Representing Suppositional Decision Theories with Sets of Desirable Gambles

Kevin Blackwell KEVIN.BLACKWELL@BRISTOL.AC.UK
Department of Philosophy, University of Bristol, UK

Abstract

The sets of desirable gambles framework has been well-studied as a tool for representing decision-making with imprecise probabilistic beliefs – under the assumption of act-state independence. The question of this paper is: can we use sets of desirable gambles to represent decisions where the states do depend (e.g., causally or probabilistically) on the acts? In particular, I investigate two possible routes for representing suppositional decision theories with sets of desirable gambles, concluding that while one route works only for the subclass of SDTs representable by general imaging, the other route can represent any SDT whatsoever. After giving a fairly flat-footed representation, I investigate whether it's equivalent to a construction directly from the local (suppositional) desirability judgments; it isn't, but this latter construction represents a different aggregation rule applied to the same "credal committee." Finally, I extend the representation to model uncertainty about the supposition rule itself, in addition to imprecise credences.

Keywords: suppositional decision theories, causal decision theory, evidential decision theory, sets of desirable gambles, imaging

1. A Motivating Example: Extortion

Consider the following scenario from Jim Joyce:

> Suppose you have just parked in a seedy neighborhood when a man approaches and offers to "protect" your car from harm for $10. You recognize this as extortion and have heard that people who refuse "protection" invariably return to find their windshields smashed. Those who pay find their cars intact. You cannot park anywhere else because you are late for an important meeting. It costs $400 to replace a windshield. Should you buy "protection"? [9, p. 115]

We might reason as follows. There are two possible states of the world that are relevant to this decision: whether my window will get broken or not; and there are two actions I might take: paying the $10 or refusing. So the payoff table looks like this:

Windshield, Payment	Broken	Not Broken
Pay	-$410	-$10
Don't Pay	-$400	0

Clearly, choosing not to pay *dominates*[1] paying on the partition {Broken, Not Broken}: if my window gets broken, I would prefer to have kept 10 extra dollars in my pocket; if my window doesn't get broken, the same is true.

What's gone wrong? The obvious problem is that whether my windshield gets broken or not *depends* on whether or not I pay; this is why it's *extortion*! But dominance reasoning only naturally applies to partitions of states that are, in some sense, *independent* of my actions; *which sense* turns out to be a matter of some controversy. The main contenders are evidential decision theory[2] (which takes ordinary probabilistic independence as the standard for when dominance reasoning is valid) and causal decision theory (which allows dominance reasoning in cases where the relevant states cannot be causally influenced by your actions). For an investigation of a causal version of the Sure Thing Principle, see [20].

Both EDT and CDT (or at least, several of the major flavors of CDT) turn out to be examples of what has been termed, following Joyce, *suppositional decision theories* [9, Chapter 6]. Informally: an SDT enjoins you to maximize expected utility *from a certain epistemic perspective* which will typically be different from the one you currently hold; in general, an SDT tells you to first *suppose* that you perform a particular action (which involves modifying your initial beliefs in a certain way characteristic of that kind of supposition) and to then calculate the expected value of that action with the suppositional credences. Different

[1] Relative to some partition, X, act A dominates B iff $\forall x \in X$, $v(A,x) \geq v(B,x)$ and there is at least one $x \in X$ for which $v(A,x)$ is strictly greater than $v(B,x)$; $v(A,x)$ represents the payoff the agent receives at the world where they perform act A and x occurs. An anonymous reviewer recommended I highlight the connection with Savage's Sure Thing Principle, which takes the inference from preference conditional on each element of some partition to unconditional preference as valid.

[2] Throughout the paper, by EDT I mean the SDT whose supposition operator is Bayesian conditionalization, as originally promoted by Richard Jeffrey [8]. An anonymous reviewer has reasonably pointed out that this terminology is misleading; there are other decision theories that could be considered "evidential" which have radically different structural assumptions than Jeffrey's theory, including [14]. My interest in discussing Jeffrey's version of EDT is mainly that Bayesian updating is a very intuitive example of a supposition rule, and so it's an interesting contrast case.

kinds of supposition procedures generate different kinds of SDTs: e.g., EDT is the decision theory that results from using ordinary Bayesian conditionalization as your supposition procedure (by stipulation; see footnote 2). There are different flavors of CDT which correspond to different kinds of supposition; arguably, the simplest to understand is with supposition given by Pearl's do-operator (which he takes to be the appropriate way of updating your beliefs in cases of causal interventions) [15].

As normally presented, both EDT and CDT assume that the agent has a *credence function*, representing precise subjective probabilistic judgments about the world; and both typically assume that the suppositional epistemic perspective the agent adopts is also a precise probability.

For decisions under act-state independence, there is a well-developed framework for representing *imprecise* beliefs that still imposes probabilistic consistency, under the assumption that the agent is trying to maximize expected utility: sets of desirable gambles. Originally introduced by Peter Williams [18], sets of desirable gambles have been well-studied by many researchers in the Imprecise Probabilities community.

The question I will be investigating in this paper is: can we use the sets of desirable gambles framework to model decision problems with act-state dependence? In particular, can we represent suppositional decision theories like EDT and CDT?

2. Preliminaries: Modeling Act-State Dependence with Gambles

To represent decision problems where the agent believes that their actions and some relevant states of the world might not be independent, it is useful to construct our outcome space in terms of separate event spaces representing, respectively, the acts under an agent's "direct" control $\mathcal{A}$, and all other states of the world $\mathcal{X}$ relevant to the decision. Then the total outcome space is $\Omega \subseteq \mathcal{A} \times \mathcal{X}$ (the Cartesian product). As long as the acts are *logically independent* of the states,[3] we can safely take $\Omega = \mathcal{A} \times \mathcal{X}$. We assume that the elements of $\mathcal{A}$ and the elements of $\mathcal{X}$ are, respectively, pairwise mutually exclusive. This lets us understand both $\mathcal{A}$ and $\mathcal{X}$ as partitions of Ω, and we can freely translate between values of variables and sets of events. E.g., if $\mathcal{A} = \{A_1, \ldots A_\ell\}$ and $\mathcal{X} = \{X_1, \ldots, X_m\}$, then $\Omega = \{(A_i, X_j) \colon A_i \in \mathcal{A}, X_j \in \mathcal{X}\}$; but we will also understand $A_i = \{\omega \in \Omega \colon (\exists X_j \in \mathcal{X})(\omega = (A_i, X_j))\}$. Hence, $\omega \in A_i$ is another way of saying that A takes the value A_i at ω. Throughout the paper, I will consider only cases where this logical independence holds; I am also restricting my attention to cases where both $\mathcal{A}$ and $\mathcal{X}$ are finite, so that Ω is also finite.

[3] Viz., it is *logically possible* that both A and X obtain, for any $A \in \mathcal{A}$ and any $X \in \mathcal{X}$.

A gamble g is typically defined as a function $g : \Omega \to \mathbb{R}$, with $g(\omega), \omega \in \Omega$ interpreted as a gain/loss in *utility* (or a commodity for which the agent has a linear utility function) that the agent will receive if and only if ω obtains.[4] $\mathcal{L}(\Omega)$ is the set of all gambles, and, e.g., $\mathcal{L}(\mathcal{X})$ is the set of all gambles defined on $\mathcal{X}$ (the set of all functions from $\mathcal{X}$ to $\mathbb{R}$). For a gamble that depends, e.g., only on which state obtains, $g \in \mathcal{L}(\mathcal{X})$, we can also understand it as a gamble on Ω: $g^*(\omega) = g^*((A, X)) = g(X)$; g^* is sometimes called the *cylindrical extension* of g. However, in the rest of the paper, I will usually just identify the cylindrical extension with the gamble itself.

Immediately, there are a couple of things we have to be careful about. The first problem is that it's unclear that gambles defined generally on $\Omega = \mathcal{A} \times \mathcal{X}$ really make sense, from either a behavioristic or epistemic interpretation. The problem is: we are assuming that the acts in a decision problem are under the agent's control in the sense that they *can make it so* that some particular $A_i \in \mathcal{A}$ obtains. So, the agent's betting behavior on gambles defined purely on acts (viz., a gamble $g : g(\omega) = g_{A_i}$ iff $\omega \in A_i$) cannot generally be understood in terms of any prior beliefs about what the agent thinks they are likely to do.[5] The only reasonable decision principle for gambles defined purely on acts is *maximax*: $g \succ h$ iff there is some $A \in \mathcal{A}$ for which $g(A) > \max_{B \in \mathcal{A}} h(B)$. I simply pick the gamble which has the highest possible payout, and then perform the act that gets me that payout.[6]

Although I will be modeling gambles on the outcome space Ω, we will only give a decision-theoretic interpretation to comparisons of gambles of certain forms. In each of the two routes for representing SDTs with gambles that will be explored below, there will be a special class of gambles that we take as *characteristic* of the agent's candidate acts. We will primarily consider comparisons between gambles from this class, so that, e.g., finding $g - h$ desirable, where g, h represent the acts A, B, respectively, will represent preferring A to B. We will also make comparisons to certain other gambles for the purposes of pricing; these gambles will always be constant over $\mathcal{X}$. In one of the two routes, we will be interested in *constant* gambles $g(\omega) = \epsilon, \forall \omega \in \Omega$.

[4] But note that ordinarily, unlike in this paper, the outcome space Ω is assumed to represent states of the world independent of the agent's options; ordinarily, the gambles themselves *are* the acts.

[5] There is also, of course, continuing debate about whether these predictions about what the agent will do, "act credences", *themselves* make sense: the negative view has been sloganized by Isaac Levi as "deliberation crowds out prediction". See, e.g., [16], [10], [7], and [13]. My only point here is that there are problems with representing an agent's prior credences about acts via gambles on a space including the acts, assuming such beliefs make sense in the first place.

[6] If the agent has a decision problem containing a mix of decisions about betting on their own actions and outcomes which depend on their actions and the state of the world, things get a bit more complicated than just maximax. Nonetheless, it seems clear that these betting decisions don't reflect antecedent beliefs about which act the agent will perform.

(I will typically identify constant gambles with the real number that represents their constant reward.) In the other, we will be interested in gambles of the form $\mathbb{1}_A\epsilon$: gambles that are called off unless act A is performed, and return a constant utility of ϵ across all $\mathcal{X}$ whenever A is performed.[7]

The second problem is the issue of utility. In this paper, I will simply be assuming that our agent has a utility function that we can define in some world-independent way, so that we can assign utilities directly on Ω. There are interesting philosophical questions about whether this makes sense, and interesting technical questions about how you might represent preferences without making this kind of assumption. For an interesting exploration of some of these issues, see [21]. But I will not be engaging with these questions at all in this paper. We will simply assume that our agent has a utility function $u : \Omega \to \mathbb{R}$ which represents all preferences possibly relevant to the decision problem. $\mathbb{U}$ represents the set of all utility functions definable on Ω.

A couple more notational notes: I will sometimes use the notation $g > h$, where g and h are both gambles. What this really means is that g (weakly) Ω-dominates h: $\forall\omega \in \Omega$, $g(\omega) \geq h(\omega)$ and $\exists\omega \in \Omega : g(\omega) > h(\omega)$. When a gamble (weakly) dominates the zero gamble, I'll call it "positive"; $\mathcal{L}_{>0}$ is the set of all positive gambles.

3. The Bridge between Desirable Gambles, Utility, and Belief

For a decision theory where we assume that acts are independent of the states of the world relevant to some decision and that our agent is attempting to maximize expected utility, there is a very natural connection between coherent sets of desirable gambles, the agent's utility function, and their probabilistic beliefs. Suppose we have an agent with: a sample space of states of the world outside of their control, $\mathcal{X}$, relevant to some decision problem; a menu of actions $\mathcal{A}$; a utility function defined on $\Omega = \mathcal{A} \times \mathcal{X}$, $u : \Omega \to \mathbb{R}$, representing how desirable they would find each world, $\omega \in \Omega$, resulting from (consisting of) performing a certain act in a certain state; and imprecise probabilistic beliefs about $\mathcal{X}$, which we can represent as a set of probability functions defined on $\mathcal{X}$, P. Further, suppose that our agent evaluates the choiceworthiness of acts by *supervaluation* according to P: the agent prefers act A to act B ($A \succ_{u,P} B$) iff $\forall p \in P, E_p(u(A,\cdot)) > E_p(u(B,\cdot))$[8] *or* act A weakly dominates B.[9]

Then, whatever P is, there is a coherent set of desirable gambles D_P that represents the agent's beliefs in the following sense: for any utility function, u,

- we can associate each act $A \in \mathcal{A}$ with a *characteristic gamble*

$$g_A(\omega) = \mathbb{1}_A(\omega)u(\omega) = \begin{cases} 0, & \omega \notin A \\ u(\omega), & \omega \in A \end{cases}; \quad (1)$$

- and for any $A, B \in \mathcal{A}$, $A \succ_{u,P} B$ iff $g_A - g_B \in D_P$.

When considering how to represent suppositional decision theories like EDT and CDT with sets of desirable gambles, there are two natural-seeming generalizations of this bridge. We might hope to either:

1. Use the same set of gambles D_P, but allow for different kinds of characteristic gambles, so that we can find some way of reading off the preferences/valuations of our suppositional decision theory from our agent's current beliefs (as represented by D_P).

2. Hold fixed the definition of the characteristic gamble, but find some other set of desirable gambles which encodes the agent's preferences/valuations from the suppositional decision theory; we don't assume that this set of gambles reflects the agent's current beliefs.

Although both approaches might seem like reasonable ideas at first, it turns out that the first approach is only possible for the special subclass of suppositional decision theories representable by *generalized imaging*. However, we will see that the second approach is tractable for *any* suppositional decision theory whatsoever.

4. Suppositional Decision Theories and Imaging

Formally, supposition is represented by some operator $s : \mathbb{P} \times \mathcal{P}(\Omega) \to \mathbb{P}$, where $\mathbb{P}$ is the set of all probability functions defined on Ω and $\mathcal{P}(\Omega)$ is the powerset of Ω; the only further constraint on s is that, for any $R \in \mathcal{P}(\Omega)$ and any $\omega \notin R$, $s(p,R)(\omega) = 0$. Bayesian conditionalization is one example: $s_B(p,R)(\cdot) = p(\cdot|R) = \frac{\sum_{\omega \in R} p(\cdot \wedge \omega)}{\sum_{\omega \in R} p(\omega)}$.

An SDT, then, enjoins you to pick the available act which maximizes expected utility *under the supposition that you perform it*; if you have utility function u and precise credence function p, you act to maximize the quantity $V(p,u,\cdot) = E_{s(p,\cdot)}(u(\cdot)) = \sum_{\omega \in \Omega} s(p,\cdot)(\omega)u(\omega)$.

In the literature on Causal Decision Theory, the various flavors of CDT are typically taken to be representable by a special kind of supposition known as *generalized imaging*.[10]

[7] $\mathbb{1}_E$, for $E \subseteq \Omega$, is the *indicator* for the event E; $\mathbb{1}_E : \Omega \to \{0,1\}$, with $\mathbb{1}_E(\omega) = 1$ iff $\omega \in E$. It picks out which worlds the event obtains in.

[8] In the context of sets of desirable gambles (where desirability is strict preference to the status quo), Walley-Sen maximality and E-Admissability both collapse to supervaluation. Walley-Sen maximality and E-Admissibility are different only for non-binary preferences.

[9] This weak dominance condition is necessary for coherence, which we will discuss in Section 6.

[10] Versions of CDT that are *explicitly* constructed with reference to imaging functions are common in philosophy, including: Lewis's, Sobel's,

Imaging was originally introduced by David Lewis, while analyzing Stalnaker conditionals [11, p. 310]; generalized imaging is a generalization due to Peter Gärdenfors [5]. For much more on supposition, generalized imaging, and SDTs, including representation theorems for general SDTs, see [9], especially Chapters 6 and 7.

A general imaging function, $f : \mathcal{P}(\Omega) \times \Omega \to \mathbb{P}$, is a map from pairs of propositions and worlds to probability functions; it must also satisfy $f(R, \omega)(\omega') = 0$ for any $R \in \mathcal{P}(\Omega)$ and any $\omega' \notin R$. I don't find general imaging functions so intuitive, but I will do my best to provide some interpretations. One way to understand them is as a generalization of Stalnakerian selection functions. Here's another: if an agent had a prior credence function, p_ω, which was certain of some world ω, then $f(R, \omega)$ represents the credence function that the agent should have suppositional on R.

We say that a supposition operator, s, is *representable by general imaging* iff there is some general imaging function f s.t. $s(p, R)(\cdot) = \sum_{\omega \in \Omega} p(\omega) f(R, \omega)(\cdot)$, $\forall p \in \mathbb{P}$ and $\forall R \in \mathcal{P}(\Omega)$. In words: s is representable by general imaging when, for any credence function p and any proposition R, the credence that $s(p, R)$ assigns to some world ω', can be represented as the p-expectation (over all worlds $\omega \in \Omega$) of $f(R, \omega)(\omega')$.

5. Bridge Type 1

Formally, a bridge of Type 1 between an agent's beliefs and an SDT would be a map $g : \mathbb{U} \times \mathcal{A} \to \mathcal{L}(\Omega)$, where $\mathbb{U}$ is the set of all utility functions on Ω, s.t. $\forall u \in \mathbb{U}$ and $\forall P \subseteq \mathbb{P}$, $g(u, A) - g(u, B) \in D_P$ iff $(\forall p \in P)(V(p, u, A) > V(p, u, B))$; and for any $\epsilon \in \mathbb{R}$, $(\forall p \in P)(V(p, u, A) > \epsilon)$ iff $g(u, A) - \epsilon \in D_P$; $(\forall p \in P)(\epsilon > V(p, u, A))$ iff $\epsilon - g(u, A) \in D_P$. That is, we would have a way of "reading off", from gambles that are included in D_P, both the agent's preferences among the acts and facts about their (utility) prices according to the SDT.

Theorem 1 *For an SDT characterized by supposition operator s, the SDT admits a bridge of Type 1 iff s is representable by general imaging – at least for suppositions about acts (viz., there is a general imaging function f such that, for any $A \in \mathcal{A}$, $s(p, A)(\cdot) = \sum_{\omega \in \Omega} p(\omega) f(A, \omega)(\cdot)$, $\forall p \in \mathbb{P}$).* [11]

Proof $\leftarrow$: Suppose s is representable by general imaging, at least for suppositions about acts. Then $V(p, u, A) = \sum_{\omega \in \Omega} s(p, A)(\omega) u(\omega) = \sum_{\omega \in \Omega} \sum_{\omega' \in \Omega} p(\omega') f(A, \omega')(\omega) u(\omega) = \sum_{\omega' \in \Omega} p(\omega') \sum_{\omega \in \Omega} f(A, \omega')(\omega) u(\omega)$.

Define $g : (u, A) \mapsto \sum_{\omega \in \Omega} f(A, \cdot)(\omega) u(\omega)$. Then we have $V(p, u, A) = E_p(g(u, A))$. By definition of D_P, $g(u, A) - g(u, B) \in D_P$ iff the SDT recommends A over B; this is true for any $P \subseteq \mathbb{P}$ and any $u \in \mathbb{U}$ – which is to say, there is a bridge of Type 1 between the agent's beliefs and the SDT.

$\rightarrow$: Suppose there is a bridge of Type 1 between the agent's beliefs and the SDT. First, observe that a bridge of type 1 requires that, $V(p, u, A) = E_p(g(u, A))$, for any precise p, any utility function u, and any act A. (Suppose, for reductio, there is some p, u, A for which $V(p, u, A) > E_p(g(u, A))$. Then consider $\epsilon = V(p, u, A)$. $E_p(\epsilon - g(u, A)) > 0$, so $\epsilon - g(u, A) \in D_p$. But, obviously, $\epsilon \not> V(p, u, A)$, so g doesn't represent a bridge of type 1. The proof for $V(p, u, A) < E_p(g(u, A))$ is similarly obvious.)

Next, observe that $V(p, \mathbb{I}_\omega, A) = \sum_{\omega'} s(p, A)(\omega') \mathbb{I}_\omega(\omega') = s(p, A)(\omega)$. So, $E_p(g(\mathbb{I}_\omega, A)) = s(p, A)(\omega)$. A and ω are arbtirary, so s is representable by general imaging (at least for acts), with $f(A, \omega) = g(\mathbb{I}_\omega, A)$, for any $A \in \mathcal{A}$, and any $\omega \in \Omega$. ■

In general, Jeffrey's theory cannot be represented by general imaging [1, Theorem 1, Theorem 3], so it won't always admit a bridge of Type 1; however, the major flavors of CDT can be represented by general imaging. In cases where Bayesian updating and CDT happen to generate the same supposition rule, a Type 1 model may be possible for both (as we will see in Section 10).

5.1. Bridge Type 1 and Pricing

To readers who are already fairly familiar with sets of desirable gambles, the comparisons $g(u, A) - \epsilon$ and $\epsilon - g(u, A)$ might seem familiar. In the normal representation of the act-state independent case, we find that an agent is willing to buy a gamble g for a (utility) price ϵ whenever $g - \epsilon$ is desirable; similarly, an agent is willing to sell g for ϵ whenever $\epsilon - g$ is desirable.

This familiarity is misleading. In the ordinary setting, ϵ is a constant gamble over states of the world; in our setting the constant gamble ϵ represents a utility that is constant over *both acts and states*. It's an odd feature of the bridge of type 1 that these are the relevant comparisons, which requires some explanation; for more on what I think is strange

and Rabinowicz's versions of CDT; see [12], [17]. Other authors have defined CDT in terms of counterfactuals that can be analyzed with imaging functions, e.g., Gibbard & Harper [6]. J. Dmitri Gallow has a very nice constructive proof that causal intervention on a Causal Bayesian Network is always representable by a general imaging function [4]; we will apply this recipe in the example in Section 10 below.

[11] This theorem is basically a lazier version of a theorem proved by Snow Zhang and cited by Andrew Bacon [1, Theorem 3]; I'm including my proof only because the proof of Zhang's theorem hasn't been published anywhere. She has asked that I emphasize that her result is a straightforward corollary of [5, Theorem 1].

about this, see Subsection 6.1. Unfortunately, a thorough treatment of what I think is going on here would require more time and space than I am able to devote in the present paper. For now, I will just explore an odd feature of the characteristic gambles $g(u, A) = \sum_{\omega \in \Omega} f(A, \cdot)(\omega)u(\omega)$ defined for general imaging; I hope this feature will at least help illustrate the unusual concept of value that a bridge of type 1 involves.

The first thing to notice is that these characteristic gambles, radically unlike characteristic gambles as defined in Equation 1, are *not called off* when the act isn't performed. This is because $g(u, A)$ represents what we might term the "imaged value" of act A.[12] That is, if $f(A, \omega)$ picks out a particular world, ω', with certainty (viz., $f(A, \omega) = \mathbb{1}_{\omega'}$), then $g(u, A)(\omega) = u(\omega')$; if, more generally, $f(A, \omega)$ assigns nonzero probability to several worlds, $g(u, A)(\omega)$, is the expectation according to $f(A, \omega)$ of the utility of each of these worlds.

In particular, one fact about this kind of value is that, for any act A which has constant utility ϵ across all states, its imaged value is $g(u, A) = \epsilon$ – which is, again, a constant gamble across all *acts and states*.[13] This fact helps explain why bridge type 1 regards the relevant comparisons for buying prices as $g(u, A) - \epsilon$. Suppose an agent is in a decision problem where $A, B \in \mathcal{A}$, and where $u((B, X)) = \epsilon, \forall X \in \mathcal{X}$. As observed above, $g(u, B) = \epsilon$ according to any supposition rule representable by general imaging. So whenever B has constant utility ϵ in states, $g(u, A) - g(u, B) = g(u, A) - \epsilon$; which means bridge type 1 represents act A as preferred to the state-constant act B iff $g(u, A) - \epsilon \in D_P$.

6. Bridge Type 2

Formally, a bridge of type 2 exists when, for any P there is a set of desirable gambles D s.t. $\forall u \in \mathbb{U}$, for any two acts $A, B \in \mathcal{A}$, $A \succ_{s,u,P} B$ iff $g_A - g_B \in D$, with g_A, g_B defined as in Equation 1; and for any $\epsilon \in \mathbb{R}$, $(\forall p \in P)(V(p, u, A) > \epsilon)$ iff $g_A - \mathbb{1}_A\epsilon \in D$; $(\forall p \in P)(\epsilon > V(p, u, A))$ iff $\mathbb{1}_A\epsilon - g_A \in D_P$. That is, for any P representing an agent's beliefs, we can find some set of desirable gambles, D, which encodes the recommendations that the SDT makes given P, for all utility functions; D also encodes the agent's price judgments for all acts, though in a slightly different form than a bridge of type 1 does.

To see how, let's first consider the precise case: $P = \{p\}$. Define p_{eff} as the probability function which has $p_{eff}(\cdot|A) = s(p, A)(\cdot)$, for all $A \in \mathcal{A}$, and is uniform over the acts themselves. (Viz.: for any $A, B \in \mathcal{A}$, $p_{eff}(A) = \sum_{\omega \in A} p_{eff}(\omega) = p_u(B)$.) In particular, then, $p_{eff}(\omega) = p_{eff}((A, X)) = \frac{s(p,A)(\omega)}{||\mathcal{A}||}$. If we define characteristic gambles for acts as above, it's not difficult to see that, for any two acts A, B, $V(p, u, A) > V(p, uB)$ iff $E_{p_u}(g_A - g_B) > 0$.[14]

$E_{p_{eff}}(g_A - g_B) = \frac{E_{s(p,A)}(g_A) - E_{s(p,B)}(g_B)}{||\mathcal{A}||} = \frac{V(p,u,A) - V(p,u,B)}{||\mathcal{A}||}$.

Similarly, for an imprecise (non-singleton) P, let $P_{eff} = \{q \in \mathbb{P} : (\exists p \in P)(\forall A \in \mathcal{A}, X \in \mathcal{X})(q((A, X)) = \frac{s(p,A)((A,X))}{||\mathcal{A}||})\}$. For the same reason, it's clear that for any two $A, B \in \mathcal{A}$, $V(p, u, A) > V(p, u, B)$ for all $p \in P$ iff $E_q(g_A - g_B) > 0$ for all $q \in P_{eff}$.

So, we can always represent the recommendations of an SDT by $D_{P_{eff}} = \{g \in \mathcal{L}(\Omega) : E_q(g) > 0, \forall q \in P_{eff}\}$.[15]

As for prices: note that, for any $\epsilon \in \mathbb{R}$, $(\forall p \in P)(V(p, u, A) > \epsilon)$ iff $(\forall p \in P)(E_{p_{eff}}(g_A) > \frac{\epsilon}{||\mathcal{A}||} = E_{p_{eff}}(\mathbb{1}_A\epsilon))$; which is to say, $(\forall p \in P)(V(p, u, A) > \epsilon)$ iff $g_A - \mathbb{1}_A\epsilon \in D_{P_{eff}}$. In parallel fashion, we find that for any $\epsilon \in \mathbb{R}$, $(\forall p \in P)(\epsilon > V(p, u, A))$ iff $\mathbb{1}_A\epsilon - g_A \in D_{P_{eff}}$.

The next obvious question to ask is: will $D_{P_{eff}}$ typically be coherent? As we will see in the next section, the answer is technically "no", but we can make it coherent without causing any problems. But first, we should briefly discuss the way that a bridge of type 2 prices gambles.

6.1. Pricing for Bridge Type 2

In Subsection 5.1, I cautioned against regarding the comparisons $g - \epsilon$ and $\epsilon - g$ as the natural analogues of how prices of gambles are normally determined; let me say just a little bit about why here. The gamble $-\epsilon$ represents a cost to the agent that they must pay at all possible outcomes; in our setting, that means it's not only constant over the states of

[12] Andrew Bacon refers to this quantity as the "actual value" of the act A, which he interprets as representing "the utility of the world that would have obtained if A had been true" [1, p. 3]. This interpretation assumes that $f(A, \omega)$ represents the agent's beliefs about which worlds might have resulted, when the actual world is ω, if they had performed act A; but I don't think every SDT representable with a general imaging function does (or should) interpret f this way. E.g., it seems likely that Functional Decision Theory is formally representable by general imaging. If it is, $f(A, \omega)(\omega')$ for FDT represents something more like: if the actual world is ω, the probability that *the kind of agent* who would decide to perform A would end up at world ω'; proponents of FDT are quite clear that (something like) this is how they think agents should evaluate acts even when the agent knows that they are in a state of the world where ω' *cannot result from their actions*. See the discussion of the transparent Newcomb problem and Parfit's hitchhiker in [19].

[13] To see why: by assumption, $f(A, \omega)(\omega')$ is nonzero only for $\omega' \in A$ (supposition always involves certainty the supposed proposition obtains); $u(\omega') = \epsilon$ for all $\omega' \in A$; for every $\omega \in \Omega$, $f(A, \omega)$ is a probability function (and thus must assign total probability 1); and so $g(u, A)(\omega) = \sum_{\omega' \in \Omega} f(A, \omega)(\omega')u(\omega') = \sum_{\omega' \in A} f(A, \omega)(\omega')\epsilon = \epsilon, \forall \omega \in \Omega$, no matter what else is true of f.

[14] I want to stress here, again, that this is the *only* rationale for the use of p_{eff}; we can use it to encode the agent's preferences over acts, and prices, in line with the SDT. It is not intended to represent the agent's beliefs; "eff" stands for "fake" as much as "effective".

[15] Readers who are already familiar with the sets of desirable gambles framework might notice that this set of desirable gambles won't generally be coherent; we'll get to that in the next section.

the world, but also constant over the *acts* the agent chooses which of to perform. If we want to interpret a buying price for an act as a utility that the agent would be willing to pay to perform it, $g_A - \epsilon \in D_{Pe\!f\!f}$ doesn't represent this; whether the agent prefers $g_A - \epsilon$ to the status quo depends not only on how valuable act A would be, if performed, but on the agent's beliefs about *whether they will perform it*. If we wanted to give an interpretation to $g_A - \epsilon$, it would be more like buying the *option* or *right* to perform act A, where that allows for uncertainty about whether the option will be exercised.[16] The more natural understanding of paying to perform an act is a cost ϵ the agent pays iff they perform the act: $\mathbb{I}_A\epsilon$. Whether $g_A - \mathbb{I}_A\epsilon$ is desirable or not doesn't depend on the agent's beliefs about whether they'll perform act A; it purely reflects how valuable they would expect it to be, if performed.

Similarly, $\epsilon - g_A$ doesn't properly reflect an agent's willingness to sell the rewards of act A. If they perform any other act B, they pocket ϵ and certainly won't have to pay the buyer anything, because $g_A((B, \cdot)) = 0$. Again, for basically the same reasons, I claim the natural comparison is $\mathbb{I}_A\epsilon - g_A$. Personally, I find the way that bridge type 2 prices acts much more natural than the way bridge type 1 does.

7. Coherence for Suppositional Decision Theories

The standard coherence axioms for sets of desirable gambles are (see, e.g., [2, Definition 1]):

D$_1$. $0 \notin D$

D$_2$. $\mathcal{L}_{>0} \subseteq D$

D$_3$. If $f, g \in D$, then $f + g \in D$

D$_4$. If $f \in D$ and $\lambda \in \mathbb{R}_{>0}$, then $\lambda f \in D$.

D$_1$ (together with D$_3$ and D$_2$) encodes the rational requirement of *avoiding Dutch books* (or avoiding partial loss): there is no package of bets you can offer the agent which they find (individually) desirable but which combine into a gamble that is guaranteed to never make utility for the agent and which might lose utility in some outcome.[17]

Although willingness to accept the zero gamble doesn't seem like a rational problem in the same way that accepting a Dutch book does, it is convenient to exclude the zero gamble from desirability, because it lets us associate desirability with *preference to the status quo*, with the status quo understood as the zero gamble. It doesn't make sense to (positively) prefer the status quo to itself, so on this conception of desirability the zero gamble is never desirable.

D$_3$ and D$_4$ are both closure axioms that follow from (or characterize) the way we are understanding desirability. D$_3$ is the "package principle": if any pair of gambles are individually desirable, a single gamble that generates their combined payoffs in any possible outcome must be, too. D$_4$ can be understood as a consequence of the linearity of expectation together with the idea that what our agent is trying to do is maximize expected utility (given imprecise information). These assumptions are both consonant with the way we're understanding (supervaluational) suppositional decision theories, so perhaps it isn't surprising that $D_{Pe\!f\!f}$ also satisfies these axioms.

Not only would it be bad news to find a gamble that's non-positive at all worlds desirable; it would also be a problem if, as a consequence of other gambles you find desirable, you are implicitly *committed* to accepting a non-positive gamble. This requirement is often termed *consistency*; because our closure axioms are D$_3$ and D$_4$, this amounts to the requirement that $\text{posi}(D) \cap \mathcal{L}_{\leq 0} = \emptyset$.[18] Consistency certainly seems rationally required for the decisions recommended by a suppositional decision theory; and as we will see, $D_{Pe\!f\!f}$ satisfies this requirement.

As stated, $D_{Pe\!f\!f}$ will not typically satisfy D$_2$. D$_2$ is the rational requirement to *accept partial gain*: if there is a gamble that can't lose you utility, but might possibly win you utility (and taking it is *completely free*), you would be a fool not to take it – no matter what your probabilistic beliefs are.[19] This might seem like an unassailable principle of rational decision making, but we should notice: D$_3$ and D$_2$, together with our representational assumption that $g \succ h$ iff $g - h \in D$, entails *Ω-dominance*: if for all $\omega \in \Omega$ $g(\omega) \geq h(\omega)$ and there is at least one $\omega \in \Omega$ for which $g(\omega) > h(\omega)$, then $g \succ h$.

We have already seen, in Section 1, that (state-relative) dominance reasoning (for acts) is not always correct. Should we be similarly suspicious of dominance reasoning for gambles defined on the entire possibility space (including acts)? I argue that we shouldn't for two reasons: (1) unlike state-relative dominance, dominance on the total possibility space never *positively conflicts* with the recommendations of any SDT (viz.: there *cannot* be a case where act $A \succ_{SDT} B$ but g_B Ω-dominates g_A); and (2) although adding Ω-dominance can sometimes sharpen our agent's preferences (make slightly more desirability judgments between acts

[16] But note that bridge type 2 cannot properly evaluate this, because the uniform probabilities that $P_{Pe\!f\!f}$ assigns to acts are fake. For any hope of either bridge representing buying act options correctly, I think we would need to represent it as a sequential decision problem, which is outside the scope of this paper.

[17] Satisfying both D$_3$ and D$_2$ entails that if D accepts a partial loss, D must also contain 0. For any g which is nonpositive in every component and negative in at least one, $-g$ is nonnegative in every component and positive in at least one, so $-g \in D$; and $g + -g = 0$, so $0 \in D$.

[18] $\text{posi}(D)$, for $D \subseteq \mathcal{L}$, is the set of all positively-weighted, finite, combinations of gambles in D. Put another way: $\text{posi}(D)$ is the closure of D under D$_3$ and D$_4$.

[19] Note that this assumes that there are events you might regard as possible, but having probability zero.

than the SDT alone does), it only adds a very specific kind of judgment that is both innocuous and intuitive.

Theorem 2 *If g_A, g_B are the characteristic gambles (as defined in Equation 1) for two acts A, B, respectively, then g_A Ω-dominates g_B* only if *$g_B \leq 0$ and $g_A \geq 0$; furthermore, either $g_A > 0$, $g_B < 0$, or both.*

Proof Because g_A is the characteristic gamble of act A, $g_A((B, X)) = 0$ for all $X \in \mathcal{X}$. So if $g_A(\omega) \geq g_B(\omega)$, for all $\omega \in \Omega$, $g_B((B, X)) \leq 0$ for all $X \in \mathcal{X}$. And because g_B is the characteristic gamble of B, $g_B(\omega) = 0$ for all $\omega \notin B$; so $g_B(\omega) \leq 0$ for all ω.

Similarly, $g_B((A, X)) = 0$ for all $X \in \mathcal{X}$, so if g_A Ω-dominates g_B, then $g_A((A, X)) \geq 0$ for all $X \in \mathcal{X}$; $g_A(\omega) = 0$ for all $\omega \notin A$, and so $g_A(\omega) \geq 0$ for all ω.

If g_A Ω-dominates g_B, then there must also be at least one $\omega \in \Omega$ for which $g_A(\omega) > g_B(\omega)$, which means that g_A and g_B cannot both be the zero gamble. Since $g_A(\omega) \geq 0$ and $g_B(\omega) \leq 0$ for all ω, if they can't both be identically 0, either g_A is positive for some ω, g_B is negative for some ω, or both. ■

Theorem 3 *There is no SDT, $P \subseteq \mathbb{P}$, and $u \in \mathbb{U}$ for which act A is recommended over act B but g_B Ω-dominates g_A.*

Proof First observe, from the previous theorem, that g_B can Ω-dominate g_A only if g_B has no negative component and g_A has no positive component. So, there is no $p \in \mathbb{P}$ such that $E_p(g_A) > E_p(g_B)$; $\forall p \in \mathbb{P}$, $E_p(g_A) \leq 0$ and $E_p(g_B) \geq 0$. So there is no $D_{P_{eff}}$ s.t. $g_A - g_B \in D_{P_{eff}}$. ■

So, as alluded to earlier, we have that requiring our agent to satisfy Ω-dominance will never conflict with the desirability judgments of any SDT; and the only desirability judgments that it will (sometimes) add is that some acts which cannot yield positive utility in any outcome will be dispreferred to certain acts which cannot lose utility in any outcome – and where, further, there must be at least one outcome where the "good" act wins utility or the "bad" act loses utility. As long as we take the relevant outcomes to be genuinely possible, this seems intuitive: to trade the "bad" act for the "good" is a riskless improvement.

So I think the relevant coherence axioms for suppositional decision theories are just the standard ones. That might be surprising (it surprised me), but I hope the preceding discussion was convincing.

In any case, it is very simple to adjust $D_{P_{eff}}$ to satisfy the coherence axioms: let $\overline{D}_{P_{eff}} = \{g \in \mathcal{L}(\Omega) \colon g \in \mathcal{L}_{>0} \text{ or } \exists h \in D_{P_{eff}} : g \geq h\}$.

Theorem 4 *$\overline{D}_{P_{eff}}$ is the* natural extension *of $D_{P_{eff}}$; it is the smallest set of gambles which includes $D_{P_{eff}}$ and satisfies all of* D_1-D_4.

Proof First, we'll check that $\overline{D}_{P_{eff}}$ satisfies the coherence axioms.

D_2: holds by definition.

D_3, D_4: suppose $f, g \in \overline{D}_{P_{eff}}$. So, either $f \in \mathcal{L}_{>0}$ or there's some $f' : f \geq f'$ and $(\forall q \in P_{eff})(E_q(f') > 0)$; similarly for g. If f, g are both in $\mathcal{L}_{>0}$, then so is $f + g$, so by definition $f + g \in \overline{D}_{P_{eff}}$. If only one is in $\mathcal{L}_{>0}$, assume WLOG it's g. Then $f + g > f'$, and so $f + g \in \overline{D}_{P_{eff}}$. If neither is in $\mathcal{L}_{>0}$, then consider $f' + g'$. $(\forall q \in P_{eff})(E_q(f' + g') = E_q(f') + E_q(g') > 0)$, and $f + g \geq f' + g'$, so $f + g \in \overline{D}_{P_{eff}}$.

Similarly: let $\lambda \in \mathbb{R}_{>0}$. If $f \in \mathcal{L}_{>0}$, then so is λf. If not, then $(\forall q \in P_{eff})(E_q(\lambda f') = \lambda E_q(f') > 0)$ and $\lambda f \geq \lambda f'$, so $\lambda f \in \overline{D}_{P_{eff}}$.

D_1: suppose (for *reductio*) $0 \in D_{P_{eff}}$. $0 \notin \mathcal{L}_{>0}$, so there must be some $h : 0 > h$ and $(\forall q \in P_{eff})(E_q(h) > 0)$. But if $h < 0$, $E_p(h) \leq 0$ for any probability p. Contradiction.

To verify it's the *smallest* set of gambles including $D_{P_{eff}}$ which is coherent, observe that $\overline{D}_{P_{eff}} = \{g \in \mathcal{L}(\Omega) \colon (\exists h \in D_{P_{eff}})(g \geq h)\} \cup \mathcal{L}_{>0}$. Axioms D_3 and D_2 entail Ω-dominance (any gamble which Ω-dominates a desirable gamble must also be desirable) and $\{g \in \mathcal{L}(\Omega) \colon (\exists h \in D_{P_{eff}})(g \geq h)\}$ is just the closure of $D_{P_{eff}}$ under Ω-dominance. Any coherent D which includes $D_{P_{eff}}$ must have $D \supseteq \{g \in \mathcal{L}(\Omega) \colon (\exists h \in D_{P_{eff}})(g \geq h)\}$ and $D \supseteq \mathcal{L}_{>0}$, so $D \supseteq \overline{D}_{P_{eff}}$. ■

So, for any SDT and any $P \subseteq \mathbb{P}$, we can identify a coherent set of desirable gambles, which sometimes (very slightly!) *sharpens* the recommendations that the (supervaluated) suppositional decision theory makes given our agent's beliefs: for any $u \in \mathbb{U}$ and any pair of acts A, B, if $(\forall p \in P)(V(p, u, A) > V(p, u, B))$, then $g_A - g_B \in \overline{D}_{P_{eff}}$. And if $g_A - g_B \in \overline{D}_{P_{eff}}$, then either $(\forall p \in P)(V(p, u, A) > V(p, u, B))$ or g_A Ω-dominates g_B.

8. Representation with Local Desirability Assessments?

The construction of $D_{P_{eff}}$ from an SDT and the agent's set of credence functions P is not conceptually difficult, but it might be computationally demanding. We might hope that there is a way of representing the same information directly in terms of the local, suppositional desirability assessments – that is, from the beliefs that the agent would have about $\mathcal{X}$ suppositional on each candidate act in $\mathcal{A}$.

Let's return again to the precise case: $P = \{q\}$. Here is an alternative way of constructing $D_{P_{eff}} = D_{q_{eff}}$ in this case (which yields the same result). First, consider the gambles (defined on $\mathcal{X}$) that the agent would find desirable *under the supposition* of each act. Let $m = ||\mathcal{A}||$. For each $i \in 1 : m$,

let $D_i^q = \{g \in \mathcal{L}(\mathcal{X}) \colon E_{s(q,A_i)}(g) > 0\}$ – the gambles (defined on $\mathcal{X}$) that have positive expected value according to the probability that results from revising q by supposing A_i. Let $D_{uni} = \{f \in \mathcal{L}(\mathcal{A}) \colon \sum_i f(A_i) > 0\}$ represent the gambles (defined on $\mathcal{A}$) which have positive expected utility according to the uniform probability function defined on $\mathcal{A}$: $p_{uni}(A_i) = \frac{1}{m}$. Then let $D_{s,q} = \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i^q \cup D_{uni})$.

Theorem 5 *$D_{s,q} = D_{q_{eff}}$; $h \in D_{s,q}$ iff $h \in D_{q_{eff}}$, for all $h \in \mathcal{L}(\Omega)$.*[20]

Moving beyond the precise case, then, we can generally also construct $D_{P_{eff}} = \bigcap_{p \in P} D_{s,p} = \bigcap_{p \in P} \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i^p \cup D_{uni})$.[21] The local desirability assessments for P suppositional on A_i are given by $D_i = \bigcap_{p \in P} D_i^p$. If it were true that $\bigcap_{p \in P} \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i^p \cup D_{uni}) = \text{posi}(\bigcup_i \mathbb{I}_{A_i} \bigcap_{p \in P} D_i^p \cup D_{uni})$, then we would have a very nice representation of $D_{P_{eff}}$ in terms of the local, suppositional desirability assessments.

As the reader may guess from the wording of the previous conditional: unfortunately, $D_{P_{eff}}$ is not generally equivalent to $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$. However: $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni}) \subseteq D_{P_{eff}}$. Any desirability judgment that $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$ makes, $D_{P_{eff}}$ also endorses; but $D_{P_{eff}}$ will (typically)[22] find some gambles desirable that $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$ takes no stance on. In this sense, $D_{P_{eff}}$ is generally *at least as informative as* $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$.

It turns out that $\text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$ actually represents a less informative *judgment aggregation rule* than $D_{P_{eff}}$. So far, we've been considering a supervaluated/unanimous-decision approach: act A is preferred to act B iff for every $p \in P$, $V(p,u,A) > V(p,u,B)$.

Theorem 6 *For any $P \subseteq \mathbb{P}$, $u \in \mathbb{U}$, and any two acts, $A, B \in \mathcal{A}$, $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$ iff either: (1) $\exists \epsilon_1, \epsilon_2 \in \mathbb{R}$ such that $(\forall p \in P)(V(p,u,A) > \epsilon_1 > \epsilon_2 > V(p,u,B))$; (2) $g_A = 0$ and $\forall p \in P$, $V(p,u,B) < 0$; (3) $g_B = 0$ and $\forall p \in P$, $V(p,u,A) > 0$; or (4) $\forall p \in P$, $V(p,u,A) > 0$ and $V(p,u,B) < 0$.*[23]

This aggregation rule is clearly less informative than supervaluation, because all 4 conditions entail that $\forall p \in P$, $V(p,u,A) > V(p,u,B)$, but the latter can be true without satisfying any of the four conditions; put another way, it's a *stricter* notion of preference derivable from the SDT and the agent's beliefs.

[20] For lack of space, the proof has been relegated to the supplement.

[21] This amounts to taking the marginal extension of the agent's actual suppositional assessments with the fake uniform prior over acts.

[22] There are special cases where they will be equal. One that we've already seen is for any precise $P = \{q\}$.

[23] For lack of space, the proof has been relegated to the supplement.

9. More Uncertainty

Thus far, we've been assuming that our agent might have imprecise credences about Ω: they might not be able to attach precise probabilities to how likely it is that they will perform the acts in $\mathcal{A}$, or to how likely each state of the world in $\mathcal{X}$ is. However, we have been implicitly assuming that their suppositional updating rule is precise, in this sense: s maps a pair of probability function and supposed proposition to a single probability function. In the context of, e.g., causal decision theory, we might take this to reflect a kind of precision of belief *about the laws themselves*. For instance, this could make sense if our agent was certain about the objective chances of each state resulting from any intervention to perform an act. It might also make sense if our agent had precise probabilities over some different causal hypotheses which they combined to calculate the expected chance of each state resulting from performing each act. But what if our agent has imprecise beliefs about the laws themselves?

Let $\mathbb{S}$ be the set of all functions from $\mathbb{P} \times \mathcal{P}(\Omega)$ to $\mathbb{P}$. Rather than assuming that our agent performs supposition using a particular $s \in \mathbb{S}$, we could represent our agent as having a *set of candidate supposition rules* $S \subseteq \mathbb{S}$. Just as we can aggregate the judgments of an imprecise "credal committee" by supervaluation, we could do the same with the recommendations made by any candidate supposition rule set S.

Definition 7 (SP-supervaluated SDT) *For any $s \in \mathbb{S}$, $p \in \mathbb{P}$, $u \in \mathbb{U}$, let $V(s,p,u,\cdot) = E_{s(p,\cdot)}(u(\cdot)) = \sum_{\omega \in \Omega} s(p,\cdot)(\omega) u(\omega)$.*

For an agent with imprecise probabilistic prior represented by $P \subseteq \mathbb{P}$, imprecision about supposition represented by $S \subseteq \mathbb{S}$, and utility function $u \in \mathbb{U}$, an SP-supervaluated SDT *enjoins the agent to prefer act A to act B iff $(\forall s \in S, p \in P)(V(s,p,u,A) > V(s,p,u,B))$.*

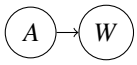
How might we represent this kind of decision theory with sets of desirable gambles? Well, the recipe I've cooked up for a bridge of Type 2 can be adapted, virtually without modification. We've shown, for any $s \in \mathbb{S}$, $p \in \mathbb{P}$, and $u \in \mathbb{U}$, that for any $A, B \in \mathcal{A}$, $V(s,p,u,A) > V(s,p,u,B)$ iff $g_A - g_B \in D_{s,p}$. So, $(\forall s \in S, p \in P)(V(s,p,u,A) > V(s,p,u,B))$ iff $g_A - g_B \in \bigcap_{s \in S, p \in P} D_{s,p}$ – which is just to say that $\bigcap_{s \in S, p \in P} D_{s,p}$ precisely encodes all of the recommendations of our SP-supervaluated SDT via a bridge of type 2. Just as when we were considering only supervaluation over P, $\bigcap_{s \in S, p \in P} D_{s,p}$ will not necessarily be coherent, but it can be made coherent in exactly the same way: let $\overline{D}_{SP} = \{g \in \mathcal{L}(\Omega) \colon g \in \mathcal{L}_{>0} \text{ or } \exists h \in \bigcap_{s \in S, p \in P} D_{s,p} : g \geq h\}$. I won't go through the proof (it's essentially identical to the proof of Theorem 4), but $\overline{D}_{SP}$ is the natural extension of $\bigcap_{s \in S, p \in P} D_{s,p}$. Just as

before, the only additional act preferences involve cases of Ω-dominance.

It's much less clear to me whether there's any sensible way of extending the idea of a bridge of Type 1 to the SP-supervaluated case.

10. A Simple Numerical Example: Extortion Revisited

Let's return to the motivating example and make it more concrete: our Agent believes that the threatening man (TM) will decide to break their windshield based on whether or not Agent pays. Suppose, for some reason, Agent is very precise about how likely TM is to break the windshield: they think that if they pay, TM will break the windshield with probability $\frac{1}{10}$; but if they don't, TM will break the windshield with probability $\frac{9}{10}$. We'll also assume that Agent is initially (almost) maximally imprecise about which act they will choose. (For simplicity, we ignore other factors that might cause the windshield to break.)[24] The causal structure of the relationship is given by this very simple graph:



The agent's choice of act is represented by the variable A, while W represents what happens to the windshield. $\mathcal{A} = \{P, DP\}$: Pay or Don't Pay; $\mathcal{W} = \{B, NB\}$: the windshield will be Broken or Not. But note this isn't *quite* a single causal Bayesian network (CBN) in the sense of [15, p. 24]; although Agent is certain of a single precise supposition rule, they have an imprecise prior over $\mathcal{A}$. As we will see, this is a case where EDT and CDT will generate the same supposition rule (restricted to supposing acts).[25] In this kind of case, we *can* build a bridge of Type 1 for both EDT and CDT (because the Type 1 model for CDT will also represent the judgments of EDT); the Type 2 models for EDT and CDT will also be the same.

Also for simplicity, we assume that Agent has linear utility in dollars, so that Agent's utility function is given by the table in Section 1.

10.1. Bridge Type 1

First, we find the coherent set of desirable gambles which represents Agent's prior beliefs. The agent's prior credal set, $\mathcal{M}$, is the set of all probabilities defined on $\Omega = \mathcal{A} \times \mathcal{W}$ which are *regular* (for all $\omega \in \Omega, p(\omega) > 0$) and which satisfy $p(B|DP) = \frac{9}{10}$ and $p(B|P) = \frac{1}{10}$ (because they are probabilities, they will also satisfy $p(NB|\cdot) = 1 - p(B|\cdot)$). Because of the regularity condition, the set of desirable gambles $D_{\mathcal{M}} = \{g \in \mathcal{L}(\Omega) : \forall p \in \mathcal{M}, E_p(g) > 0\}$ is already coherent. $D_{\mathcal{M}}$ takes a fairly simple form: $D_{\mathcal{M}} = \{g \in \mathcal{L}(\Omega) : \forall a \in (0,1), ag(P,B) + 9ag(P,NB) + (1-a)9g(DP,B) + (1-a)g(DP,NB) > 0\}$.

Next, we find the supposition operator and corresponding general imaging function dictated by the agent's beliefs about the causal structure. We can associate each precise $p \in \mathcal{M}$ with a CBN. For each $p \in \mathcal{M}$ we can define a set of interventional distributions, $\mathbf{S}^p \subset \mathbb{P}$ in the sense of [15, p. 24]: each $S^p_{\mathrm{do}(X=x)} \in \mathbf{S}^p$ represents a *causal intervention* setting the values of some subset of variables $X \subseteq \mathcal{V} = \{A, W\}$ to specific values x; changes to other variables "flow downstream": only the nodes intervened on and their descendants are affected by the intervention. This gives us a causal interventionist supposition operator: for any event E_x corresponding to some $X = x$, we define $s_{CDT}(p, E_x) = S^p_{\mathrm{do}(X=x)}$.

We are interested only in interventions on A, because this is the only variable over which the agent has direct causal control. Let's pick an arbitrary $p \in \mathcal{M}$ and consider the results of intervening on A. First, some notation: given variable V, let $PA(V)$ be the set of all Markovian parents of V. Let V_ω be the value assigned to variable V by total assignment ω; expressions like $PA(V)_\omega$ also make sense, because we can treat a set of variables as a compound variable (e.g., a vector where each component represents one of the variables in the set). Let $\mathbb{I}_{V=v}(\omega)$ be the indicator function which returns 1 iff $V_\omega = v$ and 0 otherwise; below, we will often abbreviate this with $\mathbb{I}_v$.

Pearl's general formula for interventional distributions is [15, Eq. 1.37]: for any total assignment $\omega \in \Omega$ consistent with the partial assignment $X = x$: $S^p_{\mathrm{do}(X=x)}(\omega) = \prod_{V \notin X} p(V_\omega | PA(V)_\omega)$; for ω inconsistent with x, $S^p_{\mathrm{do}(X=x)}(\omega) = 0$. Applying this formula to our very simple case, we have $S^p_{\mathrm{do}(A=a)}(\omega) = \mathbb{I}_a(\omega) p(W_\omega | A_\omega)$.

So, even though the agent's prior isn't described by a single precise CBN, there's a single causal supposition operator for acts: $s_{CDT}(p, A = a)(\omega') = \mathbb{I}_a(\omega') p(W_{\omega'} | a)$.[26] In this very simple case, intervening to set $A = a$ gives the same result as Bayesian conditionalization on $A = a$, which is why EDT and CDT make the same judgments.

To find the general imaging function which generates this supposition rule, I'll use Gallow's recipe for constructing imaging functions from interventions on a CBN [4, Appendix]. Let $ND(A)$ be the set of all Markovian non-descendants of A, $DE(A)$ the set of

[24] In the supplement to the paper, I have included a more complicated version of this case which might be more reasonable in various ways.

[25] [19] and [3] both contain very nice discussions of many famous cases where EDT and CDT come apart, also adding FDT into the mix.

[26] In this example, intervening on A washes out *all* imprecision in the agent's prior; every $p \in \mathcal{M}$ agrees about the conditional probabilities of the form $p(w|a), w \in \mathcal{W}, a \in \mathcal{A}$. I give an example of a case with a single supposition operator and more genuinely imprecise "posterior" (suppositional) beliefs in the supplement.

all Markovian descendants of A *except for* A *itself*. Then $f(A = a, \omega)(\omega') = \mathbb{I}_a(\omega') \cdot \prod_{N \in ND(A)} \mathbb{I}_{N_\omega}(\omega') \cdot \prod_{D \in DE(A)} p(D_{\omega'} | PA(D)_{\omega'})$. In this example, this recipe yields: $f(A = a, \omega)(\omega') = \mathbb{I}_a(\omega') p(W_{\omega'} | A_{\omega'})$; it's trivial to verify that $s_{CDT}(p, A = a)(\omega') = \mathbb{I}_a(\omega') p(W_{\omega'} | a) = \sum_{\omega \in \Omega} p(\omega) f(A = a, \omega)(\omega')$.

With the general imaging function in hand, we can construct the characteristic gambles of the acts for a bridge of Type 1: $g(u, A = P)(\omega') = \sum_{\omega \in \Omega} f(A = P, \omega')(\omega) u(\omega)$. Using the dollar values from Section 1 as Agent's utilities, we obtain: $g(u, P)(\omega') = u(P, B)p(B|P) + u(P, NB)p(NB|P) = -410 \cdot \frac{1}{10} - 10 \cdot \frac{9}{10} = -50$. Note that this is a constant gamble on Ω! Similarly, we find $g(u, DP) = u(DP, B)p(B|DP) + u(DP, NB)p(NB|DP) = -400 \cdot \frac{9}{10} - 0 = -360$.

That both gambles are negative is indicative of the fact that Agent is being forced to "choose the lesser of two evils"; alas, Agent cannot choose the status quo. Agent prefers P to DP iff $g(u, P) - g(u, DP) \in D_{\mathcal{M}}$. With the utility function given, $g(u, P) - g(u, DP) = -50 - -360 = 310$ and $D_{\mathcal{M}}$ includes $\mathcal{L}_{>0}(\Omega)$, so both CDT and EDT recommend Pay over Don't Pay. More generally, we could leave $g(u, P) - g(u, DP)$ as a function of utility, and solve for all utility functions which, given Agent's prior as encoded by $D_{\mathcal{M}}$ and causal beliefs as encoded by f, recommend Pay over Don't.

10.2. Bridge Type 2

We've already worked out the supposition operator (for both EDT and CDT; they are the same for this example), so this will be much quicker. Let's pick an arbitrary $p \in \mathcal{M}$ and find p_{eff}. Recall, from Section 6, the definition: $p_{eff}(\omega) = p_{eff}((A_\omega, X_\omega)) = \frac{s(p, A_\omega)(\omega)}{||\mathcal{A}||}$. $||\mathcal{A}|| = 2$ and $s(p, A = a)(\omega) = \mathbb{I}_a(\omega) p(W_\omega | a)$, so $p_{eff}(\omega) = \frac{\mathbb{I}_{A_\omega}(\omega)}{2} p(W_\omega | A_\omega) = \frac{p(W_\omega | A_\omega)}{2}$. Note again that one simple feature of this case is that all probabilities in $\mathcal{M}$ agree about the conditional probabilities of what happens to the windshield given whether the agent pays or not. So it happens that P_{eff} is a singleton, with $P_{eff} = \{q\}$ and

$$q(\omega) = \begin{cases} \frac{1}{20}, & (P, B) \\ \frac{9}{20}, & (P, NB) \\ \frac{9}{20}, & (DP, B) \\ \frac{1}{20}, & (DP, NB) \end{cases} \tag{2}$$

We find $D_{P_{eff}} = D_q = \{g \in \mathcal{L}(\Omega) \colon E_q(g) > 0\}$. $D_q = \{g \in \mathcal{L}(\Omega) \colon g(P, B) + 9g(P, NB) + 9g(DP, B) + g(DP, NB) > 0\}$. Much like $D_{\mathcal{M}}$ in the previous subsection, the regularity of q means that D_q is already coherent.

For bridge Type 2, the characteristic gambles take the much more familiar form from Eq. 1.

$$g_P(\omega) = \begin{cases} -410, & (P, B) \\ -10, & (P, NB) \\ 0, & (DP, \cdot) \end{cases} \tag{3}$$

.

$$g_{DP}(\omega) = \begin{cases} 0, & (P, \cdot) \\ -400, & (DP, B) \\ 0, & (DP, NB) \end{cases} \tag{4}$$

.

$$g_P - g_{DP} = \begin{cases} -410, & (P, B) \\ -10, & (P, NB) \\ 400, & (DP, B) \\ 0, & (DP, NB) \end{cases} \tag{5}$$

.

Agent prefers P to DP iff $g_P - g_{DP} \in D_q$; we find $-410 - 10 \cdot 9 + 400 \cdot 9 - 0 = 3100 > 0$. Both EDT and CDT recommend Pay over Don't Pay. And, like in the previous section, we could leave $g_P - g_{DP}$ as a function of utility and solve for all utility functions such that $g_P - g_{DP} \in D_q$; in both cases, we would get the same result.

11. Conclusion

We have seen that it is indeed possible to represent any suppositional decision theory with sets of desirable gambles, while modeling both imprecise credences and even uncertainty about what supposition procedure the agent should use. In the special case where the SDT is representable by general imaging, we have seen that a special representation (bridge of type 1) is possible, although it is less clear how to extend this to cases of uncertainty about the supposition procedure.

There are many topics I was not able to cover in this paper that I would be interested in exploring in future work, including: how to represent mixed acts and deliberational dynamics; sequential choice; learning and updating; and what this formalism yields in cases of exotic choice.

Acknowledgments

This research is part of the Epistemic Utility for Imprecise Probability project which is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 852677). I would like to thank my colleagues at the University of Bristol for helpful feedback from presentation of a version of this paper at a Departmental Work in Progress session. Special thanks to Arthur Van Camp and Jason Konek for helpful discussion; special thanks to Snow Zhang for sending me her Note on MEV and Linearity,

which includes a proof of [1, Theorem 3]. Thank you to the anonymous reviewers, whose helpful comments have improved the paper substantially.

References

[1] Andrew Bacon. Actual value in decision theory. *Analysis*, forthcoming. doi:10.1093/analys/anac014.

[2] Gert de Cooman and Erik Quaeghebeur. Exchangeability and sets of desirable gambles. *International Journal of Approximate Reasoning*, 53(3):363–395, 2012. Special issue in honour of Henry E. Kyburg, Jr.

[3] Kenny Easwaran. A classification of newcomb problems and decision theories. *Synthese*, 198(Suppl 27): 6415–6434, 2019. doi:10.1007/s11229-019-02272-z.

[4] J. Dmitri Gallow. Decision and foreknowledge. *Noûs*, forthcoming. doi:10.1111/nous.12443.

[5] Peter Gärdenfors. Imaging and conditionalization. *The Journal of Philosophy*, 79(12):747–760, 1982. ISSN 0022362X. URL http://www.jstor.org/stable/2026039.

[6] Allan Gibbard and William L. Harper. Counterfactuals and two kinds of expected utility. In A. Hooker, J. J. Leach, and E. F. McClennen, editors, *Foundations and Applications of Decision Theory*, pages 125–162. D. Reidel, 1978.

[7] Alan Hájek. Deliberation welcomes prediction. *Episteme*, 13(4):507–528, 2016. doi:10.1017/epi.2016.27.

[8] Richard C. Jeffrey. *The Logic of Decision (2nd ed.)*. University of Chicago Press, 1983.

[9] James M. Joyce. *The Foundations of Causal Decision Theory*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge University Press, 1999. doi:10.1017/CBO9780511498497.

[10] James M. Joyce. Levi on causal decision theory and the possibility of predicting one's own actions. *Philosophical Studies*, 110(1):69–102, 2002. doi:10.1023/a:1019839429878.

[11] David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315, 1976. ISSN 00318108, 15581470. URL http://www.jstor.org/stable/2184045.

[12] David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30, 1981. doi:10.1080/00048408112340011.

[13] Yang Liu and Huw Price. Ramsey and joyce on deliberation and prediction. *Synthese*, 197:4365–4386, 2020. doi:10.1007/s11229-018-01926-8.

[14] R. Duncan Luce and David H. Krantz. Conditional expected utility. *Econometrica*, 39(2):253–271, 1971. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913344.

[15] Judea Pearl. *Causality: Models, Reasoning and Inference (2nd ed.)*. Cambridge University Press, 2009.

[16] Wlodek Rabinowicz. Does practical deliberation crowd out self-prediction? *Erkenntnis*, 57(1):91–122, 2002. doi:10.1023/a:1020106622032.

[17] Wlodek Rabinowicz. Letters from long ago: On causal decision theory and centered chances. In Lars-Göran Johansson, Jan Österberg, and Rysiek Sliwinski, editors, *Logic, Ethics and All That Jazz: Essays in Honour of Jordan Howard Sobel*, pages 247–273. 2009.

[18] P.M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007. ISSN 0888-613X. doi:https://doi.org/10.1016/j.ijar.2006.07.019. URL https://www.sciencedirect.com/science/article/pii/S0888613X06001034. Reasoning with Imprecise Probabilities.

[19] Eliezer Yudkowsky and Nate Soares. Functional decision theory: A new theory of instrumental rationality, 2017. URL https://arxiv.org/abs/1710.05060.

[20] Marco Zaffalon and Enrique Miranda. The sure thing. In Andrés Cano, Jasper De Bock, Enrique Miranda, and Serafín Moral, editors, *Proceedings of the Twelveth International Symposium on Imprecise Probability: Theories and Applications*, volume 147 of *Proceedings of Machine Learning Research*, pages 342–351. PMLR, 06–09 Jul 2021. URL https://proceedings.mlr.press/v147/zaffalon21a.html.

[21] Jiji Zhang, Teddy Seidenfeld, and Hailin Liu. Subjective causal networks and indeterminate suppositional credences. *Synthese*, 198(Suppl 27):6571–6597, 2019. doi:10.1007/s11229-019-02512-2.