an equal claim to an indivisible good that *some* but not all of them get. (Any kind of welfarism for instance condemns a situation where *none* of the needing patients receives the kidney.) Hence, since we want to focus on what is special about the view, it seems justifiable to limit our attention to situations where one of Anna and Bob does receive the kidney.

To represent the Fair Chance Judgement in a multidimensional model we need to work with two suppositions: the supposition that P and the supposition that $\neg P$. But actually, we only need to consider two dimensions at a time, since the FCJ only orders alternatives according to what is true and what would be true under a *contrary-to-factual* supposition. That is, the judgement for instance is that a situation where the coin comes up heads and Ann receives the kidney, is made better or worse depending on whether Ann also receives the kidney under the (contrary-to-factual) supposition that the coin comes up tails. But it says nothing about whether the desirability of this situation depends on what is true under the (matter-of-factual) 'supposition' that the coin comes up

[19] The multidimensional semantics also gives truth conditions for *indicative* conditionals. But for the present purposes we only need to consider counterfactuals.

[20] The above is far from providing a complete description of the multidimensional semantics. But it should suffice for the present purposes. (Consult Bradley (2012) for the details.)

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

Actuality	Possible Situations	
w_1	$\langle w_1, \theta, w_3\rangle$	$\langle w_1, \theta, w_4\rangle$
w_2	$\langle w_2, \theta, w_3\rangle$	$\langle w_2, \theta, w_4\rangle$
w_3	$\langle w_3, w_1, \theta\rangle$	$\langle w_3, w_2, \theta\rangle$
w_4	$\langle w_4, w_1, \theta\rangle$	$\langle w_4, w_2, \theta\rangle$

TABLE 3. Space of possibilities 2

Bad	Good
$\langle w_1, \theta, w_3\rangle$	$\langle w_1, \theta, w_4\rangle$
$\langle w_2, \theta, w_4\rangle$	$\langle w_2, \theta, w_3\rangle$
$\langle w_3, w_1, \theta\rangle$	$\langle w_3, w_2, \theta\rangle$
$\langle w_4, w_2, \theta\rangle$	$\langle w_4, w_1, \theta\rangle$

TABLE 4. The good, the bad

heads. Hence, for a situation where P is true at the actual world, we need not represent the 'counter'-world under the supposition that P; and similarly for the situation where $\neg P$ is true. So although semantically we are now working with a three-dimensional model, we only need to worry about two dimensions at a time.

Let θ denote the world we need not consider at each time. If P is true, then θ is the 'counter'-world under the supposition that P (which is the second world as I am setting up the n-tuples) but if $\neg P$ is true, then θ is the 'counter'-world under the supposition that $\neg P$ (which is the third world as I am setting up the n-tuples). Table 3 thus represents the space of possible situations we need (call the whole space $\mathcal{W}$).

I will assume that it makes no difference, according to the FCV, whether Ann or Bob actually receives the kidney, since by assumption their situation is symmetrical in all ways that are relevant for the decision of who should receive the kidney. Hence, the fair situations are equally good when Ann receives the kidney as when Bob receives the kidney and similarly for the unfair situations. The FCJ thus orders the situations into two equivalence classes, represented in Table 4, where every situation from the 'Good' class is better than every situation from the 'Bad' class, but any two situations within the same class are equally good.

What is common to the ('Bad') situations in the left column is that the person who actually receives the kidney would also have received it had the coin come up differently. In the situations in the right column, however, whoever actually receives the kidney would not have received it had the coin landed differently.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

The Fair Chance Judgement can now be formulated as follows:[21]

$$\begin{aligned} &\langle w_1,\theta,w_3\rangle \sim \langle w_2,\theta,w_4\rangle \sim \langle w_3,w_1,\theta\rangle \sim \langle w_4,w_2,\theta\rangle \\ (10)\qquad &\prec \langle w_1,\theta,w_4\rangle \sim \langle w_2,\theta,w_3\rangle \sim \langle w_3,w_2,\theta\rangle \sim \langle w_4,w_1,\theta\rangle \end{aligned}$$

and the FCJ can be represented by a function V that satisfies:[22]

$$\begin{aligned} &V(\langle w_1,\theta,w_3\rangle) = V(\langle w_2,\theta,w_4\rangle) = V(\langle w_3,w_1,\theta\rangle) = V(\langle w_4,w_2,\theta\rangle) \\ (11)\ &< V(\langle w_1,\theta,w_4\rangle) = V(\langle w_2,\theta,w_3\rangle) = V(\langle w_3,w_2,\theta\rangle) = V(\langle w_4,w_1,\theta\rangle) \end{aligned}$$

There will certainly be many functions satisfying 11 (and thus representing 10): any *ordinal* utility function, defined over a set of world-triples, can represent this ordering.[23] So to some extent we have reached the aim of making the FCV compatible with (modal) consequentialism. For we have found a way of showing that there is a function whose assignment of values to situations corresponds to whether the FCV deems the situation fair. And we do not have to worry about clashes with the Rectangular Field Assumption, since the assumption is not needed for the existence of such a function.

I have however not yet shown that there is a *Jeffrey* desirability function that represents 10. But we can do so by construction. Let V and Pr assign values to the basic situations (i.e. the ordered triples) in $\mathcal{W}$. The functions extend to any *proposition* r (i.e. to any set of situations) according to the following rules:

$$(12)\qquad V(r) = \sum_{\langle w_i,w_j,w_k\rangle \in r} V(\langle w_i,w_j,w_k\rangle)\cdot Pr(\langle w_i,w_j,w_k\rangle \mid r)$$

$$(13)\qquad Pr(r) = \sum_{\langle w_i,w_j,w_k\rangle \in r} Pr(\langle w_i,w_j,w_k\rangle)$$

Then by construction V is a Jeffrey desirability function: the desirability of a proposition, according to this function, is a weighted average of the

[21] Each possibility in 10 is a proposition, and in fact a conjunction of a factual and a counterfactual proposition. $\langle w_4,w_2,\theta\rangle$ for instance is the proposition that $\neg p \wedge \neg q \wedge p \boxrightarrow \neg q$, $\langle w_4,w_1,\theta\rangle$ the proposition that $\neg p \wedge \neg q \wedge p \boxrightarrow q$, etc. Hence, 10 is equivalent to:

$$\begin{aligned} &(p \wedge q \wedge \neg p \boxrightarrow q) \sim (p \wedge \neg q \wedge \neg p \boxrightarrow \neg q) \sim (\neg p \wedge q \wedge p \boxrightarrow q) \\ &\sim (\neg p \wedge \neg q \wedge p \boxrightarrow \neg q) \prec [(p \wedge q \wedge \neg p \boxrightarrow \neg q) \sim (p \wedge \neg q \wedge \neg p \boxrightarrow q) \\ (9)\qquad &\sim (\neg p \wedge q \wedge p \boxrightarrow \neg q) \sim (\neg p \wedge \neg q \wedge p \boxrightarrow q)] \end{aligned}$$

See Appendix 2 for a suggestion of how to calculate the Jeffrey-desirability of a counterfactual proposition.

[22] A function V is said to represent a judgement J just in case V assigns a higher value to A than to B if and only if A is better than B according to J, but the same value to A and B if and only if they are equally good according to J.

[23] Further details about such value functions are discussed in Bradley and Stefánsson (2015).

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

desirabilities of the different ways in which the propositions can come true (i.e. the different situations compatible with the proposition), where the weights are given by the appropriate conditional probabilities.

Now let's see whether this function can represent the Fair Chance Judgement, as formulated in 10. Recall the two equivalence classes of propositions (sets of n-tuples) induced by the FCJ. Let us call the 'good' equivalence class G and the 'bad' equivalence class $\neg G$. We can stipulate that:

1. $\forall \langle w_i, w_j, w_k \rangle \in G: \quad V(\langle w_i, w_j, w_k \rangle) = 1$
2. $\forall \langle w_l, w_m, w_n \rangle \in \neg G: \quad V(\langle w_l, w_m, w_n \rangle) = -1$

Then it is clear that V represents the FCJ, as formulated in 10: for any two basic situations, $\langle w_i, w_j, w_k \rangle$ and $\langle w_l, w_m, w_n \rangle$, we have: if $\langle w_i, w_j, w_k \rangle \sim \langle w_l, w_m, w_n \rangle$ according to FCJ, then $V(\langle w_i, w_j, w_k \rangle)$ and $V(\langle w_l, w_m, w_n \rangle)$ are both either -1 or 1; but if $\langle w_i, w_j, w_k \rangle \prec \langle w_l, w_m, w_n \rangle$ according to FCJ, then $V(\langle w_i, w_j, w_k \rangle) = -1$, $V(\langle w_l, w_m, w_n \rangle) = 1$. So we have constructed a Jeffrey desirability function that represents the FCJ.

G and $\neg G$ are also propositions (sets of n-tuples of worlds), and by the above stipulation: $V(G) = 1$ and $V(\neg G) = -1$. But for any arbitrary proposition r:

$$\begin{aligned} V(r) &= V(G) \cdot Pr(G \mid r) + V(\neg G) \cdot Pr(\neg G \mid r) \\ &= Pr(G \mid r) - Pr(\neg G \mid r) \end{aligned} \tag{14}$$

Up to now I have been focusing only on propositions that are either completely fair or completely unfair; i.e. propositions that are either subsets of G or $\neg G$ but do not overlap the two sets. And I have formulated the FCJ as only having something to say about propositions that are either completely fair or unfair. But we can easily construct propositions that overlap the two sets. Let's call such propositions 'mixed'. $m = \{\langle w_1, \theta, w_3 \rangle, \langle w_1, \theta, w_4 \rangle\}$ is a mixed proposition, for instance, since the first of its elements is an unfair situation but the second is fair. Intuitively, a proposition like m is a *biased* lottery. In $\langle w_1, \theta, w_3 \rangle$ Ann gets the kidney no matter how the coin lands. So since this situation is possible given m, but no situation that gives Bob a greater chance than Ann is possible given m, m is biased towards Ann.

Now take any two mixed propositions m_1 and m_2: suppose each is a set of two triples, one of which is an element of G but the other of $\neg G$. According to V, as I have constructed it, $V(m_1) \leq V(m_2)$ just in case $Pr(G \mid m_1) \leq Pr(G \mid m_2)$. It is natural to think of $Pr(G \mid m_i)$ as measuring how unbiased m_i is: if $Pr(G \mid m_i) = 1$ then m_i is not at all biased but if $Pr(G \mid m_i) = 0$ then m_i is maximally biased. So the more (less) biased a proposition is, the lower (higher) value V assigns to it. Although I have

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

said nothing about how the Fair Chance View judges biased lotteries, this is exactly what we should want.[24] For any mixed propositions m_1 and m_2, the former should be better than the latter, on this view, if and only if it is less biased; but if they are equally biased, then they should be equally good (or bad). So the function I have constructed does not only capture the intuition that a situation where Ann (Bob) receives the kidney is fair only if Bob (Ann) had a chance; it also captures the intuition that situations where they both had an *equal* chance at receiving the kidney are more fair than situations where their chances were unequal.

7. SOME IMPLICATIONS

The framework developed in the last section has the advantage over the other modal consequentialist theories I have been discussing in that it explicitly models the desirabilistic relationship between what is and what could have been, which seems at the heart of the Fair Chance View. In addition, it has the following advantage over Broome's suggestion for how to make consequentialism compatible with the preference for tossing the coin. Broome's solution consists in adding a primitive fairness property to the outcome space. My suggested solution however consists in enriching decision theory to include counterfactual prospects in such a way that the fairness property *emerges* as a relationship between actual and counterfactual outcomes. Thus, I contend, my solution better explains what orthodox decision theory lacks when it comes to capturing common intuitions about fair distribution of chances, such as the intuition underlying the FCV.

Broome might of course complain that I have myself added some primitive variable to decision theory, namely the (counterfactual) supposition operator. But those who are already motivated by the Fair Chance View, or more generally recognize that what could have been is often important for the evaluation of actual outcomes, hopefully agree that this extra complexity is more than offset by the benefit of being able to formally represent desirabilistic dependencies between facts and counterfactuals.

Using the multidimensional framework to represent the Fair Chance Judgement has the interesting implication that the extra value generated by the truth of the relevant counterfactual does not supervene on the non-modal facts. The table representing the 'goodness partition', for instance, has each actual world in both the 'Good' and the 'Bad' column. The situations $\langle w_1, \theta, w_3 \rangle$ and $\langle w_1, \theta, w_4 \rangle$ for instance share all non-modal facts and differ only in what *would* be true if $\neg P$ were. But the latter is

[24] In fact this needs to hold for the FCJ to satisfy the basic (e.g. Bolker-Jeffrey) rationality axioms.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

fair whereas the former is not, which suggests that fairness does not supervene on non-modal facts. More generally, if the multidimensional semantics, formulated as Bradley suggests, is the right semantics for counterfactuals, then counterfactuals don't supervene on non-modal facts.[25] But then if the moral value of a situation partly depends on what counterfactuals are true, as the FCV states, then the moral value of a situation does not supervene on its non-modal facts, contrary to what Broome (1991: 114–115) claims.

The non-modal consequentialist will without a doubt point out that the failure of fairness to supervene on non-modal facts is merely an artefact of our model. And she might argue as follows. This failure of counterfactuals to supervene on non-modal facts, according to the version of the multidimensional semantics I have been using here, is a reason for looking for a different semantics for counterfactuals, if we want to insist that fairness is partly determined by what counterfactuals are true. According to the best known semantics for counterfactuals – i.e. the Stalnaker–Lewis semantics (see e.g. Stalnaker 1968; Lewis 1986) – counterfactuals *do* supervene on (and are implied by) factual propositions.[26] Let f_1 be the set of factual propositions that (on this semantics) imply the counterfactual $\neg A \mathbin{\square\!\!\rightarrow} B$ and f_2 the set of factual propositions that imply the counterfactual $\neg A \mathbin{\square\!\!\rightarrow} \neg B$. What really explains our judgement that $V(A \wedge B \wedge \neg A \mathbin{\square\!\!\rightarrow} \neg B)$ can be different from $V(A \wedge B \wedge \neg A \mathbin{\square\!\!\rightarrow} B)$, the non-modal consequentialist might argue, is the fact that for us, $V(A \wedge B \wedge f_2)$ might be different from $V(A \wedge B \wedge f_1)$. To put the point in a less abstract manner, although the truth of a particular counterfactual is *one* difference between a situation where Ann receives the kidney fairly and one where Ann receives the kidney unfairly, what *really* makes the moral difference is the set of factual propositions that *implies* the relevant counterfactual.

I do not want to argue against the view that counterfactuals and other modal facts supervene on non-modal facts. However, there are various arrangements of non-modal facts that can make true the particular modal facts we are interested in. For instance, there are various ways of making it true that Ann and Bob have an equal chance of receiving the kidney. Many of these arrangements of non-modal facts are equally good, from a moral perspective. And what makes them morally good, is the fact that they all entail the relevant modal fact; in the case we are considering, the

[25] The same is true given Hannes Leitgeb's recent semantics for counterfactuals (Leitgeb 2012a, 2012b). See, however, Stefánsson (2014) for a formulation of the multidimensional semantics where everything supervenes on non-modal facts.

[26] I am assuming here that both the possible worlds that serve as truth makes for counterfactuals and the relevant similarity relation is entirely determined by the facts of the actual world (as was clearly the intention of Lewis at least).

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

fact that Ann and Bob have an equal chance. In other words, the *reason* all these different arrangements of non-modal facts are morally good, if the supervenience thesis is true, is that they entail this particular modal fact.

Moreover, and more generally, we should make a distinction between facts that *carry* value and facts on which the carriers of value supervene.[27] Even if it is true, as Humeans claim, that all facts supervene on the non-modal facts, that does not, by itself, mean that these non-modal facts are the carriers of value. Perhaps the following analogy will help. Every intrinsic (as opposed to relational) property of a painting is determined by how the atoms that make up the painting are arranged. Hence, the aesthetic qualities of the painting supervene on this arrangement of atoms. These aesthetic qualities, most people think, carry some value, that is not carried by the atoms that make up the painting. Similarly, whether or not Bob could have received the kidney may be determined by the non-modal facts of our world. But that does not mean that this particular counterfactual carries no value over and above these non-modal facts. A plausible consequentialist theory must take that into account.

ACKNOWLEDGEMENTS

I would like to thank Campbell Brown, Lara Buchak, Alan Hájek, Julian Jonker, James Joyce, Wlodek Rabinowicz, Katie Steele, Weyma Lübbe and three referees for *Economics and Philosophy* for helpful comments on earlier versions of this paper. Special thanks to Richard Bradley for very helpful discussions about all parts of the paper. This paper was presented at Formal Ethics 2012, The Tenth Annual Berkeley-London Graduate Conference and The Sixteenth Annual BPPA Conference. I am grateful to the participants of these conferences for their questions and comments.

REFERENCES

Arneson, R. J. 1990. Liberalism, distributive subjectivism, and equal opportunity for welfare. *Philosophy and Public Affairs* 19: 158–194.

Bolker, E. D. 1966. Functions resembling quotients of measures. *Transactions of the American Mathematical Society* 124: 292–312.

Bradley, R. 2012. Multidimensional possible-world semantics for conditionals. *Philosophical Review* 121: 539–571.

Bradley, R. and H. O. Stefánsson 2015. Counterfactual desirability. *British Journal for the Philosophy of Science* (forthcoming).

Broome, J. 1991. *Weighing Goods*. Oxford: Basil Blackwell.

Buchak, L. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Diamond, P. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy* 75: 765–766.

Hammond, P. 1987. Consequentialism and the Independence axiom. In *Risk, Decision and Rationality*, ed. B. Munier, 503–516. Dordrecht: Springer.

[27] I thank Wlodek Rabinowicz for suggesting this wording.

Hammond, P. 1988. Consequentialist foundations for expected utility theory. *Theory and Decision* 25: 25–78.
Harsanyi, J. C. 1982/1977. Morality and the theory of rational behaviour. In *Utilitarianism and Beyond*, ed. A. Sen and B. Williams, 39–62. Cambridge: Cambridge University Press.
Jeffrey, R. 1990/1983. *The Logic of Decision*. Chicago, IL: The University of Chicago Press.
Leitgeb, H. 2012a. A probabilistic semantics for counterfactuals. Part A. *Review of Symbolic Logic* 5: 26–84.
Leitgeb, H. 2012b. A probabilistic semantics for counterfactuals. Part B. *Review of Symbolic Logic* 5: 85–121.
Lewis, D. 1986. *Counterfactuals*. Oxford: Blackwell (revised edition).
Samuelson, P. A. 1952. Probability, utility, and the independence axiom. *Econometrica* 20: 670–678.
Savage, L. 1972. *The Foundations of Statistics*, Revised edition. New York, NY: Dover Publication.
Stalnaker, R. 1968. A theory of conditionals. In *Studies in Logical Theory*, ed. N. Rescher Oxford: Blackwell.
Stefánsson, H. O. 2014. Humean supervenience and multidimensional semantics. *Erkenntnis* 79: 1391–1406.

APPENDIX 1: THE BOLKER-JEFFREY AXIOMS

Let Ω be an atomless Boolean algebra of propositions from which the impossible proposition has been removed and '$\preceq$' a binary relation on Ω. (The relations I have been using, '$\prec$' and '$\sim$' are defined from '$\preceq$' as follows: $p \prec q$ iff $p \preceq q \wedge \neg(q \preceq p)$; $p \sim q$ iff $p \preceq q \wedge q \preceq p$.) Then any preferences that satisfies the following (Bolker-Jeffrey) axioms can be represented as maximising the value of a (Jeffrey-) desirability function:

Completeness. For any $p, q \in \Omega$: $p \preceq q$ or $q \preceq p$.
Transitivity. For any $p, q\ r \in \Omega$: if $p \preceq q$ and $q \preceq r$ then $p \preceq r$.
Averaging. If $p, q \in \Omega$ are mutually incompatible, then

- if $p \prec q$ then $p \prec p \vee q$ and $p \vee q \prec q$, but
- if $p \sim q$ then $p \sim p \vee q$ and $p \vee q \sim q$

Impartiality. If $p, q \in \Omega$ are mutually incompatible and $p \sim q$, then if $p \vee r \sim q \vee r$ for some r that is mutually incompatible with both p and q but $\neg(r \sim p)$, then $p \vee r \sim q \vee r$ for every such r.
Continuity. Suppose $p \prec s \prec q$ or $p \prec i \prec q$, where s and i are respectively the supremum and the infimum of a chain of propositions. Then there will be a member r of the chain such that if t is a member of the chain that is implied by (or implies) r, then $p \prec t \prec q$.

APPENDIX 2: DESIRABILITY OF COUNTERFACTUALS

Recall that according to Jeffrey's formula, the desirability of a proposition p is a weighted average of the different ways p_i in which it can become true, where the weights are given by $Pr(p_i \mid p)$. What is then the desirability of the counterfactual $p \mathrel{\square\!\!\rightarrow} q$? Given the multidimensional semantics, the question is equivalent to the

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press

question of the desirability of the set of situations (i.e., set of n-tuples of worlds) where $p \boxrightarrow q$ is true.

The answer obviously depends on our semantics for counterfactuals, which determines the different ways in which a counterfactuals can become true. But suppose, as I did above, that we want to be as 'liberal' as possible in this regard, and e.g. allow for the possibility that $p \boxrightarrow q$ is true even though $p \wedge \neg q$ is. In other words, we do not accept *Weak Centring* for counterfactuals (which is logically equivalent to not accepting Modus Ponens for counterfactuals). Then the (Jeffrey) desirability of $p \boxrightarrow q$ is given by:

$$\begin{aligned} Des(p \boxrightarrow q) = {} & Des(p \wedge q \wedge (p \boxrightarrow q)) \cdot Pr(p \wedge q \mid p \boxrightarrow q) \\ & + Des(p \wedge \neg q \wedge (p \boxrightarrow q)) \cdot Pr(p \wedge \neg q \mid p \boxrightarrow q) \\ & + Des(\neg p \wedge q \wedge (p \boxrightarrow q)) \cdot Pr(\neg p \wedge q \mid p \boxrightarrow q) \\ & + Des(\neg p \wedge \neg q \wedge (p \boxrightarrow q)) \cdot Pr(\neg p \wedge \neg q \mid p \boxrightarrow q) \end{aligned} \tag{15}$$

Even those who accept Weak Centring for counterfactuals should be able to accept the above formula. On their view, $Pr(p \wedge \neg q \mid p \boxrightarrow q) = 0$, which means that the second summand makes no difference to the desirability of $p \boxrightarrow q$.

BIOGRAPHICAL INFORMATION

H. Orri Stefánsson was, when writing this article, a PhD student at the London School of Economics and Political Science, working mainly on the foundations of decision theory and philosophical issues surrounding counterfactuals. He is currently a Postdoctoral Research Fellow at Collège d'études mondiales (Paris), where he works on the foundations of risk analysis and the philosophy of risk and chance.

https://doi.org/10.1017/S0266267115000152 Published online by Cambridge University Press