Some Rational Constraints on Self-Locating Belief

Cian Dorr

Oxford, 25 January 2008

1 A Bayesian framework

Necessarily, if one has ideally rational belief-forming dispositions, there is a probability distribution **C** (one's "priors") and a function **E** from times to true propositions (one's "evidence function"), such that one is disposed to have as one's credence distribution at any time t the result of conditionalising **C** on $\mathbf{E}(t)$.

2 What kind of diachronic coherence is involved in having ideal dispositions?

> *Symmetric Conditionalisation (SC)* Ideally: whenever one has credences Cr and evidence-function **E**, $\mathrm{Cr}_{t+\Delta}(A|\mathbf{E}(t)) = \mathrm{Cr}_t(A|\mathbf{E}(t+\Delta))$

3 Irreducibly self-locating belief

"$A^{(\Delta)}$" $=_{df}$ "A will be true of me Δ units of time from now / was true of me $-\Delta$ units of time ago"

> *Self-locating Symmetric Conditionalisation (SSC)* Ideally: whenever one has credences Cr and evidence-function **E**, $\mathrm{Cr}_{t+\Delta}(A^{(-\Delta)}|\mathbf{E}(t)^{(-\Delta)}) = \mathrm{Cr}_t(A|\mathbf{E}(t+\Delta)^{(\Delta)})$

4 What SSC rules out

(i) *An unreasonable way of caring about your weight.*

(ii) *The Sleeping Beauty problem.*

(1) $\mathrm{Cr}_{\mathrm{Sunday}}(\mathrm{Heads}) = 1/2$ [premise]

(2) $\mathrm{Cr}_{\mathrm{Sunday}}(\mathrm{Heads}|\mathbf{E}(\mathrm{Monday})^{(1)}) = 1/2$ [1]

(3) $\mathrm{Cr}_{\mathrm{Monday}}(\mathrm{Heads}^{(-1)}|\mathbf{E}(\mathrm{Sunday})^{(-1)}) = 1/2$ [2, SSC]

(4) $\mathrm{Cr}_{\mathrm{Monday}}(\mathrm{Heads}|\text{It's Monday}) = 1/2$ [3]

(5) $\mathrm{Cr}_{\mathrm{Monday}}(\mathrm{Heads}|\text{It isn't Monday}) \approx 0$ [premise]

(6) $\mathrm{Cr}_{\mathrm{Monday}}(\text{It's Monday}) < 1$ [premise]

(7) $\mathrm{Cr}_{\mathrm{Monday}}(\mathrm{Heads}) < 1/2$ [4, 5, 6]

(iii) *Infinite coin tosses.*

Your life is infinite in both directions; every day you have exactly the same evidence; every day a coin is tossed; you never get to see the result.

$n\ldots$:	Heads today $+n$ and every subsequent day, but not on some previous day.
$\ldots n$:	Heads today $+n$ and every previous day, but not on some subsequent day.
Cr:	Your credence function (same each morning)
T:	$1\ldots \vee \ldots -1$

(1) $\mathrm{Cr}(1\ldots|1\ldots\vee\ldots 0) = \mathrm{Cr}(0\ldots|0\ldots\vee\ldots -1)$ [by SSC]

(2) $$\frac{\mathrm{Cr}(1\ldots|T)}{\mathrm{Cr}(1\ldots|T) + \mathrm{Cr}(\ldots 0|T)} = \frac{\mathrm{Cr}(0\ldots|T)}{\mathrm{Cr}(0\ldots|T) + \mathrm{Cr}(\ldots -1|T)} \quad [1]$$

(3) $\mathrm{Cr}(1\ldots|T)\,\mathrm{Cr}(\ldots-1|T) = \mathrm{Cr}(0\ldots|T)\,\mathrm{Cr}(\ldots 0|T)$ [2]

(4) $\mathrm{Cr}(1\ldots|T) > 0$ and $\mathrm{Cr}(\ldots-1|T) > 0$ [premise]

(5) $\mathrm{Cr}(0\ldots|T) = \mathrm{Cr}(1\ldots|T)$ and $\mathrm{Cr}(\ldots 0|T) = \mathrm{Cr}(\ldots-1|T)$ [3,4]

(6) $\mathrm{Cr}(0\ldots|1\ldots) = \mathrm{Cr}(\ldots 0|\ldots-1) = 1$

(iv) *"Branching worlds" interpretations of quantum mechanics.*

5 SSC is equivalent to a constraint on rational priors

[Evidential] Harmony Any ideally rational belief-forming dispositions can be represented by a $\langle \mathbf{C}, \mathbf{E}\rangle$ such that, for any A, X, Y [where X and Y could be values of $\mathbf{E}$ at two instants separated by Δ]: $\mathbf{C}(A|XY^{(\Delta)}) = \mathbf{C}(A^{(-\Delta)}|X^{(-\Delta)}Y)$.

Uniformity Any ideally rational belief-forming dispositions can be represented by a $\langle \mathbf{C}, \mathbf{E}\rangle$ such that, for any X, Y: $\mathbf{C}(XY^{(\Delta)}) = \mathbf{C}(X^{(-\Delta)}Y)$.

6 Units of time

Life at variable speed: You are to spend six seconds in a room with a digital clock. For one of the seconds—either the first or the last, you're not sure which—a special field will be turned on which will undetectably cause all physical processes in the room to happen five times faster.

If we understand "units of time" objectively, SSC seems to have the absurd consequence that when the clock reads five you should become much more confident that the field was turned on for the first second. We can avoid this by taking the units to measure Lewisian "personal time".

7 Generalising?

Notation: $A^{(R)} =_{df}$ 'A is true relative to the person-stage R-related to me-now'.

General Uniformity Any ideally rational belief-forming dispositions can be represented by a $\langle \mathbf{C}, \mathbf{E}\rangle$ such that, for any necessarily one-one relation R between person-stages and any X, Y, $\mathbf{C}(XY^{(R)}) = \mathbf{C}(X^{(R^{-1})}Y)$

Yellow Brick Road: Along the infinite Yellow Brick Road, every third house is red. Every house is always occupied, and every night the people are moved around in such a way that each person spends every second day in a red house. The houses are locked, so no-one can ever see what colour their house is. How confident should you be that you are in a red house?

8 What about hard cases for personal identity?

Tolerant Uniformity Any ideally rational belief-forming dispositions can be represented by a $\langle \mathbf{C}, \mathbf{E}\rangle$ such that, for some *admissible* family $\mathbf{F}$ of one-one relations between person-slices, for every $R \in \mathbf{F}$ and any X, Y, $\mathbf{C}(XY^{(R)}) = \mathbf{C}(X^{(R^{-1})}Y)$

9 Fission cases