The Instability of Savage's *Foundations*: The Constant-Acts Problem

Edward Elliott

1. Introduction

Savage's seminal work, *The Foundations of Statistics* (1954), is centred around one of the most well-known and admired representation theorems ever developed. David Kreps described Savage's theorem as the "crowning glory of choice theory" (1988, p. 120); and likewise, in summarising his widely-cited review of over two dozen expected utility theorems, Peter Fishburn had this to say:

> Savage's [theorem] is suitable for a wide variety of situations, its axioms are elegant and intuitively sensible, and its representation-uniqueness result is easily connected to assessment techniques [...] I regard it as one of the best. (1981, p. 194)

The admiration for Savage's work shows through in its influence; it would not be unfair to characterise axiomatic decision theory since 1954 as a series of footnotes to Savage.[1] The majority of decision-theoretic representation theorems that exist today—for both classical expected utility theory and non-classical theories of decision-making—are based upon the same basic formal system as Savage's, usually with only minor tweaks here and there.

Despite all this—or perhaps because of it—Savage's *Foundations* has also generated an enormous amount of criticism. At the forefront of this critique is the so-called *constant acts problem*.[2] As we will see, exactly what this problem *is*, and how much of a problem it might be, are themselves difficult questions to answer. I will begin in §2 by describing Savage's formal framework and theorem in some detail. Then, in §3, I will look at the problem of constant acts.[3]

2. Savage's Foundations

My exposition of Savage's work will be in three parts. In §2.1, I begin with a discussion on the notion of an *act*, one of the core concepts underlying the standard interpretation of Savage's theorem (and other theorems based on the same framework). In §2.2, I will then discuss three technical notions involved in the statement of the theorem—i.e., *states*, *outcomes*, and

[1] Savage himself was greatly influenced by Bernoulli (1738), Ramsey (1931), de Finetti (1931, 1964), and von Neumann and Morgenstern (1944), amongst others.

[2] Two further complaints that are commonly made against Savage's theorem, particularly within the economics literature, are (i) that he requires his set of states to be uncountable, and (ii) the so-called *problem of small worlds*. I will not discuss either of these problems here. See (Joyce 1999, pp. 70-7, 110-13) for a thorough discussion of the latter.

[3] This paper is an excerpt from Chapter 5 of my thesis, *Representation Theorems and the Grounds of Intentionality*. In it, I argue that (contrary to a common opinion) the "constant acts problem" does not clearly present a problem for many important philosophical applications of Savage's (and related) representation theorem(s). I also argue, however, that there are other difficulties with Savage's framework which are not easily avoided.

act-functions. Then, in §2.3, I outline Savage's preference conditions and say a few words about the final representation result.

2.1 Acts

It will be helpful to begin our exposition with a three-fold distinction between *behaviours*, *bodily movements*, and *acts*, due to Dretske (1988). We will say that a *bodily movement* includes any process by which an external change occurs in the physical state of some animate body. So, for instance, should you raise your hand, that is one kind of movement, but it would also be a movement if an external force—say, another agent, or a mysterious force of nature—were to compel your arm to raise without your intending it so. Likewise, if you were to trip and fall, you will have again undergone some form of bodily movement.

A *behaviour* is a specific kind of bodily movement—*viz*., a movement of a living body produced specifically by some cause or causes *internal* to that body. To use Dretske's example,

> A rat moving its paw is a process in which the movement of the rat's paw is brought about by activities occurring *in* the rat. If the same paw movement is produced by an external cause, the rat's paw moves, but the rat doesn't move it. There is no *rat* behaviour. (Dretske 1990, p. 783)

Note that a behaviour just needs to be produced by *internal* causes; it need not be *intentional*. If the rat's paw movements were caused by a malfunction in the rat's limbic system, this would still be a kind of rat behaviour. Presumably, a great deal of our behaviour is non-intentional—involuntary tics, snoring, hard-wired reactions to a source of pain, blinking, and breathing are in most cases unintentional behaviours.

We can then characterise an *act* as any behaviour which is the voluntary result of an intention—a *deliberate* or *purposeful* behaviour. On an intuitive level, acts are the basic *objects of choice* in any decision situation, the things we choose between when we are deciding what to do. For instance, when bored, one might choose to *read a book* or *go fishing*; at night, one might *go out for drinks*, *go to bed*, or *stay in watching cartoons*; in a game of poker, *hold 'em* or *fold 'em*.

Two things to note about acts. First, there are limits to the kinds of acts we might perform: we cannot choose to *run faster than the speed of light*, for instance, nor to *negate the influence of gravity upon one's body*. Such things we could not realise even if we intended to, so they are not acts available to us for choice. More generally, we can roughly characterise an act as *available to S* if S would in fact perform it should she so intend. And second, acts can be described with different degrees of specificity. For example, to *read Moby Dick* is one way to *read a book*, but it is not the only way; and both are specific ways to *do something*.

Given this three-fold distinction, we can say a little bit about the interpretation of Savage's preference relation, $\succcurlyeq$. In Savage's own words, the relata of $\succcurlyeq$ are described as acts, with $\succcurlyeq$ itself being described as follows: "Loosely speaking, [α] $\succcurlyeq$ [β] means that, if [the agent] were required to decide between [α] and [β], no other *acts* being available, he would decide on [α]"

(1954, p. 17, emphasis added). The vast majority of the literature has followed Savage in treating the decision-theoretic $\succcurlyeq$ in this way, and for present purposes we can as well.[4]

It is worth pausing very briefly to note that Savage (and many who have followed him) considered his theorem to be supplying a fully behavioural definition of preferences and, consequently, of degrees of belief and utilities. Within the behaviouristic tradition originating with Samuelson (1938, 1948)—still very much alive today in certain fields of enquiry—$\succcurlyeq$ is understood as somehow directly encoding a subject's *behavioural dispositions* in different kinds of choice situations without any need of recourse to a prior of the subject's intentional mental states. If this were so, the relata of $\succcurlyeq$ would have to be taken as *behaviours* generally (as opposed to acts more specifically), and characterised in wholly non-intentional terms. However, it is clear on reflection that this way of understanding $\succcurlyeq$ is a non-starter—not only because of the well-known difficulties associated with providing the required non-intentional specifications of $\succcurlyeq$ and its relata, but (moreover) because the resulting theorem will have us attributing *every* behaviour a subject makes to some *choice* of theirs—indeed a rational calculation of expected utilities—*even when this clearly isn't the case*.[5] We will therefore in all that follows continue treating the relata of $\succcurlyeq$ as *acts* proper, rather than mere *behaviours*.

Suppose then that we have a non-empty set, $\mathcal{A}'$ = {α, β, γ, …}, containing a range of acts available to some subject *S* in an unspecified decision situation. The members of this set will form the basic relata of our preference relation, the basic objects of choice. As not every act in $\mathcal{A}'$ can be realised by the decision-maker, $\mathcal{A}'$ should be understood as containing act types rather than tokens.[6] This set will be critically important for understanding Savage's framework, but before it can be put to work in that regard we will need get more precise about how $\mathcal{A}'$ is supposed to be specified.

First of all, $\mathcal{A}'$ should be specified in such a way that *S* must perform at least one act in $\mathcal{A}'$, and such that the performance of any one such act in $\mathcal{A}'$ should preclude the performance of any other. Or, equivalently, $\mathcal{A}'$ should be such that *S* is guaranteed to perform exactly one of its members.[7] Thus, for example, if *read Moby Dick* were in $\mathcal{A}'$, then *read a book* could not be, because the latter is compatible with the former; but *read the Odyssey* might yet be. The reason for this restriction, very roughly, is that we will later want to characterise *S* as choosing amongst the acts in $\mathcal{A}'$, so (i) we had better not be double counting any of her options, and (ii) we had better not leave out any potential objects of choice.

Secondly, the acts in $\mathcal{A}'$ ought to be characterised at a rather fine level of specificity. Exactly *how* fine is something which can only be stated precisely after we have characterised the other elements of Savage's theorem, but for now the rough idea is simply that acts should be specified *as finely as makes a difference to the things we care about*. For example, it would not do to let one of the acts in $\mathcal{A}'$ be *do something*, for there are many different ways by which

[4] There are problems already with this formulation, some of which are discussed thoroughly in (Maher 1993, pp. 12-15).

[5] Just below, I will also note a number of constraints on $\succcurlyeq$'s relata which also highlight the difficulty (if not impossibility) of providing a fully non-intentional interpretation of that relation.

[6] Alternatively, one could think of $\mathcal{A}'$ as a set of propositions which specify that *S* performs one of the acts available to her—there are no deep issues that arise from construing $\mathcal{A}'$ as a set of acts or propositions about acts.

[7] On the construal of $\mathcal{A}'$ as a set of propositions about which act *S* performs, this then turns into the requirement that $\mathcal{A}'$ is a partition of some (as yet unspecified) space of possibilities.

someone might *do something* and these different ways will in general have a very great impact upon how the world turns out in ways that we care about—one could *read a book* or *jump off a cliff*, say, and these two actions will have drastically different consequences for us. Indeed, it would not do to let one of the acts in $\mathcal{A}'$ be just *read a book*, because whether the result is enjoyable or not enjoyable will depend on the specific book read. Continuing this path of reasoning, it should become immediately clear that the acts in $\mathcal{A}'$ have to be specified very finely indeed—perhaps more so than we ourselves are even capable of doing without a great deal of reflection.

Finally, every act α in $\mathcal{A}'$ should be such that *S* is certain that she would perform α, if she were to intend as such. For instance, *S* might *intend* to *travel to New York*, but whether she *succeeds* or not depends on a number of factors outside of her control which could, for all she knows, prevent her from arriving. On the other hand, in most cases she can, say, *reach for the nearest object*, and she can be sure that she will succeed in doing so should she so choose.

The motivation for this final condition can be made evident with the help of an example:[8]

> Before Jill is a red button, above which is a sign reading 'PRESS ME FOR $100!' Jill knows that she can push the button easily, and also knows that the button will only do something if it's pushed—however she is not certain *what* it will do. As a matter of fact, the sign is accurate and pushing the button will cause $100 to pop up from a hidden compartment, free for her to take with no strings attached. Jill could do with the money, but she does not believe the sign: she knows that a prank-centred TV show is in town, and is (for good reason) rather more confident that she is on camera, and that pushing the button will only result in her receiving a painful electric shock or some other cruel outcome. Jill chooses to leave the button alone.

Clearly, Jill *would* have been able to press the button had she so intended, and she knows this. Furthermore, if she had so intended, she *would* have received $100 as a result of pressing the button. It would be admissible to let $\mathcal{A}'$ be {*push the button*, *leave the button alone*}. But it would be problematic if we were to characterise $\mathcal{A}'$ as {*receive $100 by pushing the button*, *leave the button alone*}. Jill needs the money, and if she *knew* that she could *receive $100 by pushing the button* then she most certainly would have chosen that option rather than preferring to leave the button alone. She did not push the button because she did not know that receiving $100 was one of her options.[9] If acts are characterised as those things which *S would* perform if she intended to, *such that S* is certain that she would be successful if she so chose, then *receive $100 by pushing the button* (and *receive painful electric shock by pushing the button*) will not be amongst Jill's available acts—but *push the button* and *leave the button alone* will be. This seems to be as things should be—otherwise it would be exceedingly odd that Jill's choices reveal a preference for not pushing the button over receiving $100.

[8] A similar case to this is Brian Hedden's 'Raging Creek' example, in (Hedden 2012, pp. 347-8).

[9] If the reader is uncomfortable with treating *receives $100 by pushing the button* as an act, alternative examples which make essentially the same point are easy to come by. Ultimately, all that is required is a mismatch between the acts that are actually available to an agent and the acts she believes are available, where her choices would have been very different had she been aware of the facts regarding her available options.

In the rest of this subsection, I want to raise three issues that arise from the foregoing restrictions on $\mathcal{A}'$. It would be possible to skip ahead to §2.2 without losing anything of great importance to understanding the rest of the paper.

The first issue is that requiring that Jill is *certain* of her capacity to perform any act in $\mathcal{A}'$ may rule out too much—there are very few acts which Jill is *absolutely* certain she can perform. Note, in particular, that this constraint is in tension with the first constraint mentioned, that $\mathcal{A}'$ be specified such that *S must* perform at least one of its members. In effect, the two constraints together imply that for any act *S* performs, there must be some way of describing that act such that *S* was certain she could have performed it, had she so intended. This is by no means obviously possible. Nevertheless, something like this restriction is required to make sense of the fact that in Savage's theorem (and all similar theorems), preference-rational agents are implicitly modelled as being certain of their capacity to perform any of the acts over which they have preferences: this is why the expected utility of performing an act α is calculated through consideration of α's—and *only* α's—potential outcomes (as described below). If the agent gave some substantial credence to the thought that by *intending* to perform α, she might instead end up performing β, then presumably some consideration of β's possible outcomes should play a proportionate role in her deliberations about whether to try to do α.

Another problem case arises when *S* is *certain* that she can perform α, but—as a matter of fact—if she were to try, she would fail. One can be certain about falsehoods! Perhaps Jill is mistakenly certain that pushing the button would destroy the universe. In this case, it is entirely unclear whether *destroy the universe by pushing the button* should be included in Jill's range of available acts—Jill herself seems to think it is, and that fact has an explanatory role to play in her decisions. To avoid answering this question, Sobel (1986) has suggested that rational agents can never be certain of a falsehood. I have some doubt that this is true, but in any case, the same cannot be said for the ordinary person on the street. A weaker suggestion would be that if rational agents are certain they can perform α, then they can. However, on any natural conception of an act, this still seems too strong—and it does not help us to characterise $\mathcal{A}'$ for non-rational agents.

These and similar considerations lead Schwarz (MS, pp. 7-11) to suggest that a broadly Savagean decision theory is best thought of *not* as a theory about preferences over *acts*—conceived of as things like *go to the park*, *get a drink from the fridge*, and so on—but instead as a theory about preferences over *intentions*—specifically, intentions to act in different ways. Hedden (2012) defends a nearby view, though he casts his position in terms of 'decisions' rather than intentions. It seems at least somewhat more plausible, for example, to suggest that the ordinary agent has complete epistemic access to what *intentions* she might form, so that she can reliably be certain that if she decided to *intend* to perform α, then she would be successful in intending as such.

Fitting this into the Savagean framework, one *might* understand the forming of an intention as a special kind of *mental act*—the kind of thing which, as with our ordinary acts, will have variable consequences depending on the different ways the world might turn out, of which we are uncertain.[10] If this turns out to be the best way to interpret Savage's system, then it again puts the lie to the purported *behaviouristic* construal of his formalism.

[10] Note that on this construal, *intend to read a book* and *intend to read Moby Dick* can, at least on one reading of the former, be treated as distinct and mutually exclusive acts described at the same level of specificity. Our

The suggested interpretation is not without its own problems, though—in particular, we are left wanting a characterisation of when $\alpha \succcurlyeq \beta$ holds, given that α and β are now being construed as two *intentions*. Certainly, it would be difficult to see it as a disposition to form the one intention rather than the other *under the condition that only one of those two intentions can be formed*. For one thing, it would be a far off possible world indeed where one is only able to form one of only two intentions. Moreover, the kinds of things which tend to limit our ability to form intentions are our beliefs about what we are able to do, so the foregoing characterisation of $\succcurlyeq$ would likely have us considering a person's preferences at counterfactual scenarios where the subject's beliefs are *radically* different than they are at the actual world (where she presumably believes she can perform a great many acts). It is doubtful that any information we could glean from such scenarios would be very helpful in telling us anything about our degrees of belief and utilities.

2.2 States, Outcomes, and Act-functions

An act will, in general, have a range of different *outcomes*, depending on the different ways the world might be. If you were to read a book that you have never come across before, then you might either *become entertained* or *become annoyed*, depending on the (presently unknown to you) contents of its pages; and if you were to go fishing, you might *catch a fish* or *catch nothing*, depending on what's in the water.

Let $\mathcal{O} = \{o_1, o_2, o_3, \ldots\}$ contain descriptions of *each* of the possible outcomes that might arise given any act in $\mathcal{A}'$, focused in particular on describing those states of affairs that the our subject *S cares about*. (*S* will probably not care, for instance, that if she were to *go fishing*, then *she will (still) have an even number of pencils in her office*, so we can leave that out of the description of the outcome.) As Savage describes the outcomes in $\mathcal{O}$, "They might in general involve money, life, state of health, approval of friends, well-being of others, the will of God, or anything at all about which the person could possibly be concerned" (Savage 1954, p. 14). For reasons to be clarified below, the descriptions ought to be not only maximally specific with respect to what *S* cares about, but they also be mutually exclusive. Since exactly one act in $\mathcal{A}'$ must be performed, the set of outcomes will therefore be jointly exhaustive of the possibilities.

Next, we will need a set of the *states*, $\mathcal{S} = \{s_1, s_2, s_3, \ldots\}$, upon which the different outcomes of *S*'s acts depend. Like $\mathcal{O}$, $\mathcal{S}$ should be a *partition* of some possibility space;[11] i.e., a collection of propositions such that exactly one is true. Savage does not explicitly describe states in much detail, and the description that he does give is not very helpful. There are, however, two critically important properties that we need to assume states have if Savage's theorem is to have a plausible interpretation *qua* decision theory, which I will outline now (and motivate a little later).

First of all, states should be *independent* of whatever act the agent might choose to perform. In the literature, this property of states is referred to as *act-independence*. As Allan Gibbard and William Harper (1978) have pointed out, Savage's system is compatible with (at least) two notions of independence being applied in the precisification of this requirement. The first

intentions themselves can be given with different levels of specificity, but this should not be confused with the sense noted above in which different acts can be described at different levels of specificity.

[11] I will leave it open *which* space, though it is best characterised in terms of doxastic or epistemic possibilities.

is *evidentially independence*, where a state s is *evidentially independent* of the performance of an act α just in case S's credences that s is true under the assumption that she performs α is equal to her credences that s is true under the assumption that she does not perform α.[12] The second kind of independence they refer to as *causal*, though it would be better termed *counterfactual independence*. A state s is *counterfactually independent* of the performance of an act α just in case s would hold if α were performed, and s would hold if α were not performed.

For the purposes of the present exposition, it's not important which of these two notions of independence is used. I will, however, note a consequence of applying either—namely, that states must be *logically independent* of acts:

> **Definition 1: Logical independence**
> A state s is *logically independent* of the performance of an act α iff s is consistent with α being performed and α not being performed

This allows us to define an important property of states within Savage's system, which we'll call *act-independence*:

> **Definition 2: Act-independence**
> A state s is *act-independent* (with respect to a specification of $\mathcal{A}'$) iff s is logically independent of the performance of any $\alpha \in \mathcal{A}'$

As Savage requires that every state in $\mathcal{S}$ is act-independent, no state can entail that a particular act in $\mathcal{A}'$ is chosen (or not chosen).

Secondly, states should be *outcome-functional*:

> **Definition 3: Outcome-functionality**
> A state s is *outcome-functional* (with respect to a specification of $\mathcal{A}'$ and $\mathcal{O}$) iff the performance of any s-compatible $\alpha \in \mathcal{A}'$ at s uniquely determines that a particular outcome $o \in \mathcal{O}$ obtains

The upshot of assuming outcome-functionality is that, for each state s, there will be a function which maps every act in $\mathcal{A}'$ which can potentially be performed at s to an outcome in $\mathcal{O}$. If *every* act in $\mathcal{A}'$ is compatible with the state s, as act-independence requires, then these will be *total functions* on $\mathcal{A}'$.

Before moving on, it is worth emphasising again that act-independence and outcome-functionality are not *formal* requirements on the specification of $\mathcal{S}$, which for the purposes of the theorem itself may be characterised sparsely as *any* non-trivial partition of a non-empty set.

[12] Evidential independence is standardly characterised in terms of *probabilistic independence*; *viz.*, if $\mathcal{B}el$ is a probability function, then s is evidentially independent of the performance of α (relative to $\mathcal{B}el$) just in case $\mathcal{B}el(s|\textit{perform}\ \alpha) = \mathcal{B}el(s|\textit{don't perform}\ \alpha)$, where $\mathcal{B}el(P|Q) = \mathcal{B}el(P \,\&\, Q)/\mathcal{B}el(Q)$. If s is evidentially independent of all acts in $\mathcal{A}'$, which are by hypothesis mutually exclusive and jointly exhaustive, then for any act $\alpha \in \mathcal{A}'$, $\mathcal{B}el(s|\textit{perform}\ \alpha) = \mathcal{B}el(s|\textit{don't perform}\ \alpha) = \mathcal{B}el(s)$. I have avoided this formulation of evidential independence because of its use of conditional probabilities, the application of which raises concerns insofar as S isn't probabilistically coherent. There are some difficulties with the formulation of evidential independence given here, but the precise formulation is not important for the discussion that follows.

Rather, act-independence and outcome-functionality are two properties that we must assume the states in $\mathcal{S}$ to have, *if* Savage's theorem is to have a plausible interpretation *qua* decision-theory. Without these background interpretive assumptions, the proof of the theorem will still go through, but the theorem itself will (probably) lack any useful interpretation for the purposes of decision theory.

In Savage's framework, one can see states as the ultimate objects of uncertainty: it is from $\mathcal{S}$ that Savage constructs the domain of his $\mathcal{B}el$ function—namely, the set of *events*, $\mathcal{E} = \{E_1, E_2, E_3, \ldots\}$. Every event is a set of states, and most theorems in the Savagean tradition will treat $\mathcal{E}$ as an uncountable σ-algebra on $\mathcal{S}$. Savage, however, assumes something somewhat stronger: that *every* set of states is included in $\mathcal{E}$ (i.e., $\mathcal{E} = 2^{\mathcal{S}}$).

Although events are *technically* sets of states rather than propositions *per se*, we do no harm in treating each event as a proposition.[13] As states are pairwise inconsistent, every set of states corresponds directly to one and only one proposition, *viz*., the disjunction of each of the states in the set. We will therefore treat events as though there were just propositions. It should also be clear, given this way of characterising events as just disjunctions of states, that they inherit the event-equivalent property of act-independence from the states of which they are composed. That is, if every state is act-independent, then every disjunction of states is act-independent as well. On the other hand, events do not have anything like the outcome-functionality property that states have.

Savage's central insight was the recognition that, given the way we have characterised $\mathcal{S}$ and $\mathcal{O}$, each act in $\mathcal{A}'$ can be uniquely modelled by a function form $\mathcal{S}$ to $\mathcal{O}$. The idea is that each such function determines a unique definite description that identifies a particular act that the agent might perform—or at least a class of acts which are, from the perspective of the decision-maker, not worth distinguishing:[14]

> If two different acts had the same consequences in every state of the world, there would from the present point of view be no point in considering them two different acts at all. An act may therefore be identified with its possible consequences [at different states of the world]. (1954, p. 14)

Suppose that $\mathcal{F}$ is the function that pairs the state s_1 with the outcome o_1, s_2 with o_2, and so on; we can then say that $\mathcal{F}$ represents:

> the act α in $\mathcal{A}'$ such that, were it performed, then (if s_1 were the case, o_1 would result) & (if s_2 were the case, o_2 would result) & …

We will refer to any function from a set of states to outcomes as an *act-function*. Formally, Savage's $\succcurlyeq$ is defined not on a primitively given set of acts $\mathcal{A}$ but instead on a set of act-functions, and it is this feature which essentially characterises the influential formal paradigm that he developed. For most theorems within this paradigm, act-functions are total functions

[13] As characterised above, a state is a proposition, and an event therefore a set of propositions. A set of propositions is not itself a proposition.

[14] Of course, if the outcomes are specified in enough detail, it's highly unlikely that two acts would have the same outcomes across all states.

on $\mathcal{S}$ and often only take a finite number of values from $\mathcal{O}$, though this varies from theorem to theorem. In the literature, act-functions are more often called *Savage acts* or sometimes just *acts*; however, it will be helpful for the discussion that follows to distinguish the *functions* and the *acts* that they are supposed represent.[15]

Note that the representation of acts as *total* functions from $\mathcal{S}$ to $\mathcal{O}$ would be nonsensical if some states were logically incompatible with the performance of some acts—what sense would it make to speak of an act's outcome at a state which *implies* that the act is not performed? Likewise, outcome-functionality is required if a function from states to outcomes is to represent an act along the lines described—if, for example, α could only ever result in either o_1 or o_2, but every state in $\mathcal{S}$ left it indeterminate which of these outcomes would result, then there would be no reason to suppose that α corresponds to one function from $\mathcal{S}$ to $\{o_1, o_2\}$ rather than any other.

With the set of events specified as the set of all subsets of $\mathcal{S}$, it's worth noting that every one of Savage's act-functions can be expressed equivalently as a mapping from a set of mutually exclusive and jointly exhaustive *events* to outcomes, simply by collecting together the states with similar outcomes into a single event. For example, if $\mathcal{F}(s) = o_1$ for all states s in E, and $\mathcal{F}(s) = o_2$ for all states s in $\neg E$, then we might represent $\mathcal{F}$ as $(E, o_1 | \neg E, o_2)$. More generally, assume the following convention for representing act-functions:

> **Definition 4: Act-function notation**
> $\mathcal{F} = (E_i, o_i | \ldots | E_n, o_n)$ iff $\{E_i, \ldots, E_n\}$ is a partition of $\mathcal{S}$ and if $s \in E_i$, $\mathcal{F}(s) = o_i$, …, and if $s \in E_n$, $\mathcal{F}(s) = o_n$

This convention will be helpful in laying out Savage's preference conditions and formal results more transparently.

So far, I have been treating *acts* as a kind of conceptual primitive, with *states*, *outcomes*, and *events* being partially characterised by their relations to the acts in $\mathcal{A}'$. In Savage's formal system, however, the situation appears rather different. Savage theorem begins with just two primitive sets: $\mathcal{O}$ and $\mathcal{S}$, where all that is required of $\mathcal{O}$ is that it contains at least two members, and all that is required of $\mathcal{S}$ is that it is a non-trivial partition of some non-empty set. Sparse characterisations, to be sure, but this hides the informal properties they must have if they are to stand for collections of *outcomes* and *states* respectively. There is no *formal primitive* which corresponds to $\mathcal{A}'$. Rather, from $\mathcal{S}$ and $\mathcal{O}$, Savage constructs the set which we will label $\mathcal{A} = \{\mathcal{F}, \mathcal{G}, \mathcal{H}, \ldots\}$, which on Savage's construction contains *all* total functions from $\mathcal{S}$ to $\mathcal{O}$ (i.e., $\mathcal{A} = \mathcal{O}^{\mathcal{S}}$).

The order of this construction is somewhat misleading—suggesting as it has to many that acts can be straightforwardly *defined* in terms of states and outcomes, where the latter can be specified independently of any account of what the set of available acts are. This is not at all the case, as the informal characterisation of $\mathcal{S}$ above should by now have made clear. States are characterised as necessarily consistent with the performance of any act in $\mathcal{A}'$ and such that the performance of any act in $\mathcal{A}'$ determines a unique outcome in $\mathcal{O}$. There is no sense to be

[15] Making this terminological distinction also helps to highlight the fact that there are many different ways of understanding *what* act-functions represent, and *how* they represent. There are a large number of interesting questions here, some of which I will discuss in §3.

made of $\mathcal{S}$ as containing states and outcomes *as they were described above* without some way of specifying of $\mathcal{A}'$ that does not simply define the latter in terms of the former. There is, therefore, a sense in which the set of acts proper, $\mathcal{A}'$, is a kind of *informal primitive* which underlies any decision-theoretic interpretation of Savage's formal framework.

Given that the states in $\mathcal{S}$ are act-independent and outcome-functional (with respect to a choice of $\mathcal{A}'$ and $\mathcal{O}$), it's clear that every act in $\mathcal{A}'$ can be uniquely represented by a particular act-function in the manner described above. It is far less clear, however, that every possible act-function in $\mathcal{A}$ corresponds a member of $\mathcal{A}'$. Nevertheless, as mentioned, Savage assumes that *all* possible act-functions are in $\mathcal{A}$. This includes, famously, *constant act-functions*. That is, for each outcome o in $\mathcal{O}$, there is a constant act-function in $\mathcal{A}$ that maps every state in $\mathcal{S}$ to o. Because of their importance, it will be helpful to have special notation for constant act-functions:

Definition 5: Constant act-functions
$\underline{o} = \mathcal{F}$ iff $\mathcal{F}(s) = o$ for all $s \in \mathcal{S}$

Assuming that the outcomes are specified rather finely—as they must be, for reasons we will return to shortly—it's extremely doubtful that any constant act-function could serve to represent anything *real* that an agent might choose to do: what acts are there which would bring about any given outcome, regardless of how the world turns out to be? Nothing in the pre-theoretic, intuitive construal of the space of possible acts seems to have this character. In a nutshell, this one way of understanding the *problem of constant acts*.

Constant act-functions play a number of important roles in Savage's theorem. For instance, Savage uses preferences between constant act-functions to construct a relative utility ranking upon the set of outcomes, which eventually gives rise to the utility function $\mathcal{D}es$—the idea being simply that the subject prefers the constant act $\underline{o}_1$ to $\underline{o}_2$ just in case she attaches a higher utility to o_1 than to o_2. This simple idea then finds application in Savage's definition of a relative confidence relation, $\succcurlyeq^b$, defined on the space of events. The ability to construct $\succcurlyeq^b$ from $\succcurlyeq$ is crucial for establishing the existence of Savage's $\mathcal{B}el$ function. In the literature, the definition Savage gives has come to be known as Savage's principle of *Coherence*:

Definition 6: Coherence
For all $E_1, E_2 \in \mathcal{E}$, $E_1 \succcurlyeq^b E_2$ iff, for any $o_1, o_2 \in \mathcal{O}$, if $\underline{o}_1 \succcurlyeq \underline{o}_2$ then $(E_1, o_1 \,|\, \neg E_1, o_2) \succcurlyeq (E_2, o_1 \,|\, \neg E_2, o_2)$

This highly influential principle is *prima facie* intuitive—at least on the assumption that $(E_1, o_1 \,|\, \neg E_1, o_2)$ and $(E_2, o_1 \,|\, \neg E_2, o_2)$ *actually correspond to things the agent can choose to do*. Suppose that the agent finds o_1 more desirable than o_2. Then, if she is given a choice between two acts which each might result in either o_1 or o_2 but under different circumstances, our subject should prefer the act which, from her perspective, has the greater likelihood of resulting in o_1, and the smaller likelihood of resulting o_2. If she finds E_1 more likely than E_2 then, accordingly, she should find $(E_1, o_1 \,|\, \neg E_1, o_2)$ to be the more desirable act than $(E_2, o_1 \,|\, \neg E_2, o_2)$.

Of course, the foregoing reasoning implicitly rests the assumption that o_1 obtaining under any state in E_1 is exactly as valuable for the subject as o_1 obtaining under any state in E_2, and likewise for o_2 in $\neg E_1$ and o_2 in $\neg E_2$. However, suppose that the following scenario occurs:

(a) S considers E_1 to be exactly as likely as E_2, i.e., $E_1 \sim^b E_2$
(b) S prefers the constant act $\underline{o}_1$ to the constant act $\underline{o}_2$
(c) S is generally indifferent between o_2 given $\neg E_1$ and o_2 given $\neg E_2$
(d) S finds o_1 substantially more desirable on average if it obtains in one of the states in E_1 than if it obtains in one of the states in E_2

Such a situation seems coherent; yet, presumably, the rational choice for S in this kind of case would be to prefer $(E_1, o_1 | \neg E_1, o_2)$ to $(E_2, o_1 | \neg E_2, o_2)$, despite the fact that $E_1 \sim^b E_2$. Although both acts have an equal subjective likelihood of resulting in o_1 and o_2, for the former the outcome o_1 is much more desirable to S because it obtains in the right kinds of states. If o_1 can have a different subjective value for the agent if it obtains in any of the states in E_1 than it does if it obtains in any of the states in E_2, and similarly for o_2, then the justification for Coherence falls apart.[16]

Thus, it is frequently noted in the literature that Savage's theorem requires that outcomes are *state neutral*, where an outcome o is *state neutral* (relative to an agent S and specification of states $\mathcal{S}$) just in case S's utility for o does not depend on the state $s \in \mathcal{S}$ in which it's realised. However, simply requiring state neutrality is not *quite* enough to fully justify Coherence, which requires that the choice between $(E_1, o_1 | \neg E_1, o_2)$ and $(E_2, o_1 | \neg E_2, o_2)$ depends *solely* on the (presumed constant) values for o_1 and o_2, and the relative likelihoods of E_1 and E_2. To begin with, note that state neutrality does not yet rule out that the utility of an outcome may depend upon the *act* which gave rise to it. Thus, something stronger than state neutrality is needed, which I will call *context neutrality*:

Definition 7: Context neutrality
An outcome o is *context neutral* (relative to an agent S and a choice of $\mathcal{S}$ and $\mathcal{A}'$) iff S's utility for o depends neither on the state $s \in \mathcal{S}$ in which it's realised nor on the act $\alpha \in \mathcal{A}'$ from which it originates

Even the assumption of context neutrality is not quite enough, though, for it's conceivable that acts *themselves* could be objects of utility independently of their potential consequences. Thus Savage is forced to make an assumption about how agents value *acts*; namely, that they have no *intrinsic* preferences between acts, or preferences which don't depend upon the possible outcomes that the act might have. Without this assumption, it could be the case that the subject prefers o_1 to o_2, finds E_1 more likely than E_2, yet has such a strong *intrinsic* distaste for the act represented by $(E_1, o_1 | \neg E_1, o_2)$ that she is disposed to prefer $(E_2, o_1 | \neg E_2, o_2)$ instead *despite* its having the smaller likelihood of resulting in the best outcome.

Without these two assumptions, Savage's system becomes highly implausible, both descriptively and normatively. A natural thought here is that if agents care about the specific acts

[16] The same can be said for the definition of null events, and for the conditions **SAV3**, **SAV4** and **SAV5**, all discussed below.

they perform, then that such-and-such an act was performed can be built into the description of the outcomes that obtain. Indeed, the most straightforward way to ensure the aforementioned requirements hold is to treat outcomes as conjunctions of states and acts. If outcomes are characterised in this way, then context neutrality is ensured and we don't need to assume that agents have no intrinsic preferences for acts. However, this move does not sit well with other aspects of Savage's system (Joyce 1999, p. 56). Note, first of all, that since every outcome gets paired with every state by at least one act-function, and assuming that every act-function represents an act in $\mathcal{A}'$, it follows that states must be *outcome-independent* in the following sense:[17]

> **Definition 8: Outcome-independence**
> A state s is *outcome-independent* (with respect to a specification of outcomes, $\mathcal{O}$) iff s is logically consistent with any outcome $o \in \mathcal{O}$

For example, an outcome o cannot imply that a particular state s does *not* obtain, since (it is assumed that) there is some act the agent could perform which would bring about o if s were to be the case. Secondly, since every outcome is in the range of multiple act-functions, no outcome can imply that a *particular* act was chosen (though every outcome will imply that some range of acts was *not* chosen).

Thus, if the descriptions in $\mathcal{O}$ are intended to specify the various things the decision-maker may care about, the implication here is that the decision-maker has no intrinsic interest in what act she performs. (This is, of course, also in the background of Savage's assertion that two acts with the same outcomes at all states are not worthy of being distinguished.) Roughly put, Savage assumes that, from the decision-maker's perspective, *only potential outcomes matter*: the final decision model is one where the choice between acts depends wholly upon the credence-weighted utility of the outcomes; utilities for states and for acts *themselves* don't figure in the representation, which has a utility function defined only for the relatively limited set of propositions $\mathcal{O}$.

A number of authors have objected to the assumption of state neutrality—and by extension, context neutrality. (See, for instance, Karni, Schmeidler *et al.* 1983, Schervish, Seidenfeld *et al.* 1990, Bradley 2001.) I will not go over those complaints here; though I will say that *if* context neutrality is to be considered problematic, then this can only be because it is in tension with other parts of Savage's system—context neutrality by itself seems hardly problematic. Context neutrality forces outcomes to be rather fine-grained, and it is because of this that the problem of constant acts exists (see §3.1 for a more in-depth explanation of this point). To see this, consider an example of Jamie Dreier's:

> I would rather have money as a gift from Boris than money stolen from Boris. The two outcomes must be distinguished. No one could plausibly accuse me of having intransitive preferences on the grounds that I preferred \$100 as a gift from Boris to \$5 as a gift from Boris, and \$5 as a gift from Boris to \$100 stolen from Boris. (1996, p. 257)

[17] As with act-independence, events will inherit their own form of outcome-independence from states.

Here, Dreier is highlighting the important distinction between characterising outcomes in a coarse-grained way,

o_3 = *obtain \$100 from Boris*

And characterising them in a relatively fine-grained way,

o_1 = *obtain \$100 as a gift from Boris*
o_2 = *obtain \$100 stolen from Boris*

Most would value o_1 over o_2. However, an act whose outcome could be coarsely described as simply o_3 may actually have outcomes manifest in particular as either o_1 or o_2, depending on the state of the world in which it's performed. Likewise, two distinct acts which both result in o_3 given at a particular state may, more specifically, result in o_1 on the one hand or o_2 on the other. As the example highlights, the coarse-grained description of outcomes does not sit well with the presumption of context neutrality: the value of an outcome depends on the context in which it obtains. But the upshot of all this is that the more context that is built into the specification of the outcome, the less its value will depends on outside factors; in the limit, everything of potential relevance will be captured in the specification and context neutrality assured. The example does suggest, however, that context neutrality is plausible only insofar as the outcomes in $\mathcal{O}$ are specified in rather great detail.[18]

The following summarises the essential points to keep in mind:

(1) $\mathcal{A}'$ = {α, β, γ, …} is a set of mutually exclusive *acts*, such that *S* must perform exactly one of its members. Every act should also be such that the decision-maker is certain that she would perform the act, if she were to so choose. It is assumed that agents have no intrinsic preferences between acts.
(2) $\mathcal{O}$ = {o_1, o_2, o_3, …} is a set of *outcomes*; that is, a set of mutually exclusive and jointly exhaustive propositions about the consequences of performing an act at a state. For Savage's system to have a plausible interpretation *qua* decision theory, then the outcomes in $\mathcal{O}$ must be context-neutral and thus very fine-grained, and they cannot imply that a particular act was chosen or that a particular state obtains.
(3) $\mathcal{S}$ = {s_1, s_2, s_3, …} is a set of *states*; that is, a set of mutually exclusive and jointly exhaustive propositions. For Savage's system to have a plausible interpretation *qua* decision theory, the states in $\mathcal{S}$ must be act-independent in either the causal or evidential sense, and therefore logically independent of what acts are performed; they must also be outcome-functional. Together with the assumption that $\mathcal{A} = \mathcal{O}^{\mathcal{S}}$, the foregoing implies that states are outcome-independent.
(4) $\mathcal{E}$ = {E_1, E_2, E_3, …} is a set of *events*; that is, (effectively) a set of propositions equivalent to disjunctions of states. Events inherit act-independence and outcome-independence properties from the states they are constructed out of.

[18] Note that given outcome-functionality, context neutrality then implies that $\mathcal{S}$ must be correspondingly fine-grained.

(5) $\mathcal{A} = \{\mathcal{F}, \mathcal{G}, \mathcal{H}, \ldots\}$ is the set of all *act-functions*; that is, the set of all total functions from $\mathcal{S}$ to $\mathcal{O}$. Such functions are *intended* to represent acts in $\mathcal{A}'$, by specifying the act's outcomes under different states.

(6) $\succcurlyeq$ is primitively defined on $\mathcal{A}$, and usually given an interpretation based on choice-dispositions.

2.3 Savage's theorem

With all this in mind, we can now outline Savage's theorem and the structure of its proof. The theorem has seven preference conditions in the original formulation, though I will follow Joyce (1999) in explicitly listing the purely structural assumption that Savage needs to make about $\mathcal{A}$:

SAV0 $\mathcal{A} = \mathcal{O}^{\mathcal{S}}$

It's possible to weaken **SAV0** (and drop Savage's seventh preference axiom, **SAV7**) if we only desire the representation to hold for finitely-valued act-functions. In what follows, let $\mathcal{F}_E$ refer to the restriction of $\mathcal{F}$ to E. (Thus $\underline{o}_E$ is the restriction of $\underline{o}$ to E.) Furthermore, the *mixture* of $\mathcal{F}$ and $\mathcal{G}$, $\mathcal{F}_E \cup \mathcal{G}_{\neg E}$, is an act-function $\mathcal{H}$ such that $\mathcal{H}(s) = \mathcal{F}(s)$ for all $s \in E$, and $\mathcal{H}(s) = \mathcal{G}(s)$ for all $s \notin E$. We can now state the weakened act-richness assumption as follows:

SAV0' $\mathcal{A}$ is the set of all finite-valued functions from $\mathcal{S}$ to $\mathcal{O}$; i.e., for any outcome $o \in \mathcal{O}$, $\underline{o} \in \mathcal{A}$, and for all $\mathcal{F}, \mathcal{G} \in \mathcal{A}$, and any $E \in \mathcal{E}$, $\mathcal{F}_E \cup \mathcal{G}_{\neg E} \in \mathcal{A}$

SAV0' says that $\mathcal{A}$ contains not only all constant act-functions, but also all act-functions that can be constructed therefrom *via* a finite number of mixings. Note that, although $\mathcal{O}$ may contain an infinite number of outcomes, each act-function in $\mathcal{A}$ is only ever associated with a finite number of outcomes.

The first two real preference conditions are straightforward weak order and non-triviality requirements on $\succcurlyeq$:

SAV1 $\succcurlyeq$ on $\mathcal{A}$ is complete and transitive

SAV2 $\underline{o}_i \succ \underline{o}_j$ for some $o_i, o_j \in \mathcal{O}$

The transitivity of $\succcurlyeq$ is an obvious necessary condition for the kind of representation that Savage aims to achieve, whereas the completeness of $\succcurlyeq$ is required for Savage's strong uniqueness result (amongst other things). **SAV2** is a simple non-triviality condition.

The remaining preference conditions require a bit of work to spell out. We first extend $\succcurlyeq$ to restricted act-functions:

Definition 9: $\succcurlyeq$ for restricted act-functions
$\mathcal{F}_E \succcurlyeq \mathcal{G}_E$ iff $\mathcal{F}^* \succcurlyeq \mathcal{G}^*$ whenever $\mathcal{F}_E = \mathcal{F}^*_E$, $\mathcal{G}_E = \mathcal{G}^*_E$, and $\mathcal{F}^*_{\neg E} = \mathcal{G}^*_{\neg E}$

Furthermore, define the set of *null events*, $\mathcal{N}$, as:

Definition 10: Null events
$\mathcal{N} = \{E \in \mathcal{E}: \mathcal{F} \sim \mathcal{G}$ whenever $\mathcal{F}_{\neg E} = \mathcal{G}_{\neg E}\}$

The members of $\mathcal{N}$ are the events which will receive a $\mathcal{B}el$ value of 0 in the final representation. Again, the idea behind this is highly intuitive: if any two act-functions are considered equivalent for the purposes of decision-making whenever they only differ in their outcomes with respect to states $s \in E$ for some event E, then what happens in those states must be considered utterly irrelevant from the point of view of the decision-maker. Assuming basic rationality, this would come to pass just in case the subject had zero confidence in one of those states obtaining. The background assumption, of course, is that agents have no interest the outcomes of their acts at states they consider utterly unlikely to be true.[19]

Savage's next two preference conditions express his so-called *sure-thing principle*. For all $\mathcal{F}, \mathcal{G}, \mathcal{F}^*, \mathcal{G}^* \in \mathcal{A}$, $E \in \mathcal{E}$, and $o_1, o_2 \in \mathcal{O}$,

SAV3 If $\mathcal{F}_E = \mathcal{G}_E$, $\mathcal{F}^*_E = \mathcal{G}^*_E$, $\mathcal{F}_{\neg E} = \mathcal{F}^*_{\neg E}$, and $\mathcal{G}_{\neg E} = \mathcal{G}^*_{\neg E}$, then $\mathcal{F} \succ \mathcal{F}^*$ iff $\mathcal{G} \succ \mathcal{G}^*$

SAV4 If $E \in \mathcal{E} - \mathcal{N}$, then $\underline{o}_E \succ \underline{o}^*_E$ iff $\underline{o} \succ \underline{o}^*$

Savage's famous example of his (controversial) principle goes as follows:

> A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he knew that the Republican candidate were going to win, and again finds that he would. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, as we would ordinarily say. It is all too seldom that a decision can be arrived at on the basis of this principle, but except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance. (1954, pp. 21-2)

More specifically, **SAV3** says that whether $\succcurlyeq$ holds between two act-functions does not depend on those states which have identical consequences for the two acts. This seems plausible for any rational agent, given the assumption that the states are act-independent. **SAV4**, on the other hand, sets up a correspondence between outcome preferences (i.e., preferences over constant act-functions) and restricted act-function preferences for non-null events.

The next condition is especially important for the sensibility of Coherence. Say that $\mathcal{F}_E \equiv o$ iff $\mathcal{F}_E(s) = o$ for all s in E. Then, for all relevant acts and events,

[19] I will not delve into the plausibility of this assumption here, though I will note that it is not obviously true. See (Bradley and Stefansson forthcoming) for related discussion.

SAV5 If $\underline{o}_1 \succ \underline{o}_2$, $\mathcal{F}_E \equiv o_1$, $\mathcal{F}_{\neg E} \equiv o_2$, $\mathcal{G}_{E^*} \equiv o_1$, $\mathcal{G}_{\neg E^*} \equiv o_2$, and similarly for $\underline{o}^+_1$, $\underline{o}^+_2$, $\mathcal{F}^+$, $\mathcal{G}^+$, then $\mathcal{F} \succ \mathcal{G}$ iff $\mathcal{F}^+ \succ \mathcal{G}^+$

This condition, in conjunction with Coherence, ensures that $\succcurlyeq^b$ is a weak order on $\mathcal{E}$. To recall, Coherence tells us that a subject finds E_1 strictly more probable than E_2 just in case, for any pair of outcomes o_1 and o_2, whenever she prefers o_1 to o_2, she prefers the act-function $(E_1, o_1 \mid \neg E_1, o_2)$ over $(E_2, o_1 \mid \neg E_2, o_2)$. We interpret this as the one act having a higher subjective likelihood of resulting in the better outcome, and a lower likelihood of resulting in the worse outcome. **SAV5** says that *any* time a subject prefers $(E_1, o_1 \mid \neg E_1, o_2)$ to $(E_2, o_1 \mid \neg E_2, o_2)$ for *some* o_1, o_2 such that $o_1 \succ o_2$, then for *all* pairs of outcomes o_3, o_4 such that $o_3 \succ o_4$, the agent will prefer $(E_1, o_3 \mid \neg E_1, o_4)$ to $(E_2, o_3 \mid \neg E_2, o_4)$. In light of how we are interpreting the agent's behaviour, **SAV5** can be read as a basic condition of coherent decision-making upon an agent: if, in one instance, she is disposed to choose as if she considers E_1 more likely than E_2, then she ought to choose as such in all instances. Without this condition, a subject's preferences may fail to determine any well-defined qualitative probability relation at all, rendering Coherence effectively useless as a definition.

Savage's final two preference conditions are that, for all $\mathcal{F}, \mathcal{G} \in \mathcal{A}$ and $E \in \mathcal{E}$,

SAV6 If $\mathcal{F} \succ \mathcal{G}$ then there is a finite partition $\boldsymbol{T}$ of $\mathcal{S}$ such that for all $E \in \boldsymbol{T}$, $\mathcal{F}^*_E \equiv o_1$ and $\mathcal{F}^*_{\neg E} = \mathcal{F}_{\neg E}$ only if $\mathcal{F}^* \succ \mathcal{G}$; and $\mathcal{G}^*_E \equiv o_1$ and $\mathcal{G}^*_{\neg E} = \mathcal{F}^*_{\neg E}$ only if $\mathcal{F} \succ \mathcal{G}^*$

SAV7 If $\mathcal{H}_E \equiv \mathcal{G}(s)$, then $\mathcal{F}_E \succ \mathcal{H}_E$ only if $\mathcal{F}_E \succcurlyeq \mathcal{G}_E$; and $\mathcal{H}_E \succ \mathcal{F}_E$ only if $\mathcal{G}_E \succcurlyeq \mathcal{F}_E$

SAV6 is a very strong structural condition which in effect requires that no outcome is either infinitely desirable or infinitely undesirable. In conjunction with the other preference conditions, it plays an important role in the derivation of a probability function $\mathcal{B}el$ that represents $\succcurlyeq^b$. **SAV7** is also very strong, but as noted above, it's not required if we limit our attention to finitely-valued act-functions.

With these conditions set out, Savage proves the following theorem:[20]

Theorem 1: Savage's theorem

If **SAV0–SAV7** hold of $<\mathcal{S}, \mathcal{E}, \mathcal{N}, \mathcal{O}, \mathcal{A}, \succcurlyeq>$, then there is a probability function $\mathcal{B}el: \mathcal{E} \mapsto [0, 1]$, and a function $\mathcal{D}es: \mathcal{O} \mapsto \mathbb{R}$, such that for all $o_1, o_2 \in \mathcal{O}$, all $E, E_1, E_2 \in \mathcal{E}$, and all $(E_i, o_i \mid \ldots \mid E_n, o_n), (E_j, o_j \mid \ldots \mid E_m, o_m) \in \mathcal{A}$,

(i) $\underline{o}_1 \succcurlyeq \underline{o}_2$ iff $\mathcal{D}es(o_1) \geq \mathcal{D}es(o_2)$

(ii) $E_1 \succcurlyeq^b E_2$ iff $\mathcal{B}el(E_1) \geq \mathcal{B}el(E_2)$

(iii) If $0 < \lambda < 1$, then $\mathcal{B}el(E_1) = \lambda.\mathcal{B}el(E)$, for some $E_1 \subseteq E$

(iv) $(E_i, o_i \mid \ldots \mid E_n, o_n) \succcurlyeq (E_j, o_j \mid \ldots \mid E_m, o_m)$ iff $\sum_i^n \mathcal{B}el(E_i).\mathcal{D}es(o_i) \geq \sum_j^m \mathcal{B}el(E_j).\mathcal{D}es(o_j)$

Furthermore, $\mathcal{B}el$ is unique and $\mathcal{D}es$ is bounded and unique up to positive linear transformation

[20] This statement of Savage's theorem has been slightly modified for purposes of readability. For a more precise formulation, see (Fishburn 1970).

A thorough statement of the proof of Theorem 1 can be found in (Fishburn 1970, Ch. 14).

The strong statement of Savage's uniqueness condition, while technically accurate, is somewhat misleading. Savage *does* prove that, *given a choice of* $\mathcal{O}$ and $\mathcal{S}$, if **SAV0–SAV7** are satisfied then $\succcurlyeq$ can be given an expected utility representation where $\mathcal{B}el$ on $\mathcal{E}$ is unique and $\mathcal{D}es$ on $\mathcal{O}$ is unique up to positive linear transformation. The strength of this uniqueness condition is often considered a substantial point in favour of Savage's theorem. It is *prima facie* valuable to have a theorem which supplies us with a unique credence function. The problem here is that both $\mathcal{B}el$ and $\mathcal{D}es$ have their uniqueness conditions only relative to the choice of $\mathcal{S}$ and $\mathcal{O}$. This much is obvious for $\mathcal{D}es$, as it is a function defined on $\mathcal{O}$ and so necessarily changes its character whenever $\mathcal{O}$ is altered. But, as Schervish, Seidenfeld *et al*. (1990) show, the character of $\mathcal{B}el$ is also strongly dependent *not only* on how $\mathcal{S}$ (and hence $\mathcal{E}$) is specified, but also on how $\mathcal{O}$ is specified: if it turns out that there are multiple, equally viable ways of characterising the space of states and outcomes, then Savage's strong uniqueness results are to some illusory (see also Levi 2000, p. 399).

A huge number of decision-theoretic representation theorems are formulated within a framework very similar to Savage's own. As Krantz *et al*. put it in their monumental *Foundations of Measurement*,

> In general, a rough sort of consensus exists about the primitive terms to be employed in the formulation of the problem of decision making under risk or uncertainty. Nearly everyone seems to agree that there are chance events to which probabilities adhere, consequences which exhibit utilities, and decisions that are more or less arbitrary associations of consequences to events. (1971, p. 411)

That is to say, a great many representation theorems (then and today) begin with $\mathcal{S}$ and $\mathcal{O}$, and define $\succcurlyeq$ on a collection $\mathcal{A}$ of act-functions. Most theorists working within the paradigm Savage created define $\mathcal{A}$ as the set of *all* total functions from $\mathcal{S}$ to $\mathcal{O}$. Others have taken $\succcurlyeq$ to be defined on only a proper subset of $\mathcal{O}^{\mathcal{S}}$ (e.g., Richter 1975, Wakker and Zank 1999, Casadesus-Masanell, Klibanoff *et al.* 2000), or on partial functions from $\mathcal{S}$ to $\mathcal{O}$ (e.g., Luce and Krantz 1971, Luce 1972, Roberts 1974, Narens 1976).[21]

Importantly, these theorems include not only those for classical expected utility theory, but a very wide range of non-expected utility theories as well. Indeed, the vast majority of NCU theorems belong to the Savage paradigm. Savage's own theorem, as a (or *the*) classical expected utility theorem, is limited to probabilistic $\mathcal{B}el$ functions. On the other hand, the huge variety of representation styles that can and have been arrived at through the use of Savage's framework—many of which allow for *non*-probabilistic $\mathcal{B}el$ functions—should be encouraging to proponents of characterisational representationism. Unfortunately, though, there are a number of issues that arise from the use of the framework itself, to which we now turn.

[21] Suppes (1969) and Fishburn (1967) diverge from the general trend by characterising their basic objects of preference as ordered pairs of Savage-style act-functions (i.e., the option space is a subset of $\mathcal{O}^{\mathcal{S}} \times \mathcal{O}^{\mathcal{S}}$), which are supposed to represent even-chance bets with the performances of different acts as prizes. The theorem of (Kochov 2015) has a rather unique formal structure, but its basic relata for $\succcurlyeq$ can be accurately described as "multiperiod counterparts of Savage act[-function]s" (p. 240).

3. Constant act-functions and imaginary acts

I will begin my critical discussion with what is easily the most frequently cited objection to Savage's system, which Fishburn (1981) calls the *constant acts problem*: it's implausible that constant act-functions can serve to represent anything that an ordinary agent could choose to *do*. If $\mathcal{A}$ is supposed to represent the space of acts available to the agent in her current situation, then constant act-functions are an anomaly—functions which represent nothing in the real world that the agent could have preferences between.

Fishburn gives the following illustration of the problem. Let the outcome o be *Carrying an umbrella on a bright and sunny day*, and the event E be *It rains*. Then, every s in E is a state in which it rains, and any act-function which maps an s in E to o is pairing an outcome with a state that is inconsistent with it. "In fact, the natural set of [outcomes] that could occur under one state may be disjoint from the set that could occur under another state" (1981, p. 162). Note that, on this way of describing the issue, the problem appears to be that constant act-functions may pair outcomes with incompatible states, thus apparently representing acts which are literally impossible to perform.[22] If s and o are logically inconsistent, then not even an omnipotent god could make it the case that s and o. Suppes and Luce (1965, p. 299), Karni (1993), and Maher (1993, pp. 182-5) give a similar account of the constant acts problem as involving inconsistent state and outcome pairings.[23]

However, the issues here are somewhat more subtle than they are often made out to be. Constant act-functions do give rise to difficulties for characterisational representationism, but exactly what these difficulties may be depends on how we interpret the relevant formalisms. Let us therefore look again in depth at the origins of the constant acts problem, before we turn to how the problem might be dealt with.

3.1 The basis of the problem

The complaint about constant act-functions is usually levelled at **SAV0**, or its weaker counterpart **SAV0'**, wherein the character of $\mathcal{A}$ is formally specified. However, we must be careful not to lay all the blame on Savage's act-richness assumption—it is part of the problem, of course, but it's not the whole story. In fact, there are three independent factors which together lead to the constant acts problem, as I will now argue.

If taken purely on their own, **SAV0** and **SAV0'** are hardly problematic—each merely characterises $\mathcal{A}$ as a subset of $\mathcal{O}^{\mathcal{S}}$. What **SAV0**/**SAV0'** can be taken to require *in context* therefore depends on how the states in $\mathcal{S}$ are characterised, how the outcomes in $\mathcal{O}$ are characterised, and what the act-functions in $\mathcal{A}$ are intended to represent. Let us begin the interpretation of $\mathcal{A}$. As in §2, we will assume that every act-function is assumed to correspond to something an agent might *do*. Let us call this the *Act–Function Correspondence* assumption, which can be stated as follows:

[22] Indeed, if we make our outcomes so fine-grained that each outcome entails a conjunction of the form (*s obtains and* α *was performed*), as some are wont to do, then *every* finite-valued act-function in Savage's system will pair at least one outcome with an incompatible state.

[23] Joyce (1999, pp. 107-8) also supposes that **SAV0** implies the existence of act-functions which pair together incompatible states and outcomes, but interprets the constant acts problem as arising primarily from the conjunction of the completeness requirement (entailed by **SAV1**) and **SAV0**. This is because he drops the behavioural interpretation of $\succcurlyeq$ for another interpretation compatible with preferences over non-existent acts. See §3.3.

Act–Function Correspondence
There exists a natural, one-one correspondence between the set of act-functions $\mathcal{A} \subseteq \mathcal{O}^{\mathcal{S}}$ and the space of available acts $\mathcal{A}'$ such that every $(E_1, o_1 \,|\, \ldots \,|\, E_n, o_n) \in \mathcal{A}$ represents a unique act (or set of acts with the same pattern of consequences) in $\mathcal{A}'$ which, if performed, would result in o_1, if any $s \in E_1$ were the case, …, and o_n if any $s \in E_n$ were the case

As we've seen, Act–Function Correspondence requires that states are at least *logically* act-independent, and outcome-functional; if states did not have these properties, the representation of acts using act-functions would make little sense.

SAV0/**SAV0'** and Act–Function Correspondence are not yet enough to get us a problem—we still need to specify the nature of the outcomes. To see this, note that it's consistent with Savage's formalism that the outcomes in $\mathcal{O}$ are very coarse-grained. Suppose, then, that $\mathcal{O}$ contains only two extremely non-specific outcomes, o_1 and o_2. For instance, let o_1 and o_2 be very long, mutually exclusive disjunctions of the more specific states of affairs that we would ordinarily consider the outcomes of a decision to be. In this case, there does not appear to be anything unusual about constant act-functions: $\underline{o}_1$ and $\underline{o}_2$ could be construed simply as acts (or a collection of acts) which result in one or another disjunct becoming true—and such 'acts' are ubiquitous. The problem with this, of course, is that characterising $\mathcal{O}$ this way conflicts with the informal requirement of context neutrality—without which Savage's preference conditions and his principle of Coherence become highly implausible. For similar reasons, we can assume that any *useful* representation of acts as functions from $\mathcal{S}$ to $\mathcal{O}$ should make use of rather fine-grained outcomes.

We now have enough for the constant acts problem to arise. Generally speaking, there is a deep tension within Savage's system between the following triad:

(1) **SAV0**/**SAV0'**
(2) Act–Function Correspondence
(3) Fine-grained outcomes

A theorist could reasonably pick any two of these to adopt, but trying to justify all three at once is difficult. Let us assume (3) in all that follows. In this case, the constant acts problem becomes clear: **SAV0**/**SAV0'** implies that $\mathcal{A}$ has a particular kind of formal structure; Act–Function Correspondence in turn requires that $\mathcal{A}'$ must have the same structure. The existence of constant act-functions in $\mathcal{A}$, however, seems incompatible with Act–Function Correspondence. One of these needs to go.

There are two lessons that I wish to draw here. The first is that it is *slightly* misleading to express the problem as being about the *compatibility* of some states and outcomes. There would still be cause to worry about Act–Function Correspondence *even if* there were no mutually incompatible pairs of states and outcomes, and the problematic act-functions are by no means limited only to those which pair together incompatible states and outcomes. On *any* natural conception of acts and outcomes, immensely implausible that there is an act we can perform such that, *regardless of how the world turns out to be independently of our decision*, one and only one fine-grained outcome will obtain. Now, this *may* be because the set of potential outcomes $\mathcal{O}_1, \mathcal{O}_2 \subseteq \mathcal{O}$ that may result from any available act at two distinct states s_1 and s_2 respectively only partially overlap, if they overlap at all—indeed, this would seem to be the so in any ordinary case: some states just don't play nicely with some outcomes. However, even supposing that *every* state is consistent with the same range of outcomes, there would still be no good reason to think that $\mathcal{A}'$ has the kind of structure imposed upon it by the conjunction of **SAV0**/**SAV0'** and Act–Function Correspondence. Which outcomes can arise in which states depends on the range of acts available to the agent at the time of the decision, and **SAV0**/**SAV0'** places rather implausible constraints on what that range of acts must always look like. *The* problem, therefore, is not simply that:

> In virtually any realistic problem that is formulated in the Savage mode, some consequences will be incompatible with some states or events, as is "carry an umbrella on a bright, sunny day" with "rain". (Fishburn 1981, p. 162)

Rather, the problem is the unjustified and implausible imposition of a particular structure upon $\mathcal{A}'$.[24]

The second thing to note is that constant act-functions are only a very small part of a much broader problem. For example, essentially the same worries that arise for constant act-functions can be raised for what we might call bifurcating act-functions, or act-functions of the form $(E, o_1 \mid \neg E, o_2)$, which recall are crucially important for the definition of $\succcurlyeq^b$; and likewise for trifurcating act-functions $(E_1, o_1 \mid E_2, o_2 \mid E_3, o_3)$, and so on. Most (if not all) act-functions which range over only a small number of distinct finely-individuated outcomes will be just as problematic as constant act-functions, and for essentially the same reasons. I will refer to any act-function which lacks a corresponding act in $\mathcal{A}'$ as an *imaginary act-function*.[25] *Any* imaginary act-function causes as much trouble for Savage as a constant act-function does—at least to the extent that constant act-functions cause troubles at all. At best, the constant functions are simply the most salient example of the underlying issue.

If one wants to avoid the bigger issues at the heart of the constant acts problem, then it is clear that one must do much more than just remove constant act-functions from $\mathcal{A}$. The presence of *imaginary* act-functions in $\mathcal{A}$ is problematic *inasmuch* as $\mathcal{A}$ is supposed to represent $\mathcal{A}'$. This seems to leave us with only two options. On the one hand, one might retain Act–Function Correspondence and try to develop a theorem around a more realistic representation of $\mathcal{A}'$. On the other hand, one could drop Act–Function Correspondence, offering instead an

[24] In Fishburn's example, *It rains* is an event—but given an outcome set $\mathcal{O}$ that includes *Carrying an umbrella on a bright and sunny day*, there cannot be any such event *in* $\mathcal{E}$. As noted in §2.2, states must be act-independent, outcome-functional, and thus, in light of **SAV0**/**SAV0'** and Act–Function Correspondence, events must be outcome-independent. Of course, rain could still *occur*—the point is that there can be no event *in* $\mathcal{E}$ which corresponds to that proposition *if Carrying an umbrella on a bright and sunny day* already exists in $\mathcal{O}$. To apply Savage's system, we are not free to pick and choose as we like our states, outcomes, and events, but must do so within tightly constrained limits. As I will argue below, this fact itself leads to further problems with Savage's framework.

[25] Maher (1993, p. 183) refers to these as *uninterpretable acts*.

alternative interpretation of the system which somehow makes sense of imaginary act-functions. In the remainder of this section, I will consider the viability of each of these options in turn.

3.2 Doing without imaginary act-functions

Given Act–Function Correspondence, **SAV0** and even the weaker **SAV0'** are clearly too strong. The problem here is not just that *ordinary* agents could not have preferences satisfying the conditions, but rather that it would not even make sense to assert of *anyone* that their preferences satisfy the conditions. To say that these act-richness assumptions are *false* is to say that $\succcurlyeq$ is formally required to have a domain which it *does not*, and in fact *cannot*, have.

Some have thought to respond to the problem of constant acts by weakening those act-richness assumptions. As noted earlier, Luce and Krantz (1971) were able to obtain a representation result without requiring the use of constant act-functions, which they consider an important benefit of their approach.[26] However, we have seen that simply removing constant act-functions from $\mathcal{A}$ is inadequate as a response to the broader problem with imaginary act-functions. Luce and Krantz retain still very strong assumptions about the structure of their set of act-functions, which by their own admission seem to imply the presence of imaginary act-functions. This is the basis of Joyce's (1999, pp. 108-10) critique of Luce and Krantz's theorem, and I will not add anything further to it here.[27]

There is a general reason for this failure: like Savage, Luce and Krantz attempt to formally construct their set of act-functions $\mathcal{A}$ using just $\mathcal{S}$ and $\mathcal{O}$ but independently of any knowledge or specifications regarding the space of available acts $\mathcal{A}'$. It is unreasonable to begin with an arbitrary partition $\mathcal{S}$ and an equally arbitrary set of outcomes $\mathcal{O}$, and expect to work backwards from there to arrive at a plausible reconstruction of the space of available acts. $\mathcal{A}'$ may correspond to a proper subset of some collection of act-functions (defined for *some* ways of construing $\mathcal{S}$ and $\mathcal{O}$), but the *formal* character of this subset will depend heavily on the nature of $\mathcal{A}'$ itself. There may, for instance, be one available act $(E, o_1 | \neg E, o_2)$ but no $(E, o_2 | \neg E, o_1)$, or vice versa—but there is no way to know this, if all that is given is $\mathcal{S}$ and $\mathcal{O}$. If Act–Function Correspondence is ever to be justified, the *formal* construction of the space of act-functions needs to begin with $\mathcal{A}'$.[28]

[26] See also (Gaifman and Liu MS) for a recent attempt at *minimising*—but not altogether removing—the use of constant act-functions within a Savagean framework. Gaifman and Liu's theorem requires that there are at least two constant act-functions. Although much weaker than **SAV0**, it's not at all clear that their replacement condition (or the more general assumptions they need to make about the structure of their set of act-functions) is consistent with Act–Function Correspondence. Again: simply removing some problem cases from the picture is not equivalent to removing the problem *simpliciter*.

[27] A further problem with Luce and Krantz's formalisation is that many of their act-functions are very difficult to interpret *as* acts (or anything else in the vicinity). See (Krantz and Luce 1974), (Spohn 1977), and (Fishburn 1981) for discussion.

[28] I am unaware of any Savage-like theorems which take the path I am suggesting, though it is briefly discussed by Fishburn (1970, pp. 164-7). Balch and Fishburn (1974, see also Balch 1974, Fishburn 1974) develop a theorem which begins with a primitive set of acts $\mathcal{A}'$ and a set of act-independent states $\mathcal{S}$, with outcomes defined as act-event pairs. Their theorem belongs to the class of lottery-based theorems, which I critique in a different work.

On the flip side, however, as I will now argue, it is very difficult (if not impossible) to develop a Savage-like representation theorem *without* making some rather strong, and ultimately implausible, assumptions about $\mathcal{A}$. There are multiple reasons for this, though in what follows I will focus upon what appears to me the most troubling: the difficulty in developing well-defined orderings on $\mathcal{E}$ and $\mathcal{O}$, needed to construct $\mathcal{B}el$ and $\mathcal{D}es$ respectively, without an appeal to imaginary act-functions.

Fishburn has argued that, without appealing to constant act-functions, "there is no natural way of defining preferences on [outcomes] in terms of preferences on acts" (1970, 166). In Savage's system, however, preferences over constant act-functions form a crucial part of constructing the $\mathcal{D}es$ function—recall that, in his representation,

$$\underline{o}_1 \succcurlyeq \underline{o}_2 \text{ iff } \mathcal{D}es(o_1) \geq \mathcal{D}es(o_2)$$

Thus, Fishburn suggests that to do without constant act-functions, a theorist would need to develop a dual-primitive theorem, with $\succcurlyeq$ defined on $\mathcal{A}$ and a separate preference relation $\succcurlyeq^u$ defined on $\mathcal{O}$. As it turns out, though, with some imagination it is possible to characterise relative utilities between outcomes in terms of preferences between act-functions without appealing to constant act-functions at all. It will be instructive to see why this alternative characterisation still seems to end up requiring an appeal to imaginary act-functions.

The basic idea here is dominance reasoning: an outcome o_1 is more desirable than another outcome o_2 for an agent S iff $\mathcal{F} \succ \mathcal{G}$, when $\mathcal{F}$ and $\mathcal{G}$ only differ, *with respect to the states that S gives some credence to*, in that $\mathcal{F}$ is sometimes paired with o_1 at some states while $\mathcal{G}$ is paired with o_2 at those same states. In this case, with respect to what the agent considers possible, $\mathcal{F}$ represents an act which is identical to the act represented by $\mathcal{G}$ but for the possibility of resulting in o_1 instead of o_2 at some states—and if $\mathcal{F} \succ \mathcal{G}$, this is presumably then because o_1 is preferred to o_2.

In order to spell this idea out formally, we will first need a notion of *nullity* for states. As a consequence of Definition 10, any subset of a null event is also null, including any singleton events $\{s\}$, for $s \in E \in \mathcal{N}$. Given this, say that a state is *null* iff it belongs to an event E and E is null in the sense of Definition 10; the state is *non-null* otherwise. Now let $\mathcal{S}' \subset \mathcal{S}$ be a set of non-null states. We can now define a relative utility ranking $\succcurlyeq^u$ as follows:

Definition 11: $\succcurlyeq^u$ without constant acts
$o_1 \succcurlyeq^u o_2$ iff $\mathcal{F} \succcurlyeq \mathcal{G}$ whenever, for some set of non-null states $\mathcal{S}'$,
(i) If $s \in \mathcal{S}'$, then $\mathcal{F}(s) = o_1$ and $\mathcal{G}(s) = o_2$
(ii) For all non-null $s \notin \mathcal{S}'$, $\mathcal{F}(s) = \mathcal{G}(s)$

Assuming that outcomes are context neutral, the right-to-left direction of Definition 11 seems plausible for any rational agent—the dominance principle it embodies is one of the most intuitive precepts of folk decision theory. Furthermore, this definition does away with any need for constant act-functions.

However, there seems to be no good reason to think that the space of available acts will have the structure required for the general applicability of Definition 11. There are two distinct issues here.[29] The first arises as a result of the appeal to Definition 10 in the definition of null states. As almost any event in $\mathcal{E}$ can be null, and because we cannot presume to know *a priori* what events the agent considers null or non-null, the general application of Definition 10 already imposes quite strong restrictions upon the character of $\mathcal{A}$. That is, for any *potentially* null event E, Definition 10 requires that we will be able to find at least two act-functions which differ for some state(s) in E but which are identical with respect to all states in $\neg E$. There is no good reason to suppose that such acts will always be available.

Now, perhaps this first issue could be solved using another definition of nullity; or, alternatively, we might even assume that $\mathcal{N}$ is given to us for free as a primitive. This will not be enough, because a closely related issue arises for Definition 11 itself. In particular, in order to ensure that the left-to-right direction always holds for any potential subject S, it will need to be the case that for *every* way of dividing the null states from the non-null there must be act-functions $\mathcal{F}$ and $\mathcal{G}$ which satisfy the stated conditions (i) and (ii) with respect to the relevant outcomes. This is still too strong an assumption, and there is no guarantee that the space of available acts will play along. An obvious example for when Definition 11 cannot be applied (but certainly not the only one) is the case of a fatalist who is certain that whatever outcome may eventually obtain, it will obtain regardless of her choices. At every state, she believes, any of her acts will result in the same outcome, whatever that outcome may be. The fatalist prefers some outcomes over others, and is uncertain about which outcome will obtain, but there will be *no* acts available to her which have different outcomes at any states *she gives credence to*; hence, any act-function which satisfies (i) is imaginary.

Suppose, then, that both $\mathcal{N}$ and $\succcurlyeq^{u}$ are given as primitives, *not* defined in terms of preferences on act-functions. There is now the problem of defining $\succcurlyeq^{b}$, needed to construct the $\mathcal{B}el$ function, without making undue assumptions about the character of $\mathcal{A}'$. Savage's principle of Coherence appeals to bifurcate act-functions, which are usually no more plausible *qua* representations of available acts than constant act-functions. So, an alternative definition for $\succcurlyeq^{b}$ will need to be found as well.

Machina and Schmeidler (1992) present a somewhat more plausible definition of $\succcurlyeq^{b}$ within an essentially Savagean framework, as follows:

Definition 12: $\succcurlyeq^{b}$ (Machina and Schmeidler)
$E_1 \succcurlyeq^{b} E_2$ iff, if $o_1 \succ^{u} o_2$, then $\mathcal{F} \succcurlyeq \mathcal{G}$ whenever:
(i) If $s \in E_1 - E_2$, then $\mathcal{F}(s) = o_1$ and $\mathcal{G}(s) = o_2$
(ii) If $s \in E_2 - E_1$, then $\mathcal{F}(s) = o_2$ and $\mathcal{G}(s) = o_1$
(iii) If $s \notin (E_1 - E_2) \cup (E_2 - E_1)$, then $\mathcal{F}(s) = \mathcal{G}(s)$

The reasoning behind Definition 12 is very similar to the reasoning behind Coherence. Indeed, the two definitions amount to the same thing in the special case where $E_2 = \neg E_1$. If $\mathcal{F}$ and $\mathcal{G}$ satisfy the stated conditions, then the agent would prefer $\mathcal{F}$ to $\mathcal{G}$ iff she found E_1 more likely than E_2, as $\mathcal{F}$ has the greater subjective likelihood of resulting in the better outcome. The major benefit of Machina and Schmeidler's definition is that it does not make use of bifurcate act-functions—in fact, $\mathcal{F}$ and $\mathcal{G}$ may have any number of outcomes. Unfortunately, Machina and

[29] To focus in on the main problem, I will assume for now that $\succcurlyeq$ is complete on $\mathcal{A}$; in §3.4, I will discuss what can be said when that assumption is false.

Schmeidler's alternative still imposes strong constraints on the space of available acts. Before I argue this, however, I will note that it's possible to improve upon their definition in at least three ways.

To begin with, the reasoning which underlies the definition does not require something as strong as condition (ii), which makes mention of the same *outcomes* as appeared in condition (i). It would be enough that the second condition appeals to outcomes with the same *utilities* as those mentioned in (i); and since we have taken $\succcurlyeq^{u}$ as a primitive we can replace (ii) with:

(ii') If $s \in E_2 - E_1$, then $\mathcal{F}(s) = o_4$ and $\mathcal{G}(s) = o_3$, where $o_3 \sim^{u} o_1$, and $o_4 \sim^{u} o_2$

The outcome o_3 may or may not be identical to o_1, and similarly for o_2 and o_4, so (ii') is a strictly weaker condition than (ii). The second improvement is similar: with respect to condition (iii), sameness of outcomes is unnecessary—sameness of utility would be enough. (Strictly, it would be enough that the credence-weighted average of the outcomes under the states $s \notin (E_1 \cup E_2)$ is equal for $\mathcal{F}$ and $\mathcal{G}$, but there is no obvious way to specify such a condition prior to deriving the credence function.) Thus we can replace (iii) with:

(iii') If $s \notin (E_1 - E_2) \cup (E_2 - E_1)$, then $\mathcal{F}(s) \sim^{u} \mathcal{G}(s)$

Finally, it's possible to weaken the definition's requirements on $\boldsymbol{\mathcal{A}}$ if all null events are discounted from consideration. Definition 5.12 applies to *all* pairs of events E_1 and E_2, and so act-functions must be found which satisfy the definitions three conditions with respect to any pair E_1 and E_2. However, null events can be presumed to sit at the bottom of the $\succcurlyeq^{b}$ ranking (to be assigned a credence of 0), so we don't need to consider preferences over act-functions to decide where they sit with respect to $\succcurlyeq^{b}$.

The foregoing then leads to the following, improved definition of $\succcurlyeq^{b}$:

Definition 13: $\succcurlyeq^{b}$ (Machina and Schmeidler improved)
If $E \in \boldsymbol{\mathcal{N}}$, then for all $E' \in \boldsymbol{\mathcal{E}}$, $E' \succcurlyeq^{b} E$; and for all $E_1, E_2 \in \boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{N}}$, $E_1 \succcurlyeq^{b} E_2$ iff, if $o_1 \succ^{u} o_2$, then $\mathcal{F} \succcurlyeq \mathcal{G}$ whenever

(i) If $s \in E_1 - E_2$, then $\mathcal{F}(s) = o_1$ and $\mathcal{G}(s) = o_2$

(ii') If $s \in E_2 - E_1$, then $\mathcal{F}(s) = o_4$ and $\mathcal{G}(s) = o_3$, where $o_3 \sim^{u} o_1$, and $o_4 \sim^{u} o_2$

(iii') If $s \notin (E_1 - E_2) \cup (E_2 - E_1)$, then $\mathcal{F}(s) \sim^{u} \mathcal{G}(s)$

The justification for Definition 13 is essentially identical to the justifications for Definition 12 and Coherence, but it places strictly weaker requirements on the structure of $\boldsymbol{\mathcal{A}}$ than either of the latter two definitions.

It will come as no surprise that Definition 13 is still too strong. To *ensure* that $\succcurlyeq^{b}$ is always well-defined, it must be assumed that there will always be some $\mathcal{F}$ and $\mathcal{G}$ satisfying the conditions (i), (ii'), and (iii'), for any pair of non-null events E_1 and E_2 that we care to choose. And there are good reasons to think that this will not always be the case. Here is a schematic example.[30] Let E_1 be an event where, independently of any acts I might perform, many very good things occur, and let E_2 be an event where a great deal of very horrible things occur independently of any act I might perform. For simplicity, suppose that E_1 and E_2 are disjoint events. In fact, suppose that E_1 is so much better than E_2 that the very *best* possible outcome that might obtain if E_2 were true would still be worse than the very *worst* outcome that might obtain given E_1. If this is the case, however, then any act-function which satisfies (i) and (ii') *cannot* represent an available act: there *are* no acts α and β, for instance, such that α leads to o_1 at E_1, and β leads to $o_3 \sim o_1$ at E_2. According to Definition 13 then, E_1 and E_2 are *incomparable* with respect to $\succcurlyeq^b$.

A final illustration of the difficulties that come with trying to remove imaginary act-functions should suffice. As it turns out, there does appear to be a way to *systematically* construct a set of act-functions from a set of states and outcomes so as to *guarantee* act-independence, outcome-functionality, and Act–Function Correspondence. The strategy is based on a discussion of Lewis' (1981); Gibbard and Harper (1978) and Stalnaker (1972) also refer to a closely related idea, and it's critically discussed by Joyce (1999, pp. 115-19). First of all, take $\mathcal{A}'$—that is, a set of *acts* rather than act-functions—and $\mathcal{O}$ as primitive. It is assumed that the outcomes in $\mathcal{O}$ are mutually exclusive and jointly exhaustive, consistent with the performance of any act in $\mathcal{A}'$, and context neutral. $\mathcal{S}$ can now be defined as the set of all functions from $\mathcal{A}'$ to $\mathcal{O}$.

For instance, suppose there are only two available acts, α and β, and only two possible outcomes, o_1 and o_2. Then $\mathcal{S}$ contains four distinct functions:

$$s_1 = \{(\alpha, o_1), (\beta, o_1)\}$$
$$s_2 = \{(\alpha, o_1), (\beta, o_2)\}$$
$$s_3 = \{(\alpha, o_2), (\beta, o_1)\}$$
$$s_4 = \{(\alpha, o_2), (\beta, o_2)\}$$

In Lewis' terminology (1981, p. 11), each $s \in \mathcal{S}$ can be taken to represent a *dependency hypothesis*; i.e., a conjunction of counterfactuals which describes one of the different possible ways that the outcomes in $\mathcal{O}$ could causally depend upon the acts the agent might perform. For instance, s_1 can be read as *Regardless of what I do, o_1 obtains*, while s_2 is *If I do α, then o_1 will result, but if I do β, then o_2 will result*. Every dependency hypothesis is then (causally and hence logically) act-independent and outcome-functional (but *not* outcome-independent). Furthermore, given our assumptions, the set of dependency hypotheses is a partition of the relevant logical space.

With this in hand, each act in $\mathcal{A}'$ can be paired directly with an act-function in $\mathcal{A} \subset \mathcal{O}^{\mathcal{S}}$:

$$\alpha \triangleq \mathcal{F} = \{(s_1, o_1), (s_2, o_1), (s_3, o_2), (s_4, o_2)\}$$
$$\beta \triangleq \mathcal{G} = \{(s_1, o_1), (s_2, o_2), (s_3, o_1), (s_4, o_2)\}$$

[30] Thanks to Rachael Briggs for discussion here, and for help with this example. Exactly the same example also shows that Definition 12 and Coherence cannot always be applied.

The construction is such that there are *never* any constant act-functions. On the other hand, there will be *constant states*, or dependency hypotheses which imply that every act results in the same outcome. A consequence of these constant states is that the range of *every* act-function includes the entirety of $\mathcal{O}$. Moreover (as evidenced in the given example), act-functions will always evenly distribute the outcomes in $\mathcal{O}$ amongst the states in $\mathcal{S}$. For example, if there are 3 outcomes and 4 available acts, and thus $3^4 = 81$ states, each act-function will distribute each of the three outcomes to exactly 27 of those states. Thus, if there are more than 2 outcomes, we will never find *bifurcating* acts in $\mathcal{A}$ either (which figure centrally in Coherence).

Because $\mathcal{A}'$ is taken as primitive, and $\mathcal{A}$ is ultimately defined in terms of it, Act–Function Correspondence can hardly be doubted on this picture—indeed it seems about as plausible as it possibly can be. However, it also evident that none of the suggested definitions of $\succcurlyeq^u$ and $\succcurlyeq^b$ discussed above will be adequate if we adopt this framework. The Lewisian set of act-functions $\mathcal{A}$ has an interesting, and mathematically very elegant, structure to it—but it's the wrong kind of structure to guarantee that $\mathcal{N}$, $\succcurlyeq^u$, and $\succcurlyeq^b$ will always, or even *often*, be defined if Coherence, Definition 10, Definition 11, and/or Definition 13 are adopted. For example, the existence of constant states is enough to ensure that the earlier example given against Definition 13 applies; and Definition 11 cannot usefully be applied to any fatalist whose credence is distributed only over constant states. It may, of course, be possible to develop an interesting representation theorem based on this kind of construction—though I don't see how—but whatever it may turn out to be like, it will be quite different in its construction of $\mathcal{B}el$ and $\mathcal{D}es$ than anything Savage or his followers have put forward.

All of this suggests that it's very difficult—at best—to construct a Savage-like representation theorem without making some very strong assumptions about the set of act-functions, which seem implausible if Act–Function Correspondence is assumed. Savage's definitions of $\succcurlyeq^u$ and $\succcurlyeq^b$ are obviously off the table, but so are nearby suggestions. This point is borne out by other representation theorems developed within the Savage paradigm. These theorems typically require, if not constant acts, then at least a very richly structured $\mathcal{A}$ involving some imaginary act-functions. It would be an interesting project to see whether any interesting result can be achieved using the dependency hypothesis framework, but for the purposes of this discussion the key point is that no such results have been discovered—nor is it obvious than any will be found.

The presence of imaginary act-functions in $\mathcal{A}$ and Act–Function Correspondence are jointly inconsistent. So far, I have considered removing imaginary act-functions from the picture. I have argued that it seems highly unlikely that a Savage-like representation theorem will be developed under which Act–Function Correspondence is plausible. Nevertheless, removing imaginary act-functions from $\mathcal{A}$ is not the only possible response to the constant acts problem. Many authors working within the Savage paradigm are content to define $\succcurlyeq$ over imaginary act-functions, and *ipso facto* reject Act–Function Correspondence. It is to that response that I now turn.

3.3 Imaginary acts and (im)possible patterns of outcomes

Savage did not publish a response to the constant acts problem, though Fishburn (1981, pp. 162-3) reports that it "did not greatly bother Savage since he felt that the preference comparisons required by his axioms were conceptually reasonable". Exactly what Fishburn meant by this is unclear, but many have taken it to mean that Savage was content to deal with preferences over *imaginary acts*—acts which, while not actually available for the agent to perform, could still in some sense or other be imagined.[31] Others—perhaps even most who have applied the Savage framework—have expressed similar sentiments.[32] That is, the most common response to the constant acts problem is that it seems conceptually possible to *imagine* some act which gives rise to such-and-such outcomes dependent on such-and-such states of the world obtaining, even if it's granted that the outcomes might be inconsistent with the states.

Unfortunately, it is very rare that much more is said on the issue beyond the bare assertion that imaginary acts make sense and that we can have preferences over such things. This situation is unsatisfactory; as I have been stressing, the interpretation of any one element of Savage's formalism is intimately tied up with the interpretation of every other element, and the introduction of imaginary acts into the intended interpretation of $\mathcal{A}$ has important consequences elsewhere. Most importantly, the inclusion of imaginary acts is incompatible with Savage's proposed interpretation of $\succcurlyeq$: "Loosely speaking, [α] $\succcurlyeq$ [β] means that, if [the subject] were required to decide between α and β, no other acts being available, he would decide on α". It is hard to make sense of this behavioural interpretation as being even "loosely" adequate if α and/or β are imaginary acts, especially if they are acts which result in inconsistent state-outcome pairs.

Preferences between imaginary acts call for a non-behavioural construal of $\succcurlyeq$, and it's evident in the literature that those who adopt imaginary acts as part of their interpretation of Savage's act-functions forego the behavioural reading of $\succcurlyeq$ in favour of a somewhat more mentalistic construal. Indeed, Broome (1991, 1993) refers to preferences over imaginary acts as *non-practical preferences*, as whatever preferences they represent cannot be manifest in agents' dispositions to choose between available acts. And James Dreier describes the self-elicitation of non-practical preferences as follows:

> Asked whether I prefer [α] or [β], I imagine myself in a situation in which I have to choose between them. I find myself inclined to choose [α]. I report, on that basis, that I prefer [α] to [β]. (1996, p. 268)

Supposing that every act-function corresponds to some imaginable act, one could interpret $\succcurlyeq$ as encoding an agent's dispositions to *judge* that she would choose one imagined act over another. Sobel's (1997) notion of a 'pairwise preference' is described in a similar vein.

It is somewhat doubtful that we can always conceive of an act which corresponds to an arbitrarily chosen pattern of outcomes—I at least struggle to picture an act which always

[31] See, e.g., (Levi 2000, p. 398): "Savage's approach does not require that the preference ranking over potential options be a preference ranking over actual options … There is textual evidence that Savage clearly understood this." I think Levi is entirely right about this—in particular, if constant act-functions are understood as representing genuinely available acts, then decision theory becomes trivial: every agent ought to perform the constant act which results in the best possible outcome at any state (Joyce 1999). Since he obviously did not intend for his theory to be trivial, it's plausible that Savage took some of his act-functions to represent imaginary acts. However, there is also textual evidence that Savage did not fully appreciate what this meant for his supposedly 'behaviouristic' definition of credences and utilities, nor the implications that this interpretation has for the plausibility of his proposed decision rule (an example of which is discussed below). Furthermore, the interpretation conflicts sharply with how Savage introduces his decision theory in the early pages of his (1954).

[32] See (Buchak 2013, pp. 91-2) for a recent example.

brings it about that, say, *I have a glass of iced tea*, even at worlds where tea does not exist. There is, however, perhaps a more reasonable way to understand the situation, suggested by the following passage by Glen Shafer:

> [Savage] saw no reason why a person could not think about patterns of consequences corresponding to imaginary acts and formulate preferences between such patterns. In order to construct a preference between one pattern of consequences and another, it is not necessary that a person should have available a concrete act that produces this pattern, *or even that the person should be able to imagine such an act*. (1986, p. 470, emphasis added)

Instead of representing *acts*—whether real or imagined—by virtue of describing their patterns of outcomes, we might instead suppose that act-functions represent patterns of outcomes *directly*.[33] Some of these patterns may correspond to things that an agent might actually *do*, and some might correspond to things she might imagine herself doing, but many may not. It seems plausible to suppose, as Shafer suggests, that arbitrary *patterns of outcomes* are in principle available to the imagination, and that we might have preferences over such things, regardless of whether we can imagine any acts which might bring such patterns about.

One way to cash this idea out in more detail would be to let each act-function stand for an immense (possibly infinite) conjunction of counterfactuals,

$$(s_1 \,\square\!\!\rightarrow o_i) \;\&\; \ldots \;\&\; (s_n \,\square\!\!\rightarrow o_j)$$

It is then to be supposed that '$(s_i \,\square\!\!\rightarrow o_i)$' is one of the conjuncts just in case the conjunction it forms a part of corresponds to the act-function which maps s_i to o_i.[34] It could then be said that:

$(\{s_i\}, o_i \,|\, \ldots \,|\, \{s_n\}, o_n) \succcurlyeq (\{s_j\}, o_j \,|\, \ldots \,|\, \{s_m\}, o_m)$ if and only if the S prefers that $(s_1 \,\square\!\!\rightarrow o_i)$ & … & $(s_n \,\square\!\!\rightarrow o_n)$ rather than that $(s_1 \,\square\!\!\rightarrow o_j)$ & … & $(s_n \,\square\!\!\rightarrow o_m)$

Patterns of outcomes are not the kind of things that an agent *does*, nor are they the immediate objects of choice in any practical sense—so, again, this way of interpreting the elements of $\mathcal{A}$ does not sit well with a behavioural interpretation of $\succcurlyeq$. Note, however, that on this interpretation of act-functions there can be no question as to whether **SAV0** is true: every act-function can be uniquely paired with some conjunction of counterfactuals, regardless of what the decision-maker's situation happens to be like.

There are complaints that can be raised, though. As Joyce (1999, pp. 107-8) notes, it's exceedingly unlikely that anyone's preferences understood as such would satisfy **SAV1**, which requires $\succcurlyeq$ to be *complete* on $\mathcal{A}$. For one thing, there are far too many patterns of outcomes to

[33] In *Representation Theorems and the Grounds of Intentionality*, I provide an argument from another direction that the best interpretation of Savage's act-functions is in terms of patterns of possible outcomes which may or may not correspond to things the agent in question might do.

[34] Joyce (1999, pp. 62-5) argues that counterfactual conditionals would be inadequate for this way of interpreting Savage's act-functions, and instead posits a (somewhat mythical) 'Savage conditional' to play the role instead. It is orthogonal to my purposes to consider whether his argument against the use of counterfactuals is convincing, as the point I wish to make can be made just as well if we assume that every act-function represents an immense conjunction of Savage conditional statements.

imagine—uncountably many in Savage's system, as it turns out—and there seems to be no rational reason to consider all of them. This point has both descriptive and normative force. Joyce argues that completeness is not a requirement of rationality, but it's all the more clear that completeness is not even close to descriptively plausible either—and this places pressure on any version of characterisational representationism based on Savage's theorem (or a theorem which requires a similarly rich space of act-functions). Furthermore, without completeness, it's unclear whether agents would non-trivially satisfy Savage's other preference conditions. Note that almost *every* Savage-like theorem assumes **SAV1**; indeed it's very difficult to achieve a strong representation result without it. Those that try to do without **SAV1** appeal to a notion of coherent extendibility (discussed shortly) and have correspondingly weak uniqueness results; see, e.g., (Seidenfeld, Schervish *et al.* 1990, 1995).

Indeed, there is a tension within Savage's system, between requiring that agents have complete preferences on the space of *imaginable acts* (or *imaginable patterns of outcomes*) on the one hand, and how their decision-making behaviour is modelled on the other. The set of null events $\mathcal{N}$ is intended to characterise those propositions that the agent has no credence in, and a decision-maker who satisfies Savage's preference conditions is modelled as essentially *ignoring* null events when choosing between her options—hence she is indifferent between two act-functions if their outcomes only differ on null events (Definition 10). Introspectively, this is plausible—when deciding between options we discount the impossible (and perhaps even the exceedingly unlikely). It is odd, then, to simultaneously require of an agent a disposition to discount zero credence states when considering an acts' outcomes, while at the same time require interesting preference patterns between acts she is sure she cannot perform (or patterns of outcomes she is sure cannot be brought about).

A pair of examples may help to draw out this tension somewhat. Consider the choice between two act-functions: $f = (E_1, o_1 | E_2, o_2 | E_3, o_3)$ and $g = (E_1, o_1 | E_2, o_2 | E_3, o_4)$, where $\mathcal{D}es(o_3) > \mathcal{D}es(o_4)$ and $\mathcal{B}el(E_3) = 0$. Here are two possible patterns of reasoning. The first suggests that we ought to be indifferent between f and g, as they have exactly the same outcomes for all the states *that we have credence in*. This is the pattern of reasoning that Savage's axioms seem to require of the decision-maker. On the other hand, the second rests on the intuition that *preference comes cheap*: one only needs the very slightest of reasons to prefer one option f over another g, *all else being equal*; and the fact that f leads to a better outcome in some states is one such reason, however small it may be. In support of this reasoning, it is helpful to keep in mind that a probability of zero does *not* imply impossibility: assuming probabilism, there are contingent events that any rational credence function must assign a probability of zero.[35]

Now consider preferences between f and g, now being construed as conjunctions of counterfactuals, where the decision-maker fully recognises the logical impossibility of each such conjunction. Here are two possible patterns of reasoning that one might take towards this situation. The first suggests that we ought to be indifferent between f and g: each is logically inconsistent, each picks out exactly the same set of possible worlds (i.e., the empty set), so there is no reason to prefer one to the other. On the other hand, one might reason that *preference comes cheap*: although both f and g represent impossible propositions, one of those impossible propositions picks out a better way the world might (not) be than the other. (I will

[35] This point is made forcefully in (Hájek 2003) and (Hajek MS), but it goes back at least as far as Kolmogorov's early discussions on probabilities.

never own a spherical cube, but if I had the choice I'd prefer a spherical cube made of gold than a similarly shaped cube made of nightmares.)

The issue, of course, is that Savage's axioms require the decision-maker to adopt the first pattern of reasoning in the first case—or at least something extensionally equivalent to it—while the defence of his structural axioms **SAV0**/**SAV0'** being considered now appeals to the second pattern of reasoning in our second case. There is not a straightforward contradiction here, but there certainly does seem to be a tension: one can either take the hard-nosed approach by treating all null propositions as identical for the purposes of forming our preferences, or one can follow the intuition that preferences come cheap. Taking the former approach fits best with how Savage wants us to understand dominance reasoning, but would lead to uninteresting preferences over the set of constant act-functions (and over imaginary act-functions more generally). Taking the latter approach will allow us more freedom for forming preferences over act-functions even when they represent impossible propositions, but will not sit well with the substantive constraints that Savage places on those preferences.

By dropping Act–Function Correspondence and reconstruing the interpretation of $\mathcal{A}$ as either a space of imaginable acts or arbitrary patterns of outcomes, all that has been achieved is the exchange of one problem for a host of others. While **SAV0**/**SAV0'** seems salvageable under the re-interpretation, it comes at the cost of making **SAV1** almost certainly false, and doubt can be cast on whether the remaining preference conditions can be non-trivially satisfied. There is, however, one further response to the constant acts problem which I will consider briefly, which seems to me the strongest response available.

3.4 Coherent extendibility

Suppose that $\succcurlyeq$ is incomplete on $\mathcal{A}$, however $\succcurlyeq$ and $\mathcal{A}$ are supposed to be interpreted. This may be because $\succcurlyeq$ is given a behavioural interpretation and can only be coherently understood as holding between act-functions which correspond to available acts, and so is not defined on act-functions which don't correspond to available acts. (That is, $\mathcal{A}$ might be taken to represent the union of $\mathcal{A}'$ with some set $\mathcal{A}^*$ of purely fictional entities, where any behavioural preference relation would be defined only for pairs taken from the subset $\mathcal{A}'$.) Alternatively, we may suppose that $\succcurlyeq$ is incomplete on $\mathcal{A}$ because $\succcurlyeq$ is defined in terms of preferences between patterns of outcomes, but the agent only has preferences for a limited number of such patterns.

In any case, if $\succcurlyeq$ is incomplete on $\mathcal{A}$ then **SAV1** is false, then many of Savage's other preference conditions may be only trivially satisfied, and $\succcurlyeq$ is likely too impoverished to guarantee that $\succcurlyeq^u$ and $\succcurlyeq^b$ are complete on $\mathcal{O}$ and $\mathcal{E}$ respectively. Nevertheless, there may be an *extension* of $\succcurlyeq$, call it $\succcurlyeq^+$, which *does* satisfy all of Savage's conditions. Define an extension $\succcurlyeq^+$ of $\succcurlyeq$ as any superset of $\succcurlyeq$; thus $\succcurlyeq^+$ agrees with $\succcurlyeq$ regarding all those elements of $\mathcal{A}$ for which $\succcurlyeq$ is defined. If any extension of $\succcurlyeq$ conforms to Savage's conditions, then Theorem 5.1 entails that it can be given an expected utility representation. This fact could prove useful for dealing with the issues raised §§3.2–3.

Say that $\succcurlyeq$ is *coherently extendible* if it has at least one extension $\succcurlyeq^+$ which *does* satisfy Savage's conditions (or the preference conditions of whatever theorem we are considering). It is not at all obvious that the preferences (however understood) of ordinary agents *are* coherently extendible with respect any contemporary Savage-like theorem's preference conditions—but if they are, then the path is open for the advocate of characterisational representationism to attempt a characterisation of credences and utilities in terms of the representations that the theorem supplies for the extended relations $\succcurlyeq^{+}$.

In most cases, if $\succcurlyeq$ is coherently extendible at all, then there will be a large number of extensions which satisfy the stated conditions, and something would have to be said about this fact—though, this is not obviously a problem: so long as a theorem gives us substantial restrictions on the range of available interpretations, it need not have the standard (and very strong) uniqueness condition. One could appeal to further information to filter between alternative extensions of an agent's $\succcurlyeq$, thus (assuming the theorem in question has strong uniqueness results) arriving at a single expected utility representation of the agent's preferences. Alternatively, it could be argued that agents' credences (and likewise their utilities, *mutatis mutandis*) are best represented by a *set* of probability functions—*viz.*, the set determined by each coherent extension of her preference relation. This idea is not new; in the literature a set of probability functions designed to represent an agent's total credence state is called her *representor*. For discussion, see (Levi 1974), (Williams 1976), (Jeffrey 1983), (Walley 1991), and (van Fraassen 1990, 1995).

Appealing to coherent extensions of $\succcurlyeq$ seems to me the best hope we have for dealing with potentially incomplete preference systems—both for theorems within the Savage paradigm, and other theorems besides. For the strategy to be successful, of course, the preferences of ordinary agents must be coherently extendible to begin with—but it hardly seems like an impossible task to construct preference conditions such that this is possible. Moreover, $\succcurlyeq$ will have to be defined on *enough* of $\mathcal{A}$, however it ends up being interpreted, so that the range of possible coherent extensions is substantially restricted. This may not be so, for instance, if act-functions correspond to *infinite* conjunctions of counterfactuals, as in §3.3—in which case, an ordinary agent may have *no* preferences over $\mathcal{A}$, so *every* way of satisfying the relevant preference conditions will be a coherent extension of her $\succcurlyeq$-ranking, and the representor will be utterly uninteresting *qua* model of her credences and utilities. Similar things are likely to be true if $\succcurlyeq$ is defined on the union of $\mathcal{A}'$ with some set $\mathcal{A}^*$ of purely fictional entities, *if* $\mathcal{A}^*$ constitutes the very large majority of $\succcurlyeq$'s domain.

Let me summarise where things stand with the constant acts problem in relation to characterisational representationism. Admitting imaginary act-functions into $\mathcal{A}$ and assuming Act–Function Correspondence is not a coherent possibility. Thus, the proponent of characterisational representationism might try to retain Act–Function Correspondence while reconstructing $\mathcal{A}$ from the ground up *à la* Lewis, or she might drop Act–Function Correspondence and supply some alternative interpretation of $\mathcal{A}$ and $\succcurlyeq$.

Either option is consistent with appealing to a notion of coherent extendibility to achieve a final representation of an agent's credences and utilities. Appealing to coherent extensions will in general mean giving up on using the theorem to construct a unique $\mathcal{B}el$ and $\mathcal{D}es$ model of the agent, but given the kinds of strong preference conditions needed to attain strong uniqueness results that was likely a fool's errand in any case. The appeal to coherent extensions also suggests the possibility of retaining a choice-based interpretation of $\succcurlyeq$ even without Act–Function Correspondence. It is less clear, however, whether ordinary agents' preferences over whatever $\mathcal{A}$ represents are (a) coherently extendible to begin with, and (b) sufficiently rich so as to substantially narrow down the range of possible coherent extensions.

References

Balch, M. (1974). 'On recent developments in subjective expected utility'. In *Essays on Economic Behavior under Uncertainty*. M. Balch, D. McFadden and S. Wu, Eds. North-Holland Publishing Company: 45-54.

Balch, M. and Fishburn, P. C. (1974). 'Subjective Expected Utility for Conditional Primitives'. In *Essays on Economic Behavior under Uncertainty*. M. Balch, D. McFadden and S. Wu, Eds. North-Holland Publishing Company: 57-69.

Bernoulli, D. (1738). 'Specimen theoriae novae de mensura sortis.' *Commentarii academiae scientiarum imperialis Petropolitanae* **5**: 175-92.

Bradley, R. (2001). 'Ramsey and the Measurement of Belief'. In *Foundations of Bayesianism*. D. Corfield and J. Williamson, Eds. Kluwer Academic Publishers: 263-90.

Bradley, R. and Stefansson, H. O. (forthcoming). 'Counterfactual Desirability.' *British Journal for the Philosophy of Science*.

Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Basil Blackwell Press.

—— (1993). 'Can a Humean be moderate'. In *Value, Welfare and Morality*. R. G. Frey and C. W. Morris, Eds. Cambridge University Press: 51-73.

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

Casadesus-Masanell, R., Klibanoff, P. and Ozdenoren, E. (2000). 'Maxmin Expected Utility over Savage Acts with a Set of Priors.' *Journal of Economic Theory* **92**: 35-65.

de Finetti, B. (1931). 'Sul Significato Soggettivo Della Probabilita.' *Fundamenta Mathematicae* **17** (1): 298-329.

—— (1964). 'Foresight: its logical laws in subjective sources'. In *Breakthroughs in Statistics*. S. Kotz and N. L. Johnson, Eds. Springer: 134-74.

Dreier, J. (1996). 'Rational preference: Decision theory as a theory of practical rationality.' *Theory and Decision* **40** (3): 249-76.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

—— (1990). 'Précis of "Explaining Behavior: Reasons in a World of Causes".' *Philosophy and Phenomenological Research* **50** (4): 783-6.

Fishburn, P. C. (1967). 'Preference-based definitions of subjective probability.' *The Annals of Mathematical Statistics* **38** (6): 1605-17.

—— (1970). *Utility theory for decision making*. John Wiley & Sons.

—— (1974). 'On the Foundations of Decision Making under Uncertainty'. In *Essays on Economic Behavior under Uncertainty*. M. S. Balch, D. L. McFadden and S. Y. Wu, Eds. North-Holland Publishing Company: 1-25.

—— (1981). 'Subjective expected utility: A review of normative theories.' *Theory and Decision* **13** (2): 139-99.

Gaifman, H. and Liu, Y. (MS). *Decision Making without Miracles*. Unpublished manuscript. Retrieved from http://yliu.net/wp-content/uploads/state-dependent.pdf.

Gibbard, A. and Harper, W. L. (1978). 'Counterfactuals and Two Kinds of Expected Utility'. In *Foundations and Applications of Decision Theory*. C. A. Hooker, J. J. Leach and E. Francis, Eds. D. Reidel Publishing: 125-62.

Hajek, A. (MS). *Staying Regular?* Unpublished manuscript. Retrieved from http://philosophy.anu.edu.au/sites/default/files/Staying%20Regular.December%2028.2012.pdf.

Hájek, A. (2003). 'What Conditional Probability Could Not Be.' *Synthese* **137** (3): 273-323.

Hedden, B. (2012). 'Options and the subjective *ought*.' *Philosophy Studies* **158**: 343-60.

Jeffrey, R. C. (1983). 'Bayesianism with a human face.' *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science* **10**: 133-56.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge University Press.

Karni, E. (1993). 'Subjective Expected Utility Thoeory with State-Dependent Preferences.' *Journal of Economic Theory* **60**: 428-38.

Karni, E., Schmeidler, D. and Vind, K. (1983). 'On State Dependent Preferences and Subjective Probabilities.' *Econometrica* **51** (4): 1021-31.

Kochov, A. (2015). 'Time and No Lotteries: An Axiomatization of Maxmin Expected Utility.' *Econometrica* **83** (1): 239-62.

Krantz, D. H. and Luce, R. D. (1974). 'The interpretation of conditional expected-utility theories'. In *Essays on Economic Behavior under Uncertainty*. M. Balch, D. McFadden and S. Wu, Eds. North-Holland Publishing Company: 70-3.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. (1971). *Foundations of measurement, Vol. I: Additive and polynomial representations*. Academic Press.

Kreps, D. (1988). *Notes on the Theory of Choice*. Westview Press.

Levi, I. (1974). 'On Indeterminate Probabilities.' *The Journal of Philosophy* **71** (13): 391-418.

—— (2000). 'Review of *The Foundations of Causal Decision Theory*.' *The Journal of Philosophy* **97** (7): 387-402.

Lewis, D. (1981). 'Causal Decision Theory.' *Australasian Journal of Philosophy* **59** (1): 5-30.

Luce, R. D. (1972). 'Conditional Expected, Extensive Utility.' *Theory and Decision* **3** (2): 101-6.

Luce, R. D. and Krantz, D. H. (1971). 'Conditional expected utility.' *Econometrica* **39** (2): 253-71.

Luce, R. D. and Suppes, P. (1965). 'Preference, Utility, and Subjective Probability'. In *Handbook of Mathematical Psychology*. D. R. Luce, R. R. Bush and E. H. Galanter, Eds. Wilely: 249-410.

Machina, M. J. and Schmeidler, D. (1992). 'A More Robust Definition of Subjective Probability.' *Econometrica* **60** (4): 745-80.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

Narens, L. (1976). 'Utility-uncertainty trade-off structures.' *Journal of Mathematical Psychology* **13** (3): 296-322.

Ramsey, F. P. (1931). 'Truth and probability'. In *The Foundations of Mathematics and Other Logical Essays*. R. B. Braithwaite, Ed. Routledge: 156-98.

Richter, M. K. (1975). 'Rational Choice and Polynomial Measurement Models.' *Journal of Mathematical Psychology* **12**: 99-113.

Roberts, F. S. (1974). 'Laws of Exchange and Their Applications.' *SIAM Journal on Applied Mathematics* **26**: 260-84.

Samuelson, P. A. (1938). 'A Note on the Pure Theory of Consumer's Behaviour: An Addendum.' *Economica* **5**: 353-4.

—— (1948). 'Consumption Theory in Terms of Revealed Preference.' *Economica* **15** (60): 243-53.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover.

Schervish, M. J., Seidenfeld, T. and Kadane, J. B. (1990). 'State-dependent utilities.' *Journal of the American Statistical Association* **85** (411): 840-7.

Schwarz, W. (MS). *Options and Actions*. Unpublished manuscript. Retrieved from http://www.umsu.de/papers/options.pdf.

Seidenfeld, T., Schervish, M. J. and Kadane, J. B. (1990). 'Decisions without Ordering'. In *Acting and Reflecting: The Interdisciplinary Turn in Philosophy*. W. Sieg, Ed. Kluwer Academic Publishers: 143-70.

—— (1995). 'A Representation of Partially Ordered Preferences.' *The Annals of Statistics* **23** (6): 2168-217.

Shafer, G. (1986). 'Savage revisited.' *Statistical Science* **1** (4): 463-501.

Sobel, J. H. (1986). 'Notes on decision theory: Old wine in new bottles.' *Australasian Journal of Philosophy* **64** (4): 407-37.

—— (1997). 'Cyclical Preferences and World Bayesianisms.' *Philosophy of Science* **64** (1): 42-73.

Spohn, W. (1977). 'Where Luce and Krantz do really generalize Savage's decision model.' *Erkenntnis* **11** (1): 113-34.

Stalnaker, R. C. (1972). 'Letter to David Lewis'. In *Ifs*. W. L. Harper, G. A. Pearce and R. Stalnaker, Eds. Reidel: 151-2.

Suppes, P. (1969). 'The role of subjective probability and utility in decision-making'. In *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969* Springer: 87-104.

van Fraassen, B. (1990). 'Figures in a Probability Landscape'. In *Truth or Consequences*. J. Dunn and A. Gupta, Eds. Kluwer: 345-56.

—— (1995). 'Belief and the Problem of Ulyssess and the Sirens.' *Philosophical Studies* **77**: 7-37.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wakker, P. P. and Zank, H. (1999). 'State dependent expected utility for Savage's state space.' *Mathematics of Operations Research* **24** (1): 8-34.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall.

Williams, P. (1976). 'Indeterminate Probabilities'. In *Formal Methods in the Methodology of the Empirical Sciences*. M. Przelecki, K. Szaniawski and R. Wojcicki, Eds. Reidel: 229-46.