Supplement for "Representing Suppositional Decision Theories with Sets of Desirable Gambles"

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

Abstract

This supplement to my paper includes some proofs of theorems that I don't have room for in the main text and an additional example.

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

1. Proofs

The proofs of the following two Theorems, 5 and 6, were included in my original submission, but I've had to move them here to make space for the example in Section 10 of the paper.

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

Proof $\rightarrow$: assuming $h \in D_{s,q}$, show $h \in D_{q_{eff}}$. $h \in D_{s,q}$ iff there is some $n \in \mathbb{N}$ such that $h = \sum_{j=1}^{n} \lambda_j g_j$, where for every $j \in 1:n$, $\lambda_j \in \mathbb{R}_{>0}$ and $g_j \in D_{uni}$ or $g_j \in \bigcup_i \mathbb{I}_{A_i} D_i^q$. Recall that q_{eff} is the probability function which has $q_{eff}(\cdot|A) = s(q,A)(\cdot)$, for all $A \in \mathcal{A}$, and is uniform over the acts themselves. Then, clearly, $\forall j \in 1:n$, $E_{q_{eff}}(g_j) > 0$. And $E_{q_{eff}}(h) = \sum_{j=1}^{n} \lambda_j E_q(g_j) > 0$. So, $h \in D_{q_{eff}}$.

$\leftarrow$: assuming $h \in D_{q_{eff}}$, show $h \in D_{s,q}$. $h \in D_{q_{eff}}$ iff $E_{q_{eff}}(h) > 0$. So there is also some $\epsilon \in \mathbb{R}_{>0}$ such that $E_{q_{eff}}(h-\epsilon) > 0$. Observe that we can write $h = \sum_{i=1}^{m} \mathbb{I}_{A_i} h$. (Recall that m is the number of acts in $\mathcal{A}$.) Let $g = \sum_{i=1}^{m} \mathbb{I}_{A_i} E_{q_{eff}}(h((A_i,\cdot)) - \epsilon)$; g is a gamble which only depends on $\mathcal{A}$ and $\sum_{i=1}^{m} g(A_i) = E_{q_{eff}}(h-\epsilon) > 0$, so $g \in D_{uni}$. $E_{s(p,A_i)}(g((A_i,\cdot))) = E_{s(p,A_i)}(h((A_i,\cdot))) - \epsilon$; so, $\forall i \in 1:m$, $E_{s(p,A_i)}((h-g)((A_i,\cdot))) = \epsilon > 0$ and thus $(h-g)((A_i,\cdot)) \in D_i^q$. Finally, $h = g + \sum_{i=1}^{m} \mathbb{I}_{A_i}(h-g)((A_i,\cdot))$; $g \in D_{uni}$ and $\forall i \in 1:m$, $(h-g)((A_i,\cdot)) \in D_i^q$. So, $h \in D_{s,q} = \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i^q \cup D_{uni})$. ■

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 $V(p,u,B) < 0$.*[1]

Proof $\leftarrow$, part 1: assuming (1), show $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$. Let $h = \mathbb{I}_A \epsilon_1 - \mathbb{I}_B \epsilon_2$; by hypothesis, $\epsilon_1 - \epsilon_2 > 0$, so $h \in D_{uni}$. (In what follows, note that both g_1 and g_2 are intended to be read as gambles on $\mathcal{X}$.) Let $g_1 = g_A((A,\cdot)) - \epsilon_1$; by hypothesis, $\forall p \in P, E_{s(p,A)}(g_A) > \epsilon_1$, so $g_1 \in D_A$. Let $g_2 = \epsilon_2 - g_B((B,\cdot))$; by hypothesis, $\forall p \in P, E_{s(p,B)}(g_B) < \epsilon_2$, so $g_2 \in D_B$. And $g_A - g_B = h + \mathbb{I}_A g_1 + \mathbb{I}_B g_2$, so $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$.

$\leftarrow$, part 2: assuming (2), show $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$. $g_A = 0$, so $g_A - g_B = -g_B$. $\forall p \in P$, $V(p,u,B) = E_{s(p,B)(g_B)} < 0$, so $-g_B((B,\cdot)) \in D_B$. $g_A - g_B = -\mathbb{I}_B g_B((B,\cdot))$, so $g_A - g_B \in \mathbb{I}_B D_B \subset \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$.

$\leftarrow$, part 3: assuming (3), show $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$. $g_B = 0$, so $g_A - g_B = g_A$. $\forall p \in P$, $V(p,u,A) = E_{s(p,A)(g_A)} > 0$, so $g_A((A,\cdot)) \in D_A$. $g_A - g_B = \mathbb{I}_A g_A((A,\cdot)) \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$.

$\leftarrow$, part 4: assuming (4), show $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$. $\forall p \in P$, $V(p,u,A) = E_{s(p,A)(g_A)} > 0$, so $g_A((A,\cdot)) \in D_A$; $\forall p \in P$, $V(p,u,B) = E_{s(p,B)}(g_B) < 0$, so $-g_B((B,\cdot)) \in D_B$. $g_A - g_B = \mathbb{I}_A g_A((A,\cdot)) - \mathbb{I}_B g_B((B,\cdot)) \in \text{posi}(\mathbb{I}_A D_A \cup \mathbb{I}_B D_B) \subset \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$.

$\rightarrow$: assuming $g_A - g_B \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$, show that one of (1), (2), (3), or (4) holds. For any $h \in \mathcal{L}(\Omega)$, $g \in \text{posi}(\bigcup_i \mathbb{I}_{A_i} D_i \cup D_{uni})$ iff there is some $n \in \mathbb{N}$ such that $h = \sum_{j=1}^{n} \lambda_j g_j$, where for every $j \in 1:n$, $\lambda_j \in \mathbb{R}_{>0}$ and $g_j \in D_{uni}$ or $g_j \in \bigcup_i \mathbb{I}_{A_i} D_i$. First, observe that because each of the D_i and D_{uni} are, respectively, closed under posi, any gamble that can be represented this way can also be constructed by picking at most one gamble from D_{uni} and at most one gamble from each of the D_i – we also don't need to scale them; viz., $h = \sum_{k=1}^{m+1} \lambda_k g_k$, with $g_{m+1} \in D_{uni}$, $(\forall k \in 1:m)(g_k \in D_k)$, $(\forall k \in 1:m+1)\lambda_k \in \{1,0\}$, and $(\exists k \in 1:m+1)\lambda_k = 1$. For $h = g_A - g_B$ in particular, observe that $h(\omega) = 0$ for any $\omega \notin A \cup B$; this follows from the fact that g_A and g_B are both characteristic gambles of their respective acts. Notice also that it is pointless to consider cases where $g_{m+1}(A_i) \neq 0$ for any A_i distinct from

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

both A and B. If $g_{m+1}(A_i) > 0$ (for some A_i that isn't A or B), then to make $h((A_i, \cdot)) = 0$, we would need to add a negative g_i; but D_i is consistent, so that's impossible. If $g_{m+1}(A_i) = a < 0$, then we can certainly find a positive g_i which zeroes it out; but there's no need to, because $g^*_{m+1} = g_{m+1} - \mathbb{1}_{A_i} a$ is itself in D_{uni} anyway. So, $g_A - g_B = \lambda_A \mathbb{1}_A h_A + \lambda_B \mathbb{1}_B h_B + \lambda_{uni} h_{uni}$, with $\lambda_A, \lambda_B, \lambda_{uni} \in \{1, 0\}$ and at least one of them equal to 1, $g_A \in D_A$, etc. I'll divide the proof now into 2 cases: $\lambda_{uni} = 0$ or $\lambda_{uni} = 1$.

Suppose $\lambda_{uni} = 0$. Then at most one of $\lambda_A \mathbb{1}_A h_A$, $\lambda_B \mathbb{1}_B h_B$ can be zero. Suppose $\lambda_B \mathbb{1}_B h_B = 0$; then we have $g_B = 0$ and $g_A = \mathbb{1}_A h_A$. So (3) holds: $g_B = 0$ and $\forall p \in P$, $V(p, u, A) = E_{s(p,A)} g_A > 0$. Similarly, if $\lambda_A \mathbb{1}_A h_A = 0$, then $g_A = 0$ and $g_B = -\mathbb{1}_B h_B$; (2) holds. If $\lambda_A \mathbb{1}_A h_A$ and $\lambda_B \mathbb{1}_B h_B$ are both non-zero, then (4) holds: $g_A = \mathbb{1}_A h_A$ and $g_B = -\mathbb{1}_B h_B$; $\forall p \in P$, $V(p, u, A) = E_{s(p,A)} g_A > 0$; and $\forall p \in P$, $V(p, u, B) = E_{s(p,B)} g_B < 0$.

Suppose $\lambda_{uni} = 1$. $h_{uni} = \mathbb{1}_A \delta_1 + \mathbb{1}_B \delta_2$, with $\delta_1 + \delta_2 > 0$; this entails $\delta_1 > -\delta_2$. So there are $\epsilon_1, \epsilon_2 \in \mathbb{R}$: $\delta_1 > \epsilon_1 > \epsilon_2 > -\delta_2$. Then, $\forall p \in P$, $E_{s(p,B)}(g_B) = E_{s(p,B)}(-\lambda_B h_B - \delta_2) \leq -\delta_2 < \epsilon_2$ (recall that $\forall p \in P$, $E_{s(p,B)}(h_B) > 0$, so $E_{s(p,B)}(-\lambda_B h_B) \leq 0$); $\forall p \in P$, $E_{s(p,A)}(g_A) = E_{s(p,A)}(\lambda_A h_A + \delta_1) \geq \delta_1 > \epsilon_1$. (3) holds. ■

2. Extortion with More Complications

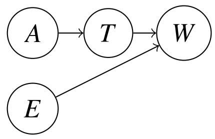
In Section 10, I develop a version of the motivating Extortion example from the beginning of the paper where we arbitrarily assume that Agent assigns precise conditional probabilities about what happens to their windshield when the Pay or Don't Pay. As I noted, this is not very natural – I just wanted to give an example, nearly as simple as possible, of how to construct both types of bridges.

Here, I'll work through a version of the case where Agent has much more imprecise suppositional beliefs. The causal structure of this example will be a little more complicated than the one in the paper, but it will still be a case where EDT and CDT ultimately agree.

As before, Agent believes that the threatening man (TM) will decide to break their windshield based on whether or not Agent pays. Suppose Agent believes: if they Pay, there's at least a 55% chance TM will want the windshield to stay unbroken; if they Don't Pay, there's at least a 55% chance TM will want the windshield to get broken. Agent thinks there's some chance that someone or something else in the neighborhood might break their windshield for reasons that are causally independent of whether they Pay or Don't; they don't have any precise estimate of how likely that is to happen, but they're extremely confident that the combined probability is less than 25%. To not make things *too* complicated: let's assume that whether TM wants the windshield to be broken doesn't depend on whether something else breaks it; if TM wants the windshield to be broken, it will be (with probability 1); and if TM doesn't want the windshield to be broken, it won't be unless something else breaks it (with probability 1).

As before, we assume that Agent initially has no idea which act they will choose; once again, we only consider probabilities in the open interval $(0, 1)$ for each act. (To properly handle allowing the true vacuous model for acts, we would need to define "primitive" conditional probabilities that can be defined even when the ratio "definition" of conditional probability – $p(A|B) = \frac{p(A \wedge B)}{p(B)}$ – is undefined. To avoid this complication, we continue to restrict to cases where the ratio formula is well-defined.)

This causal structure is given by this graph:[2]



The agent's choice of act is represented by the variable A, with $\mathcal{A} = \{P, D\}$: Pay or Don't; T represents what TM wants to happen to the windshield, with $\mathcal{T} = \{TB, TU\}$: TM want it Broken or TM wants it Unbroken; E represents something/someone else breaking the windshield, with $\mathcal{E} = \{Y, N\}$: Yes or No; and W represents what will happen to the windshield, with $\mathcal{W} = \{B, U\}$: Broken or Unbroken. This is not a single CBN: Agent's prior has imprecision concerning both A and E; more importantly, Agent has imprecise beliefs about how T impacts W. In this case, both EDT and CDT admit a single supposition operator (restricted to acts), which will also be identical. However, as we will see below, the single supposition operator conceals *different general imaging functions*, which means that it is not obvious how to construct a bridge of Type 1.

We assume that Agent has linear utility in dollars, and that the only relevant variables are, once again, A and W: Agent cares about the cost of paying, and they care about whether their windshield is broken; but they don't attach any direct utility to how TM feels about their windshield; and if their windshield is broken, they don't care whether it was by TM or something else. (Some of this might not be very realistic, but conveniently, it lets us reuse the same utility function as the example from the paper.)

To avoid repetition, in the next subsection, I construct the Type 2 model only for EDT; we will see in the following subsection that CDT admits the same supposition rule (restricted to acts and restricted to credence functions in the agent's prior), so this model is equally a Type 2 model for CDT.

[2]Note that this treats TM as not actually offering any "protection" at all, except from himself. I won't speculate about how reasonable an assumption that is, but it's what we're stipulating the agent believes.

2.1. Bridge Type 2

The total outcome space is $\Omega = \mathcal{A} \times \mathcal{X} = \mathcal{A} \times \mathcal{E} \times \mathcal{T} \times \mathcal{W}$.

Agent's initial credal set is $\mathcal{M} = \{p \in \mathbb{P} \colon p(P) \in (0,1), 0 < p(Y) < 0.25, p(TU|P) \geq 0.55, p(TB|D) \geq 0.55, p(B|TB) = 1, p(B|Y) = 1, p(B|TU,N) = 0\}$.

For EDT, supposition is just Bayesian conditionalization, so we have $s_{EDT}(p, A = a)(\omega') = \mathbb{I}_a(\omega')p(\omega'|A = a)$.

Every $p \in \mathcal{M}$ is Markov with respect to our causal graph, so for any $p \in \mathcal{M}$, we have $s_{EDT}(p, A = a)(\omega') = \mathbb{I}_a(\omega')p(E_{\omega'})p(T_{\omega'}|a)p(W_{\omega'}|E_{\omega'}, T_{\omega'})$.

$P_{eff} = \{q \in \mathbb{P} \colon (\exists p \in \mathcal{M})q(\omega) = q(A, X) = \frac{p(E_\omega)p(T_\omega|A_\omega)p(W_\omega|E_\omega,T_\omega)}{2}\}$. Put another way, every $q \in P_{eff}$ has this structure:

$$q(\omega) = \begin{cases} \frac{rs_1}{2}, & (P,Y,TU,B) \\ 0, & (P,Y,TU,U) \\ \frac{r(1-s_1)}{2}, & (P,Y,TB,B) \\ 0, & (P,Y,TB,U) \\ 0, & (P,N,TU,B) \\ \frac{(1-r)s_1}{2}, & (P,N,TU,U) \\ \frac{(1-r)(1-s_1)}{2}, & (P,N,TB,B) \\ 0, & (P,N,TB,U) \\ \frac{r(1-s_2)}{2}, & (D,Y,TU,B) \\ 0, & (D,Y,TU,U) \\ \frac{rs_2}{2}, & (D,Y,TB,B) \\ 0, & (D,Y,TB,U) \\ 0, & (D,N,TU,B) \\ \frac{(1-r)(1-s_2)}{2}, & (D,N,TU,U) \\ \frac{(1-r)s_2}{2}, & (D,N,TB,B) \\ 0, & (D,N,TB,U) \end{cases} \tag{1}$$

with $0 < r < 0.25$ and $0.55 \leq s_1, s_2 \leq 1$.

So, $D_{P_{eff}}$, which isn't coherent, is thus given by: $D_{P_{eff}} = \{g \in \mathcal{L}(\Omega) \colon (\forall (r, s_1, s_2 : 0 < r < 0.25, 0.55 < s_1, s_2 \leq 1)) rs_1 g(P,Y,TU,B) + r(1-s_1)g(P,Y,TB,B) + (1-r)s_1 g(P,N,TU,U) + (1-r)(1-s_1)g(P,N,TB,B) + r(1-s_2)g(D,Y,TU,B) + rs_2 g(D,Y,TB,B) + (1-r)(1-s_2)g(D,N,TU,U) + (1-r)s_2 g(D,N,TB,B) > 0\}$.

The natural extension is $\overline{D}_{P_{eff}} = D_{P_{eff}} \cup \mathcal{L}_{>0}(\Omega)$.

Because we're assuming the agent has the same utility function as the example from the paper (they care directly only about the costs of paying and what happens to their windshield; and they attach the same values for both of these factors as given in the paper), the characteristic gambles are very similar to the ones from Section 10 of the paper proper, just extended to this larger outcome space:

$$g_P(\omega) = \begin{cases} -410, & A = P, W = B \\ -10, & A = P, W = U \\ 0, & A = D \end{cases} \tag{2}$$

.

$$g_D(\omega) = \begin{cases} 0, & A = P \\ -400, & A = D, W = B \\ 0, & A = D, W = U \end{cases} \tag{3}$$

.

$$g_P - g_D = \begin{cases} -410, & (P,Y,TU,B) \\ -10, & (P,Y,TU,U) \\ -410, & (P,Y,TB,B) \\ -10, & (P,Y,TB,U) \\ -410, & (P,N,TU,B) \\ -10, & (P,N,TU,U) \\ -410, & (P,N,TB,B) \\ -10, & (P,N,TB,U) \\ 400, & (D,Y,TU,B) \\ 0, & (D,Y,TU,U) \\ 400, & (D,Y,TB,B) \\ 0, & (D,Y,TB,U) \\ 400, & (D,N,TU,B) \\ 0, & (D,N,TU,U) \\ 400, & (D,N,TB,B) \\ 0, & (D,N,TB,U) \end{cases} \tag{4}$$

.

Agent prefers P to D iff $g_P - g_D \in D_{P_{eff}}$; $g_P - g_D$ is not a positive gamble, so we find this is true iff $-410r - 10(1-r)s_1 - 410(1-r)(1-s_1) + 400r(1-s_2) + 400s_2 > 0$ for every r, s_1, s_2 satisfying the constraints already discussed. This is true: the infimum of this expression is 20, which is the limit as $r \to 0.25$ with s_1, s_2 both taking the minimal allowed value 0.55. So, EDT recommends Pay over Don't Pay.

2.2. CDT; No Bridge Type 1

Note, as in the example in the paper, that each $p \in \mathcal{M}$ encodes exactly one of the causal hypotheses Agent is uncertain between (viz., Agent's uncertainty about how TM will respond to being paid or not; Agent is certain about how W depends on E and T). [3]

For this reason, we can also identify a single supposition operator for CDT. If we apply Pearl's "truncated factorization" formula for calculating interventional distributions, we find the same result as for EDT in the previous subsection: $s_{CDT}(p, A = a)(\omega') = \mathbb{I}_a(\omega')p(E_{\omega'})p(T_{\omega'}|a)p(W_{\omega'}|E_{\omega'}, T_{\omega'})$.

[3] Note that the converse is not true: for any $p(TU|P)$ and $p(TB|D)$ Agent considers, there will be multiple priors over A and E that are included in $\mathcal{M}$. While supposing Agent performs a particular act will "wash out" Agent's prior uncertainty about acts, Agent will continue to be imprecise about the value of E.

However, there *isn't* a single general imaging function which generates this supposition rule. This is because the agent is uncertain about the causal impact of T on W. We could, of course, find a general imaging function corresponding to each of the conditional chances that Agent considers for how T might impact W; but this wouldn't be helpful in constructing a Type 1 bridge. The neat trick at the core of a Type 1 bridge is that the general imaging function for a single CBN depends only on the conditional chances assumed by the causal structure, which can come apart from the agent's prior. This lets us consider "characteristic gambles" encoding this causal dependence and the agent's utility function which are applicable, whatever the agent's prior might be. But in this case, different probability functions in $\mathcal{M}$ consider different conditional chances (although all are consistent with the same graph), so different elements of $\mathcal{M}$ are actually generated by distinct general imaging functions.