Death in Damascus

Ryan Doody

Indicative Conditional:

(1) If Shakespeare didn't write *Hamlet*, someone else did.

Subjunctive Conditional:

(2) If Shakespeare didn't write *Hamlet*, someone else would have.

Causal Decision Theory

The Newcomb Problem has led to the development of Causal Decision Theory, which doesn't define expected value in terms of *conditional probabilities* but rather *probabilities of (subjunctive) conditionals*.

This is sometimes called *Stalnaker's Equation*, after the philosopher Robert Stalnaker.

Causal Value: $U(\phi) = \sum_S c(\phi \mathrel{\Box\!\!\rightarrow} S) \cdot u(\phi \wedge S)$

Equivalently, we can compute the expected utility of actions using your *unconditional* credence in *dependency hypotheses*, K, which are maximally specific descriptions of the ways in which the things you care about might depend on what you do.

Lewis' Equation: $U(\phi) = \sum_K c(K) \cdot V(\phi \wedge K)$

Let's see how Causal Decision Theory is meant to work:

THE BIG TEST

	K_1	K_2	K_3	K_4
	$S \mathrel{\Box\!\!\rightarrow}$ PASS	$S \mathrel{\Box\!\!\rightarrow}$ FAIL	$S \mathrel{\Box\!\!\rightarrow}$ PASS	$S \mathrel{\Box\!\!\rightarrow}$ FAIL
	$P \mathrel{\Box\!\!\rightarrow}$ PASS	$P \mathrel{\Box\!\!\rightarrow}$ PASS	$P \mathrel{\Box\!\!\rightarrow}$ FAIL	$P \mathrel{\Box\!\!\rightarrow}$ FAIL
Study	20	0	20	0
Party	25	25	5	5

Notice that relative to the partition of dependency hypotheses ($\{K_1, K_2, K_3, K_4\}$), *Party* no longer dominates *Study*. In K_3, studying does better than partying. And if you think studying will cause you to pass, $c(K_3)$ should be high.

Causal Decision Theory: maximize U-value.

Death in Damascus

Consider the following situation:

> *Death in Damascus.* Death works from an appointment book that states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made weeks in advance on the basis of highly reliable predictions of your actions. An appointment for tomorrow has been inscribed for you; you know that it is either for Aleppo or for Damascus. You must decide now whether to stay in Damascus overnight, or ride to Aleppo to arrive tomorrow morning.

	DEATH IN ALEPPO	DEATH IN DAMASCUS
Go to Aleppo	0	10
Stay in Damascus	10	0

Suppose that $c(\text{DEATH IN X} \mid X) = 0.9$. Then,

$$\begin{aligned} V(\mathit{Aleppo}) &= 0.9 \cdot (0) + 0.1 \cdot (10) \\ &= 1 \\ V(\mathit{Damascus}) &= 0.1 \cdot (10) + 0.9 \cdot (0) \\ &= 1 \end{aligned}$$

So, according to **Evidential Decision Theory**, you should be indifferent between going to Aleppo and staying in Damascus.

Causal Decision Theory recommends going wherever you now think Death is least likely to be. But you going where you think Death is not is *evidence* that that's in fact where Death is. So, Causal Decision Theory creates *decision instability*.

What does **Causal Decision Theory** recommend?

If you're more confident that Death is in Aleppo, you should stay in Damascus. But if you think you'll stay in Damascus, that's good evidence that Death will be there. So, you should go to Aleppo. But if you think you'll go to Aleppo, that's good evidence that Death will be there. So, you should stay in Damascus. But . . .

Dicing with Death

Amend the story to include an additional option: the opportunity to flip a fair and indeterminate coin to decide between going to Aleppo and staying in Damascus. Death can reliably predict whether you'll decide to flip the coin, but Death is unable to predict the *result* of the flip—and so, if you flip the coin, Death will do no better than chance at telling where you will be tomorrow. You have to pay a small fee Δ to flip the coin. What should you do?

	Death in Aleppo		Death in Damascus	
	Heads	Tails	Heads	Tails
Go to Aleppo	0	0	10	10
Stay in Damascus	10	10	0	0
Randomize	$-\Delta$	$10-\Delta$	$10-\Delta$	$-\Delta$

Call the coin flip option *Randomize:*

$$= \begin{cases} \textit{Go to Aleppo} & \text{if Heads} \\ \textit{Stay in Damascus} & \text{if Tails} \end{cases}$$

Because Death cannot predict how the coin lands,

$$c(\textsc{Death in Aleppo} \mid \textit{Randomize}) = c(\textsc{Death in Damascus} \mid \textit{Randomize}) = 0.5$$

And so,

$$\begin{aligned} V(\textit{Randomize}) &= 0.25 \cdot (-\Delta) + 0.25 \cdot (10-\Delta) \\ &\quad + 0.25 \cdot (10-\Delta) + 0.25 \cdot (-\Delta) \\ &= 5 - \Delta \end{aligned}$$

And (so long as Δ is less than 4), because $5 - \Delta > 1$, $V(\textit{Randomize}) > V(\textit{Aleppo}) = V(\textit{Damascus})$.

Intuitively, you should pay the small fee to use the coin—doing so reduces Death from (a) an uncannily good predictor of your movements to (b) someone who can only randomly guess at them.

But Causal Decision Theory says otherwise: it's irrational to *Randomize*. Here's why:

> Suppose you think Death being either place is equally likely. Your decision of where to go in no way causally influences where Death will be. Furthermore, how the coin lands is independent of where Death is. So,
>
> $$\begin{aligned} U(\textit{Aleppo}) &= \sum_K c(K) \cdot V(\phi \wedge K) \\ &= 0.25 \cdot (0) + 0.25 \cdot (0) + 0.25 \cdot (10) + 0.25 \cdot (10) \\ &= 5 \end{aligned}$$
>
> $$\begin{aligned} U(\textit{Damascus}) &= \sum_K c(K) \cdot V(\phi \wedge K) \\ &= 0.25 \cdot (10) + 0.25 \cdot (10) + 0.25 \cdot (0) + 0.25 \cdot (0) \\ &= 5 \end{aligned}$$
>
> $$\begin{aligned} U(\textit{Randomize}) &= \sum_K c(K) \cdot V(\phi \wedge K) \\ &= 0.25 \cdot (-\Delta) + 0.25 \cdot (10-\Delta) + 0.25 \cdot (10-\Delta) + 0.25 \cdot (-\Delta) \\ &= 5 - \Delta \end{aligned}$$

$$\begin{aligned} c(\textsc{Death in Aleppo} \wedge H) &= 0.25 \\ c(\textsc{Death in Aleppo} \wedge T) &= 0.25 \\ c(\textsc{Death in Damascus} \wedge H) &= 0.25 \\ c(\textsc{Death in Damascus} \wedge T) &= 0.25 \end{aligned}$$

No matter how small the fee, Δ, Causal Decision Theory will recommend against *Randomization*. But isn't that absurd?

No matter your unconditional credences in where Death will be, $U(\textit{Randomize})$ will be lower than either (or both) $U(\textit{Aleppo})$ or $U(\textit{Damascus})$.

Let your unconditional credences in the four states be: p_1, p_2, p_3, p_4. The difference between state 1 and state 2, and between state 3 and state 4, is how the fair coin landed, so

$$p_1 = p_2 = \frac{p_1 + p_2}{2}$$
$$p_3 = p_4 = \frac{p_3 + p_4}{2}$$

And so,

$$\begin{aligned} U(\textit{Aleppo}) &= p_1 \cdot (0) + p_2 \cdot (0) \\ &\quad + p_3 \cdot (10) + p_4 \cdot (10) \\ &= 20 \cdot p_3 \end{aligned}$$

$$\begin{aligned} U(\textit{Damascus}) &= p_1 \cdot (10) + p_2 \cdot (10) \\ &\quad + p_3 \cdot (0) + p_4 \cdot (0) \\ &= 20 \cdot p_1 \end{aligned}$$

$$\begin{aligned} U(\textit{Randomize}) &= p_1 \cdot (-\Delta) + p_2 \cdot (10-\Delta) \\ &\quad + p_3 \cdot (10-\Delta) + p_4 \cdot (-\Delta) \\ &= (10 - 2\Delta) \cdot (p_1 + p_3) \end{aligned}$$

There are no values for p_1, p_2, and Δ, where *Randomize* comes out best.