Day 1: Introduction to Decision Theories

Conditionals, probability, and decision // *ESSLLI 2024*

Melissa Fusco and Matt Mandelkern

---

1 Bayesianism

Bayesianism is a family of views on which

1. rational agents have degrees of belief ("credences") that conform to the probability calculus;
2. rational agents update their credences by conditionalizing on what they learn.

(for 1:) Given a set W of *possible worlds* which determine a set $\wp(W)$ of *propositions* closed under $\{\wedge\neg\}$, let a credence function be a function that

- assigns each member $A \in \wp(W)$ a real number in [0,1], and is such that:
- if $A \Vdash B$ then $Pr(A) \leq Pr(B)$;
- $Pr(A \vee B) = Pr(A) + Pr(B) - Pr(A \wedge B)$;
- $Pr(\top) = 0$ and $Pr(\bot) = 1$.

These are *synchronic* features of an agent's credal state!

To obey these is to obey/observe *Probabilism.*

(for 2:) if Pr is the agent's credence function at t, and E is the entirety of evidence acquired between t and t^+, then the agent's credence function in arbitrary B at t^+ should be

$$Pr^+(B) = Pr(B \mid E) = \frac{Pr(B \wedge E)}{Pr(E)} = \frac{Pr(BE)}{Pr(E)}$$

Call the norm that tells you to update this way "Conditionalization". Obeying Conditionalization is a *diachronic* feature of an agent's credal state!

Visualization: cut & renormalize

→ What's an example of not updating on the *entirety* of one's evidence?

- Passport-at-the-bar case.

→ What's a subtlety of this clause about E being *the entirety*?

- College-envelopes case: is it $E = \{w_{\text{accept}}, w_{\text{reject}}\}$, or (the stronger) $E = \{w_{\text{reject}}\}$?

2 Setting and Measuring Credence

- The standard credence-betting bridge: if B is "ethically neutral" for you, then you will pay $n \times \$Pr(B)$ for a bet that pays $\$n$ if B and $\$0$ otherwise.

Why might B fail to be ethically neutral? (At least two case-types)

- It is standard to assume, *ceteris paribus*, that you are a representative member of a (large) population; more generally, it is reasonable to set credences in line with *frequences*.

'Reasonable', not 'rational'!

  - So, it is reasonable *ceteris paribus* to set *conditional* credences in line with *conditional* probabilities ("correlations").
  - Addition (**controversial!**): It is reasonable *ceteris paribus* to attribute stable *correlations* to causal relationships ("Reichenbach's Principle").

What's a universe where Reichenbach's Principle fails?

3 Decisions

We pair credence with *utility* in the calculation of *expected utility*.

> (Bike Insurance). You move to a new neighborhood with your bike (worth €100). Otto the insurance salesman suggests you buy insurance from him for €40. He points to the high number of bike thefts in the area.

This is a **decision matrix** for (Bike Insurance)

	no theft	*theft*
no insurance	€100	€0
insurance	€(100 − 40)	€(100 − 40)

Table 1: Matrix for (bike insurance).

with **acts** $A \in \boldsymbol{A}$ along the rows and **states** $S \in \boldsymbol{S}$ along the columns. We will understand all of these as partitions of propositions. At the intersection of each act and state, there is a utility value.

A naive equation for expected utility (EU):

The norm: choose some $A \in \boldsymbol{A}$ that maximizes $EU(A)$.

$$EU(A) = \sum_S Pr(S)Val(A \wedge S) \tag{1}$$

So e.g.

$$EU(\textit{no insurance}) = Pr(\textit{no theft})(€100) + Pr(\textit{theft})(€0)$$

$$\begin{aligned} EU(insurance) &= Pr(no\ theft)(€(100-40)) + Pr(theft)(€(100-40)) \\ &= [1 - Pr(theft)](€60) + Pr(theft)(€60) \\ &= €60 \end{aligned}$$

→ Is there a problem with *intrinsic enjoyment* of costs or fees?

– "At least I have peace of mind!"

Utilities are supposed to measure *non-instrumental* value.

→ can you write the standard bet on B for $n = €1$ as a decision matrix?

	B	$\overline{B}$
accept	€(1-$Pr(B)$)	$-€Pr(B)$
¬ *accept*	0	0

We add:

(Bike Insurance, pt. II). You believe theft (*theft*) is negatively correlated with purchasing insurance (*insurance*).

A second, more sophisticated equation for evidential expected utility ("*EEU*"):

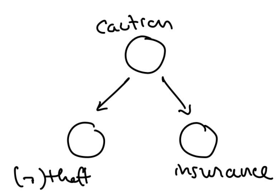
$$EEU(A) = \sum_S Pr(S \mid A) Val(A \wedge S) \tag{2}$$

In the present context:

$$Pr(theft \mid insurance) < Pr(theft \mid no\ insurance)$$

- Math fact (conglomerability): $Pr(theft \mid insurance) \leq Pr(theft) \leq Pr(theft \mid no\ insurance)$
- Another math fact (partition invariance): for any countable set $\{X_1, \ldots X_n\}$ that Pr-partitions W, if $X = \bigcup_i X_i$, then $EEU(X) = \sum_i Pr(X_i \mid X) Val(X_i)$.
  – "Partition invariance makes it possible to employ expected utility maximization in small-world decision making" (Joyce 1999, pg. 121)

(Bike Insurance, pt. III). Though you believe purchasing insurance (*insurance*) is negatively correlated with *theft*, you believe this **only because you believe there is a common cause—cautious people are less likely to expose their bikes to theft**.



- Screening-off: B screens off C from A iff, even though $Pr(C \mid A) > Pr(C)$, $Pr_B(C \mid A) = Pr_B(C)$.

→ does this entail that $\neg B$ screens of C from A?

Notation: $Pr_B(\cdot) = Pr(\cdot \mid B)$

Intuition: just be cautious! In this case, *cautious* $\in \boldsymbol{A}$, so you have control over it.
But what if the only control you have is correlational?

4 Newcomb Problems

Quoted from Ahmed (2018).

(Standard Newcomb) You must choose between taking (and keeping the contents of) (i) an opaque box now facing you or (ii) that same opaque box and a transparent box next to it containing \$1000. Yesterday, a being with an excellent track record of predicting human behaviour in this situation made a prediction about your choice. If it predicted that you would take only the opaque box ('one-boxing'), it placed \$1M in the opaque box. If it predicted that you would take both ('two-boxing'), it put nothing in the opaque box.

4.1 correlation vs. causation: two approaches

Suppose you wish to distinguish the case where A is merely correlated with (good outcome) S and the case where it is causally related.

1. K-partitions.
   Make the columns of decision matrix consist of propositions $\boldsymbol{K}$ over which you have no causal control; and maximize

$$CEU(A) = \sum_K Pr(K)Val(AK) \tag{3}$$

2. Counterfactuals/Imaging.
   The columns $\boldsymbol{S}$ of the decision matrix are anything you like (as before), but use one of:

'$Pr(A >_s S)$' is the probability of 'if A, would S'.

$$CEU(A) = \sum_S Pr(A >_s S)Val(AS) \tag{4}$$

'$Pr(S \mid\mid A)$' is the probability of S *imaged on* A.

$$CEU(A) = \sum_K Pr(S \mid\mid A)Val(AS) \tag{5}$$

- Lewis (1981) famously claimed all these approaches were equivalent.
- of note:
  - '$>_s$' is an object-language binary connective, which stands in need of a semantics.
  - '$\mid\mid$', like the '$\mid$' in '$Pr(S \mid A)$', is *not* an object-language connective of any kind, any more than '$\sum$' is.

What's a way of synthesizing the two approaches?

Imaging comes in two flavours: *sharp* and *blurred* (or *general*). Both require a *selection function* f, which takes a proposition and world as arguments.

When imaging is sharp, $f(\phi, w')$ is the unique world w to which w' wills its mass when Pr is imaged on the proposition X.

$$Pr(w \mid\mid X) := \begin{cases} 0 & \text{if } w \in \overline{X} \\ Pr(w) + \sum_{w' \in \overline{X} \mid w = \sigma(w', X)} Pr(w') & \text{if } w \in X \end{cases} \tag{6}$$

A common way of understanding the selection function f is that $f(\phi, w')$ is the *closest* or *most similar* world to w where X is true.

Similarity, however, admits of ties, as Gardenfors (1982, §1) notes. He thus defines $f(\phi, w)$ more generally as a *set* of worlds $Y \subseteq W$. The definition of general imaging additionally has recourse to a *transfer function* $T_{w,\phi} : \{v \in f(\phi, w)\} \to [0,1]$. For example, when $T_{\phi,u}(v) = .25$, then u sends exactly 25% of its probability mass to v when the probability space is imaged on ϕ. $Pr^X(w)$ is defined with the aid of $f(\cdot)$ and $T_{(\cdot)}$, as follows:

$$P^X(w) := \begin{cases} 0 & \text{if } w \in \overline{X} \\ P(w) + \sum_{w' \in \overline{X} | w \in f(X,w')} P(w') \cdot T_{w',X}(w) & \text{if } w \in X \end{cases} \quad (7)$$

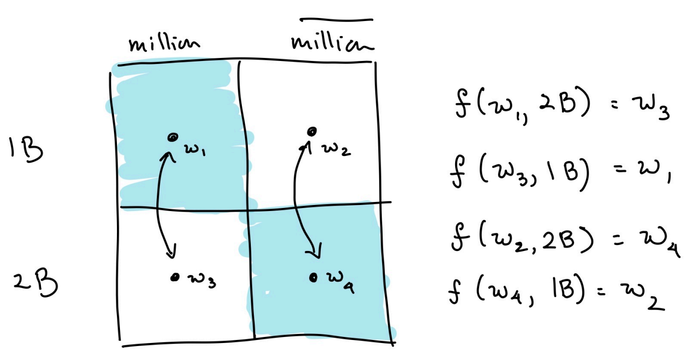
For any world w' and proposition X, we assume *at least*:

- **Success:** $f(X, w') \subseteq X$
- **Strong Centering:** if $w' \in X$, then $f(X, w') = \{w'\}$

When a world w "dies" under imaging, $\sigma(\cdot)$ (and $T_{(\cdot)}$) record how it bequeaths its probability mass to its survivors. w may dole out this mass unequally; the only requirement is that "it all goes somewhere": $\sum_{w' \in f(X,w)} T_{w,X}(w') = 1$. For this reason, Lewis influentially described imaging as a process according to which probability "is moved around" though it is "neither created nor destroyed" (1976, pg. 310).

The presumptive contrast is that when a world "dies" under conditionalisation, probability mass *is* destroyed.

Here is a picture of how imaging is standardly taken to work in Newcomb's Problem. The relevant intuition is that even if $f(A, w) \neq w$ for $A \in \boldsymbol{A}$, $f(A, w)$ is in the same K-cell as w.



The upshot—the only one that often makes its way into the decision-theory literature—is that for problems that feature correlation without causation, $Pr(S \mid A) > Pr(S)$, but $Pr(S \mid\mid A) = Pr(S)$.

Here, in Newcomb's Problem: $Pr(million \mid 1B) > Pr(million)$, but $Pr(million \mid\mid 1B) = Pr(million)$.

References

Ahmed, A. (2018). Introduction. In Ahmed, A., editor, *Newcomb's Problem*. Cambridge University Press.

Gardenfors, P. (1982). Imaging and conditionalization. *Journal of Philosophy*, 79(12):747–760.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315.

Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30.