Logic and Probabilistic Models of Belief Change

Eric Pacuit

Department of Philosophy
University of Maryland, College Park
pacuit.org

March 3, 2016

J. Joyce. *A nonpragmatic vindication of probabilism*. Philosophy of Science 65, 575603 (1998).

R. Pettigrew. *Epistemic Utility Arguments for Probabilism*. Stanford Encyclopedia of Philosophy, 2015.

R. Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, 2016.

Accuracy

Accuracy. An epistemic agent ought to approximate the truth. In other words, she ought to minimize her inaccuracy.

Accuracy

Accuracy (Synchronic expected local). An agent ought to minimize the **expected local inaccuracy** of her degrees of credence in all propositions $A \subseteq W$ by the lights of her current belief function, relative to a **legitimate local inaccuracy measure** and over the set of worlds that are currently epistemically possible for her.

Accuracy (Synchronic expected global). An agent ought to minimize the **expected global inaccuracy** of her current belief function by the lights of her current belief function, relative to a legitimate global inaccuracy measure and over the set of worlds that are currently epistemically possible for her

Measuring Inaccuracy

Alethic Vindication The ideal credence function at world w is the omniscient credence function at w, namely, v_w.

Perfectionism The accuracy of a credence function at a world is its proximity to the ideal credence function at that world.

Squared Euclidean Distance Distance between credence functions is measured by squared Euclidean distance.

B. de Finetti. *Theory of Probability*. John Wiley and Sons, 1974.

J. Pred, R. Seiringer, E. Lieb, D. Osherson, H. V. Poor, and S. Kulkarni. *Probabilistic Coherence and Proper Scoring Rules*. IEEE Transactions on Information Theory, 2009.

Consider a vector $\mathcal{E} = (E_1, \ldots, E_n)$ of events.

A **forecast** for $\mathcal{E}$ is a vector $\mathbf{f} = (f_1, \ldots, f_n)$.

Consider a vector $\mathcal{E} = (E_1, \ldots, E_n)$ of events.

A **forecast** for $\mathcal{E}$ is a vector $\mathbf{f} = (f_1, \ldots, f_n)$.

Two possible defects:

1. There may be a rival forecast $\mathbf{g}$ that guarantees a lower **penalty** than the one for $\mathbf{f}$, regardless of which events come to pass.
2. The events in $\mathcal{E}$ may be related by inclusion or partition and $\mathbf{f}$ might violate constraints imposed by probability.

Consider a vector $\mathcal{E} = (E_1, \ldots, E_n)$ of events.

A **forecast** for $\mathcal{E}$ is a vector $\mathbf{f} = (f_1, \ldots, f_n)$.

Two possible defects:

1. There may be a rival forecast $\mathbf{g}$ that guarantees a lower **penalty** than the one for $\mathbf{f}$, regardless of which events come to pass.
2. The events in $\mathcal{E}$ may be related by inclusion or partition and $\mathbf{f}$ might violate constraints imposed by probability.

de Finetti, Predd et al., Lindley, ...: The two defects are equivalent.

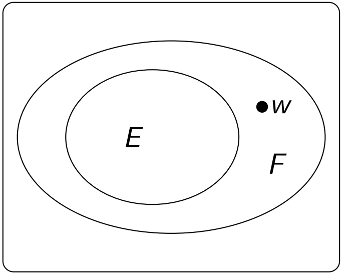
Brier Score

$$\mathcal{E} = (E, F) \text{ with } E \subseteq F$$
$$\mathbf{f} = (0.6, 0.9)$$

Brier Score

$$\mathcal{E} = (E, F) \text{ with } E \subseteq F$$

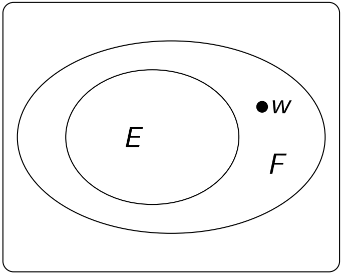
$$\mathbf{f} = (0.6, 0.9)$$



Brier Score

$$\mathcal{E} = (E, F) \text{ with } E \subseteq F$$

$$\mathbf{f} = (0.6, 0.9)$$



Penalty: $(0 - 0.6)^2 + (1 - 0.9)^2 = 0.37$

Proper Scoring Rule

$\mathcal{E} = (E, F)$ with $E \subseteq F$

$\mathbf{f} = (0.6, 0.9)$

Expected Penalty for E:
$0.6 * (1 - 0.6)^2 + 0.4 * (0 - 0.6)^2 = 0.230$

Expected Penalty for E **by lying**:
$0.6 * (1 - 0.65)^2 + 0.4 * (0 - 0.65)^2 = 0.2425$

Proper Scoring Rule

Suppose your probability for an event E is p, that your announced probability is x, and that your penalty assessed according ot the rule $(1-x)^2$ if E comes out true; $(0-x)^2$ otherwise. Then your expected penalty is uniquely minimized by choosing $x = p$.

Proper Scoring Rule

Suppose your probability for an event E is p, that your announced probability is x, and that your penalty assessed according ot the rule $(1-x)^2$ if E comes out true; $(0-x)^2$ otherwise. Then your expected penalty is uniquely minimized by choosing $x = p$.

Absolute Deviation

Expected Penalty for E:
$0.6 * |1 - 0.6| + 0.4 * |0 - 0.6| = 0.48$

Expected Penalty for E **by lying**:
$0.6 * |1 - 0.65| + 0.4 * |0 - 0.65| = 0.47$

$\mathcal{E} = (E, F)$ with $E \subseteq F$

$\mathbf{f} = (0.6, 0.9)$
$\mathbf{f}' = (0.95, 0.55)$

$\mathcal{E} = (E, F)$ with $E \subseteq F$

$\mathbf{f} = (0.6, 0.9)$
$\mathbf{f}' = (0.95, 0.55)$

Penalties:

Possibility	$\mathbf{f}$	$\mathbf{f}'$
E true, F true	0.17	0.205
E false, F true	0.37	1.105
E false, F false	1.17	1.205

S is a sample space. Subsets of S are events. Let $\mathcal{E} = (E_1, \ldots, E_n)$ be a vector of events.

S is a sample space. Subsets of S are events. Let $\mathcal{E} = (E_1, \ldots, E_n)$ be a vector of events.

A forecast is an element of $[0, 1]^n$. A forecast is **coherent** if there is a probability measure μ over S such that for all $i = 1, \ldots, n$, $f_i = \mu(E_i)$.

S is a sample space. Subsets of S are events. Let $\mathcal{E} = (E_1, \ldots, E_n)$ be a vector of events.

A forecast is an element of $[0, 1]^n$. A forecast is **coherent** if there is a probability measure μ over S such that for all $i = 1, \ldots, n$, $f_i = \mu(E_i)$.

A function $s : \{0, 1\} \times [0, 1] \to [0, \infty]$ is a **proper scoring rule** in case:

1. $ps(1, x) + (1 - p)s(0, x)$ si uniquely minimized at $x = p$ for $p \in [0, 1]$.
2. s is continuous. For $i \in \{0, 1\}$, $\lim_{n\to\infty} s(i, x_n) = s(i, x)$ for any sequence x_n from $[0, 1]$ converging to x.

Penalty

Given a proper scoring rule s, the penalty P_s based on s for forecast $\mathbf{f}$ and $w \in S$ is given by:

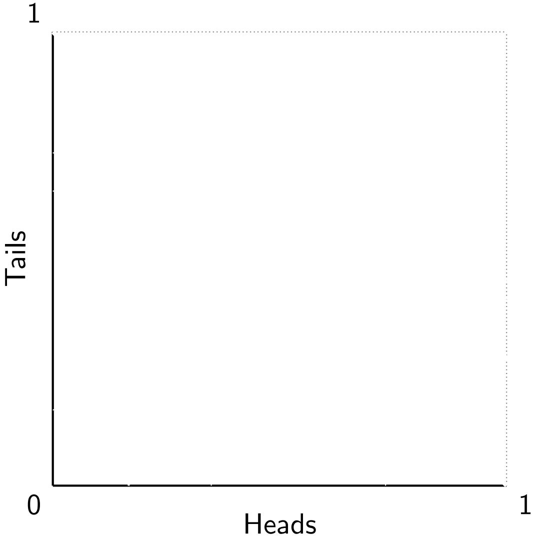
$$P_s(w, \mathbf{f}) = \sum_{i=1}^{n} s(\chi_{E_i}(w), f_i)$$

$\mathbf{f}$ is **weakly dominated** by $\mathbf{g}$ in case $P_s(w, \mathbf{g}) \leq P_s(w, \mathbf{f})$ for all $w \in S$.

$\mathbf{f}$ is **weakly dominated** by $\mathbf{g}$ in case $P_s(w, \mathbf{g}) < P_s(w, \mathbf{f})$ for all $w \in S$.

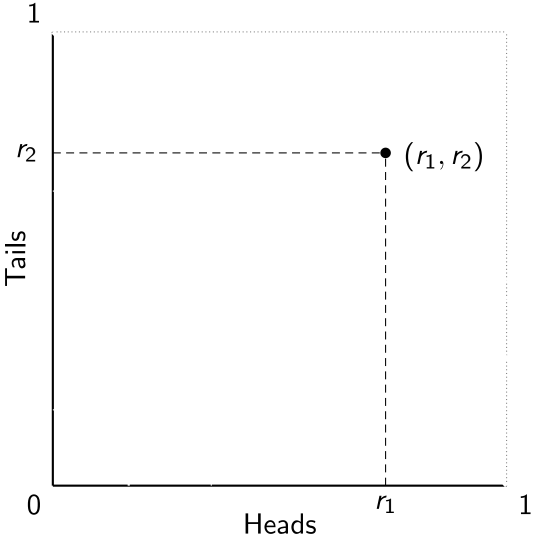
Theorem Let $\mathbf{f}$ be a forecast.

1. If $\mathbf{f}$ is coherent, then it is not weakly dominated by any forecast $\mathbf{g} \neq \mathbf{f}$
2. If $\mathbf{f}$ is incoherent, then it is strongly dominated by some coherent forecast $\mathbf{g}$



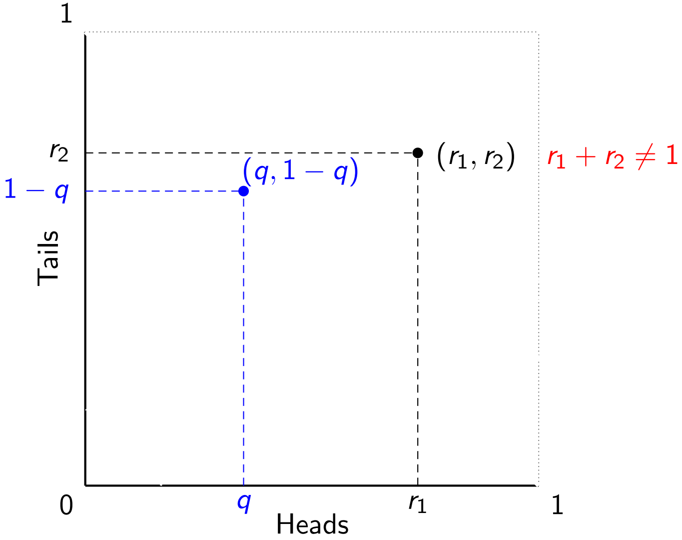
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



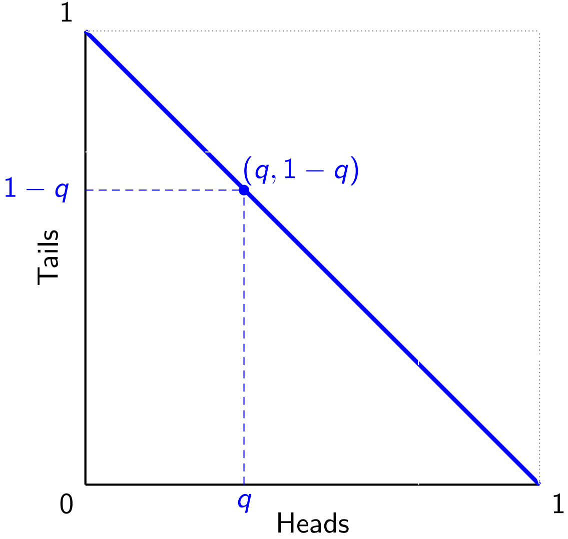
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



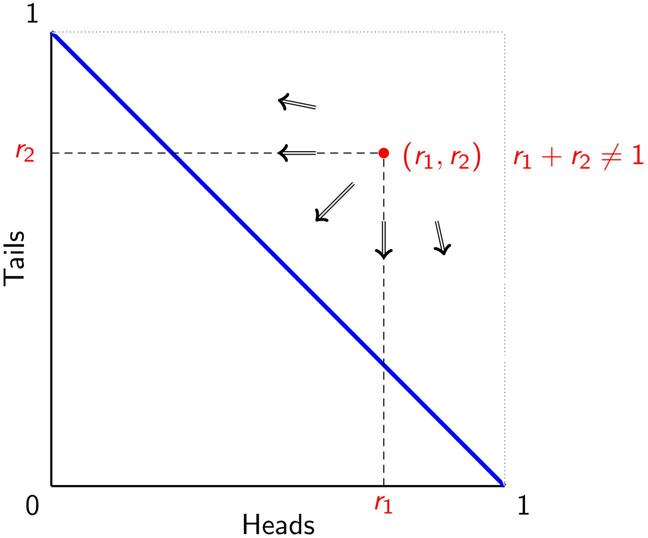
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



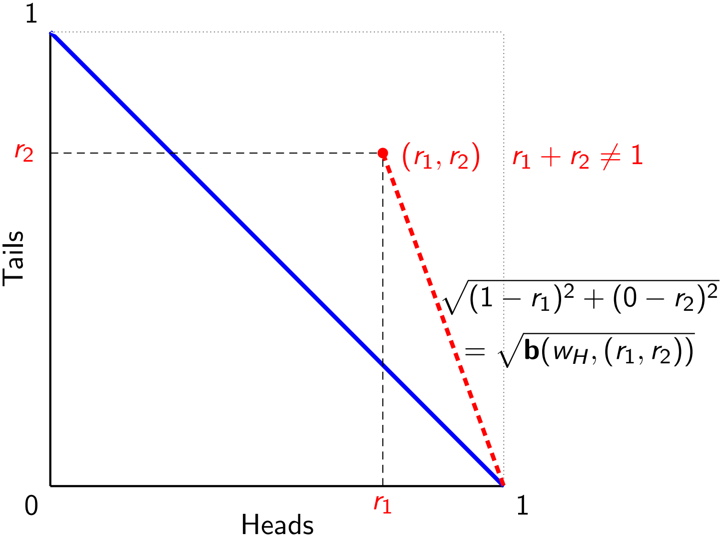
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



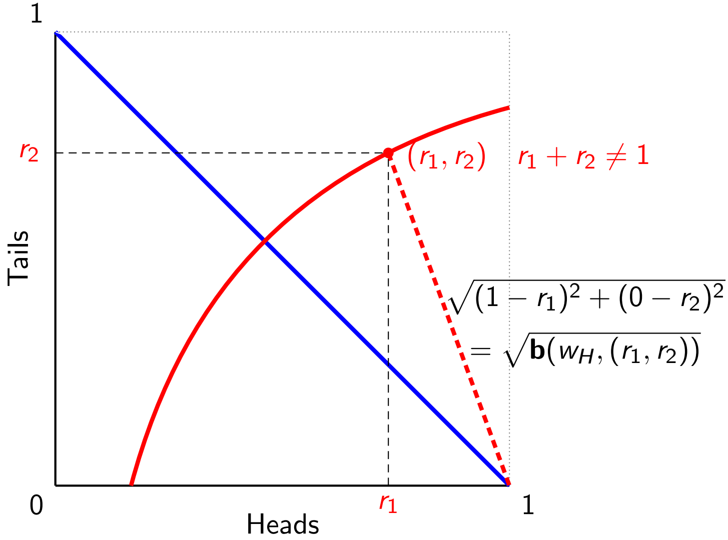
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



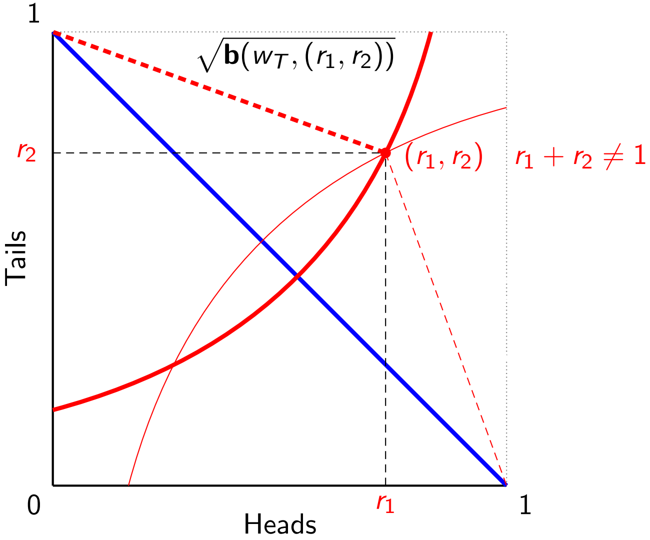
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



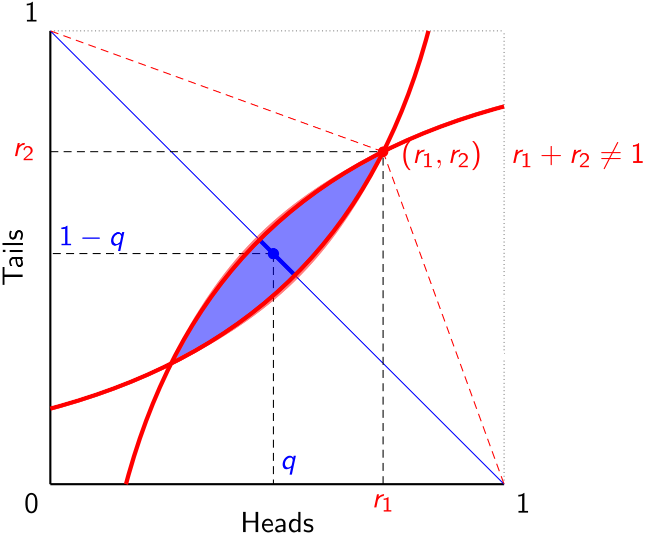
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



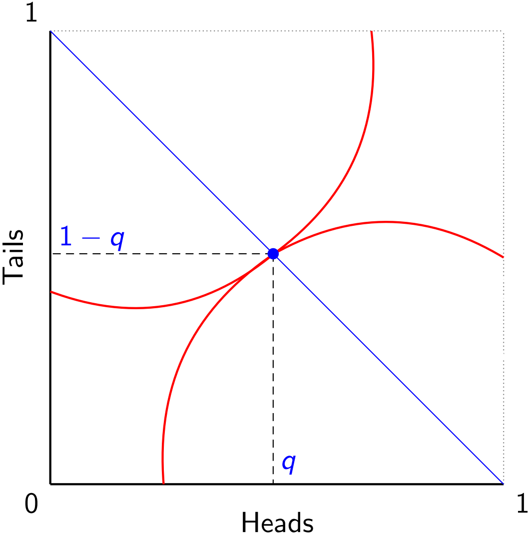
w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



w_H: The coin is facing heads up.

w_T: The coin is facing tails up.



w_H: The coin is facing heads up.

w_T: The coin is facing tails up.

J. Joyce. *A nonpragmatic vindication of probabilism*. Philosophy of Science 65, 575603 (1998).

Suppose that $\mathcal{F}$ is an algebra, $W_{\mathcal{F}}$ is the set of "ideal credences", i.e., for a state w, $v_w \in W_{\mathcal{F}}$ is defined as follows, for each $X \in \mathcal{F}$:

$$v_w(X) = \begin{cases} 1 & w \in X \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{C}_{\mathcal{F}}$ is the set of credences over $\mathcal{F}$ (non-negative real-valued functions on $\mathcal{F}$).

Suppose that $\mathcal{F}$ is an algebra, $W_{\mathcal{F}}$ is the set of "ideal credences", i.e., for a state w, $v_w \in W_{\mathcal{F}}$ is defined as follows, for each $X \in \mathcal{F}$:

$$v_w(X) = \begin{cases} 1 & w \in X \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{C}_{\mathcal{F}}$ is the set of credences over $\mathcal{F}$ (non-negative real-valued functions on $\mathcal{F}$).

Inaccuracy measure: $I : W_{\mathcal{F}} \times \mathcal{C}_{\mathcal{F}} \to [0, \infty]$ is measure for credence functions on $\mathcal{F}$.

Credal Veritism the only source of value for credences that is relevant to their epistemic status is their *gradational accuracy*, where the gradational accuracy of a credence in a true proposition is higher when the credence is closer to 1, which we might think of as the ideal or vindicated credence in a true proposition, while the gradational accuracy of a false proposition is higher when the credence is closer to 0, which we might think of as the ideal or vindicated credence in a false proposition. Thus, the only source of *disvalue* for credences is their gradational inaccuracy.

Example: Brier Score

Suppose that W is a set of states, $\mathcal{F}$ is a set of propositions, c is a credence function on $\mathcal{F}$ and for $w \in W$, $v_w : \mathcal{F} \to \{0, 1\}$ is a "valuation function". The brier score for c is:
$\mathbf{b} : W_{\mathcal{F}} \times \mathcal{C}_{\mathcal{F}} \to [0, \infty]$

$$\mathbf{b}(w, c) = \sum_{X \in \mathcal{F}} |v_w(X) - c(X)|^2$$

Joycean Inaccuracy

Structure: For each $w \in W$, $I(w, c)$ is a non-negative, continuous function that goes to infinity in the limit as $c(X)$ goes to infinity for any X.

Extensionality: At each possible world w, $I(w, c)$ is a function of nothing other than the truth-values that w assigns to propositions and the degrees of confidence that c assigns these propositions

Joycean Inaccuracy

Dominance: The accuracy of a system of degrees of belief is an increasing function of the believer's degree of confidence in any truth and a decreasing function of her degree of confidence in any falsehood.

If $c(Y) = c^*(Y)$ for every $Y \in \mathcal{F}$ other than X, then $I(w, c) > I(w, c^*)$ if and only if $|v_w(X) - c(x)| > |v_w(X) - c^*(X)|$

Normality If $|v_w(X) - c(X)| = |v_{w^*}(X) - c^*(X)|$ for all $X \in \mathcal{F}$, then $I(w, c) = I(w^*, c^*)$.

Strong Convexity: For any two distinct credence functions that are equally inaccurate at a given world, the third credence function obtained by "splitting the difference" between them and taking an equal mixture of the two is less inaccurate than either of them.

Joyce's Theorem. Suppose that $\mathcal{F}$ is an algebra and $I : W_{\mathcal{F}} \times \mathcal{C}_{\mathcal{F}} \to [0, \infty]$ is a Joycean inaccuracy measure for credence functions on $\mathcal{F}$. Now suppose that c^* is a credence function in $\mathcal{C}_{\mathcal{F}}$ that violates Probabilism. Then, there is a credence function $c' \in \mathcal{C}_{\mathcal{F}}$ satisfying Probabilism such that $I(w, c') < I(w, c)$ for all $w \in W_{\mathcal{F}}$.

Naive Dominance A rational agent will not adopt an option when there is another option that has lower disutility at all worlds.

Joyces accuracy argument for Probabilism

1. Credal Veritism + Joycean Inaccuracy
2. Naive Dominance
3. Joyce's Theorem

Therefore,

4 Probabilism

S. Moss. *Scoring Rules and Epistemic Compromise*. Mind, 120(480), pp. 1053 - 1069, 2011.

T. Seidenfeld. *Calibration, Coherence and Scoring Rules*. Philosophy of Science, 52(2), pp. 274 - 294, 1985.

R. Pettigrew. *Ch. 4: Measuring accuracy: a new account,* Accuracy and Credence. Oxford University Press, 2016.

H. Greaves and D. Wallace. *Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility*. .

	Good egg (s_1)	Bad egg (s_2)
Break into a bowl (A_1)	six egg omelet (o_1)	no omelet and five good eggs destroyed (o_2)
Break into a cup (A_2)	six egg omelet and a cup to wash (o_3)	five egg omelet and a cup to wash (o_4)
Throw away (A_3)	five egg omelet and one good egg destroyed (o_5)	five egg omelet (o_6)

$$EU^p(A) = \sum_i p(s_i)U(A(s_i))$$

	Good egg (s_1)	Bad egg (s_2)
Break into a bowl (A_1)	six egg omelet (o_1)	no omelet and five good eggs destroyed (o_2)
Break into a cup (A_2)	six egg omelet and a cup to wash (o_3)	five egg omelet and a cup to wash (o_4)
Throw away (A_3)	five egg omelet and one good egg destroyed (o_5)	five egg omelet (o_6)

$$EU^p(A_2) = p(s_1)U(o_3) + p(s_2)U(o_4)$$

	s_1	s_2	$\cdots$	s_{n-1}	s_n	
A_1			$\cdots$			$EU^p(A_1)$
A_2			$\cdots$			$EU^p(A_2)$
$\vdots$			$\vdots$			$\vdots$
A_{m-1}			$\cdots$			$EU^p(A_{m-1})$
A_m			$\cdots$			$EU^p(A_m)$

$$EU^p(A) = \sum_i p(s_i)U(s_i, A)$$

	s_1	s_2	$\cdots$	s_{n-1}	s_n
A_1	p_{s_1}	p_{s_2}	$\cdots$	$p_{s_{m-1}}$	p_{s_n}
A_2	q_{s_1}	q_{s_2}	$\cdots$	$q_{s_{m-1}}$	q_{s_n}
$\vdots$			$\vdots$		
A_{m-1}	r_{s_1}	r_{s_2}	$\cdots$	$r_{s_{m-1}}$	r_{s_n}
A_m	o_{s_1}	o_{s_2}	$\cdots$	$o_{s_{m-1}}$	o_{s_n}

For each i, x_{s_i} is a probability on S.

I.e., $p_{s_2} : S \to [0,1]$ with $\sum_i p_{s_2}(s_i) = 1$

	s_1	s_2	$\cdots$	s_{n-1}	s_n
A_1	p_{s_1}	p_{s_2}	$\cdots$	$p_{s_{m-1}}$	p_{s_n}
A_2	q_{s_1}	q_{s_2}	$\cdots$	$q_{s_{m-1}}$	q_{s_n}
$\vdots$			$\vdots$		
A_{m-1}	r_{s_1}	r_{s_2}	$\cdots$	$r_{s_{m-1}}$	r_{s_n}
A_m	o_{s_1}	o_{s_2}	$\cdots$	$o_{s_{m-1}}$	o_{s_n}

For each i, x_{s_i} is a probability on S.

$$EU^p(A) = \sum_i p(s_i)U(s_i, p_{s_i})$$

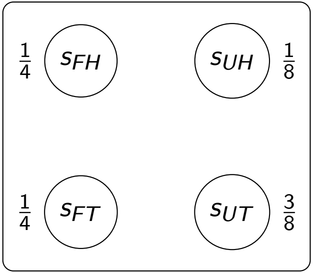
Mike has a coin. He is unsure as to whether or not it is a fair coin specifically, he assigns 50% credence to its being fair but he is (let us suppose) certain that either it is fair or it is weighted in such a way that the chances for outcomes (Heads,Tails) on a given toss are $(\frac{1}{4}, \frac{3}{4})$ respectively. The coin is about to be tossed; after observing the result of the toss, Mike will reassess his degrees of belief as to whether or not the coin is fair. He must decide in advance how the reassessment will proceed: which credence distribution he will move to if he sees heads, and which if he sees tails. We want to know how that decision should proceed.

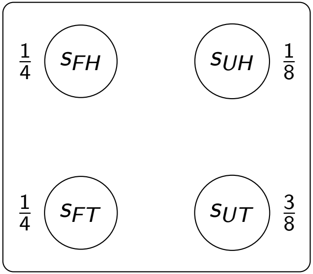
The agent contemplates a set S of (mutually exclusive and jointly exhaustive) possible states of the world; he is unsure as to which element of S obtains.

The agent contemplates a set S of (mutually exclusive and jointly exhaustive) possible states of the world; he is unsure as to which element of S obtains.

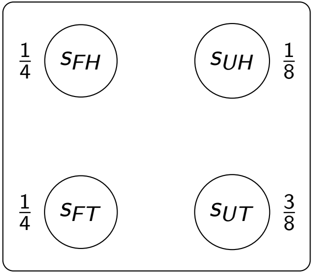
$S = \{s_{FH}, s_{FT}, s_{UH}, s_{UT}\}$

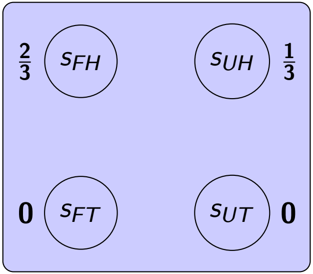
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	$p_{s_{FH}}$	$p_{s_{UH}}$	$p_{s_{FT}}$	$p_{s_{UT}}$
A_2	$q_{s_{FH}}$	$q_{s_{UH}}$	$q_{s_{FT}}$	$q_{s_{UT}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



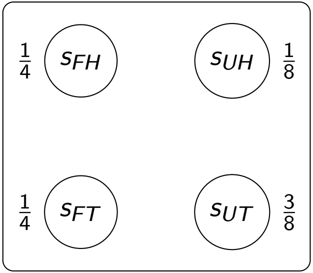


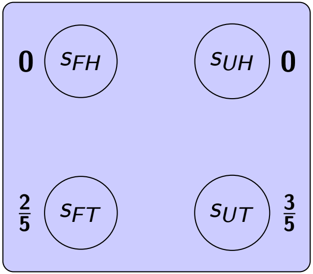
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



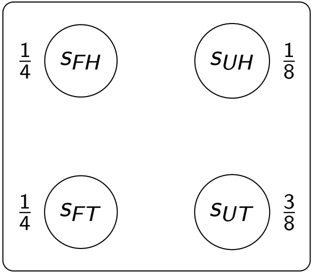


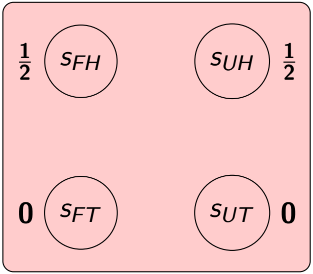
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



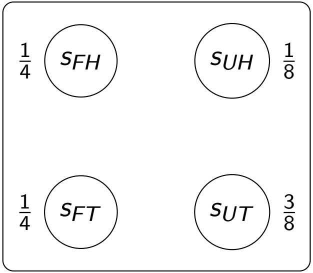


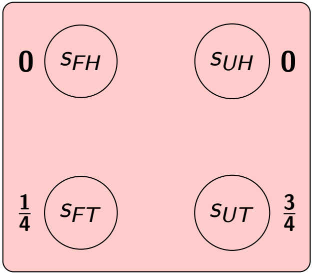
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$





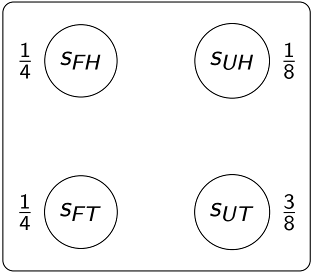
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

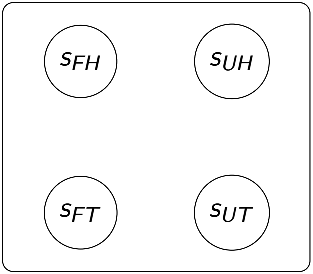




	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

$$p(s_{FH})U(s_{FH}, p_H) + p(s_{UH})U(s_{UH}, p_H)$$
$$+p(s_{FT})U(s_{FT}, p_T) + p(s_{UT})U(s_{UT}, p_T)$$

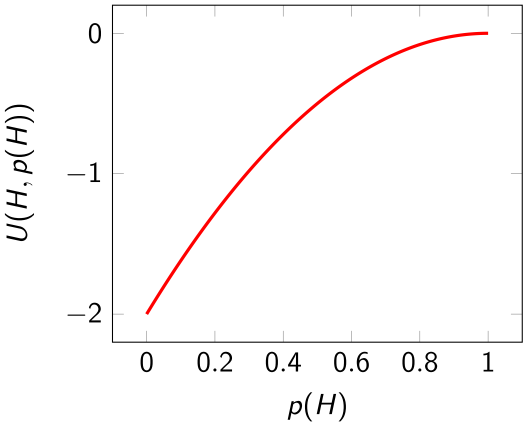


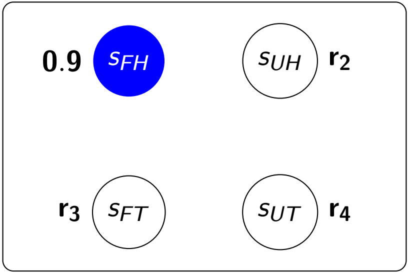


$$U(s, p_s) = -(1 - p(s))^2 - \sum_{s' \neq s} (0 - p(s'))^2$$

$$
\begin{aligned}
U(H, p(H)) &= -(1 - p(H))^2 - (0 - p(T))^2 \\
&= -(1 - p(H))^2 - (0 - (1 - p(H)))^2
\end{aligned}
$$

$$
\begin{aligned}
U(H, p(H)) &= -(1 - p(H))^2 - (0 - p(T))^2 \\
&= -(1 - p(H))^2 - (0 - (1 - p(H)))^2 \\
&= -(1 - p(H))^2 - (p(H) - 1)^2 \\
&= -(1 - 2p(H) + (p(H))^2 - ((p(H))^2 - 2p(H) + 1) \\
&= \mathbf{-2(p(H))^2 + 4p(H) - 2}
\end{aligned}
$$


0
−1
−2
0
0.2
0.4
0.6
0.8
1
$U(H, p(H))$
$p(H)$

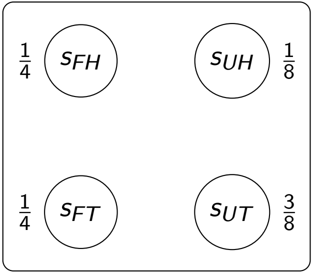


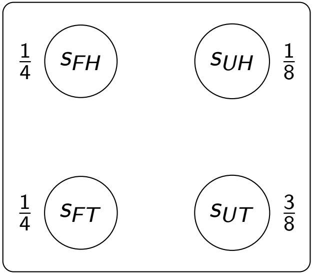
$$U(s_{FH},(0.9,0.1,0,0)) = -0.1^2 - 0.1^2 = -0.02$$

$$U(s_{FH},(0.9,0.05,0.05,0)) = -0.1^2 - 0.05^2 - 0.05^2 = -0.02 = -0.015$$

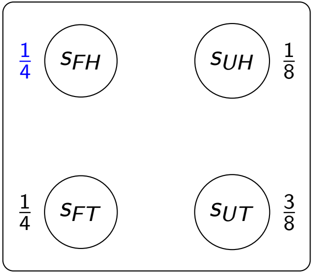
$$U(s_{FH},(0.9,0.033,0.033,0.034)) = -0.1^2 - 0.05^2 - 0.05^2 = -0.013334$$

	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	$p_{s_{FH}}$	$p_{s_{UH}}$	$p_{s_{FT}}$	$p_{s_{UT}}$
A_2	$q_{s_{FH}}$	$q_{s_{UH}}$	$q_{s_{FT}}$	$q_{s_{UT}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

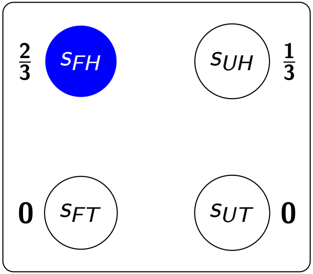




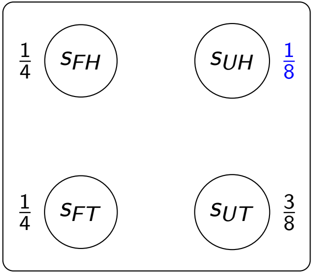
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



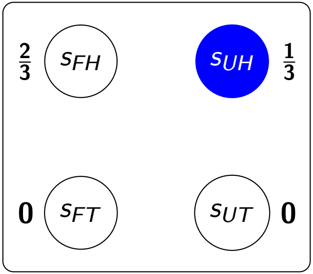
$$0.25 * (-0.22) + p(s_{UH})U(s_{UH}, p_H)$$
$$+p(s_{FT})U(s_{FT}, p_T) + p(s_{UT})U(s_{UT}, p_T)$$



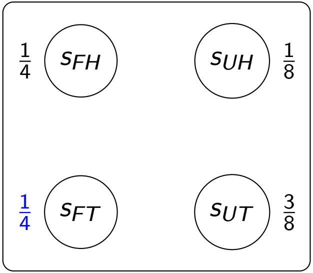
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



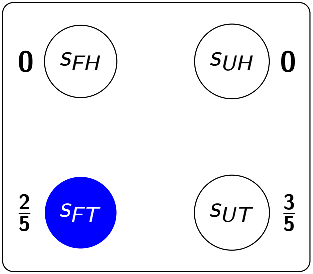
$$0.25 * (-0.22) + 0.125 * (-0.889)$$
$$+p(s_{FT})U(s_{FT}, p_T) + p(s_{UT})U(s_{UT}, p_T)$$



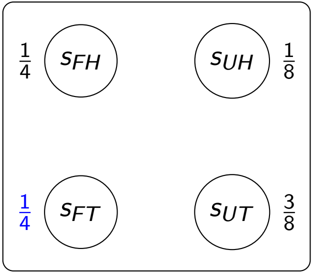
	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$



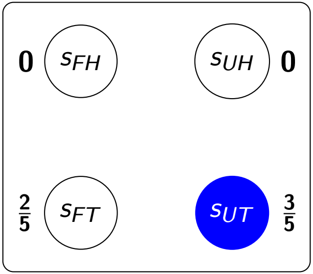
$$0.25 * (-0.22) + 0.125 * (-0.889)$$
$$+0.25 * (-0.72) + p(s_{UT})U(s_{UT}, p_T)$$



	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

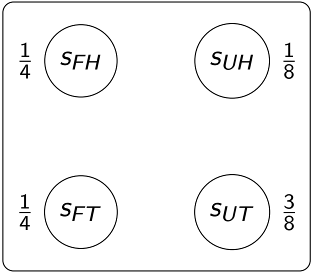


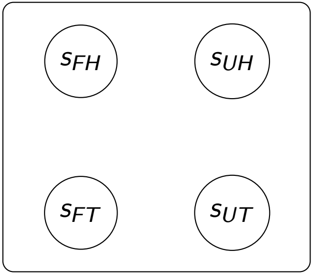
$$0.25 * (-0.22) + 0.125 * (-0.889)$$
$$+0.25 * (-0.72) + 0.375 * (-0.32)$$



	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

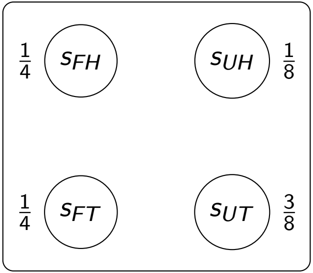
$$0.25 * (-0.22) + 0.125 * (-0.889)$$
$$+0.25 * (-0.72) + 0.375 * (-0.32)$$
$$= -0.4666667$$

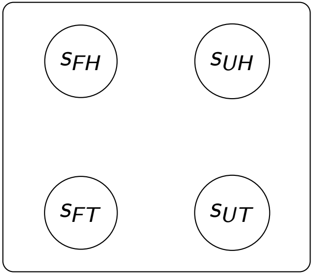




	s_{FH}	s_{UH}	s_{FT}	s_{UT}
A_1	p_H	p_H	p_T	p_T
A_2	q_H	q_H	q_T	q_T
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_{m-1}	$r_{s_{FH}}$	$r_{s_{UH}}$	$r_{s_{FT}}$	$r_{s_{UT}}$
A_m	$x_{s_{FH}}$	$x_{s_{UH}}$	$x_{s_{FT}}$	$x_{s_{UT}}$

$$0.25 * (-0.5) + 0.125 * (-0.5)$$
$$+0.25 * (-1.125) + 0.375 * (-0.125)$$
$$= -0.515625$$





constant acts

	s_1	s_2	$\cdots$	s_{n-1}	s_n
A_1	p_{s_1}	p_{s_2}	$\cdots$	$p_{s_{m-1}}$	p_{s_n}
A_2	q_{s_1}	q_{s_2}	$\cdots$	$q_{s_{m-1}}$	q_{s_n}
$\vdots$			$\vdots$		
k_q	q	q	$\cdots$	q	q
k_r	r	r	$\cdots$	r	r
$\vdots$			$\vdots$		
A_{m-1}	r_{s_1}	r_{s_2}	$\cdots$	$r_{s_{m-1}}$	r_{s_n}
A_m	p_{s_1}	p_{s_2}	$\cdots$	$p_{s_{m-1}}$	p_{s_n}

$$EU^p(k_q) = \sum_i p(s_i)U(s_i, q)$$

p **recommends** q, denoted $p \xrightarrow{R} q$, iff when the only available acts are the constant acts, k_q maximizes expected utility calculated with respect to p, i.e., for all $r \in \mathcal{P}$, $EU^p(k_q) \geq EU^p(k_r)$.

if p recommends no distribution distinct from q, then p **strongly recommends** q.

if p recommends p, then p is **self-recommending**. Iff, in addition, p recommends no distribution distinct from p, say that p is **strongly self-recommending**. Iff p is not self-recommending, say that p is **self-undermining**.

U is **everywhere stable** iff, according to U, every probability distribution is self-recommending.

U is **everywhere strongly stable** iff, according to U, every probability distribution is strongly self-recommending.

U is **partly stable** iff, according to U, some probability distributions are self-recommending and others are self-undermining.

U is **nowhere stable** iff, according to U, every probability distribution is self-undermining.

Conditionalization from prior p^*: given experiment $\mathcal{E}$, **Cond** is defined by

$$\textbf{Cond}: \qquad \text{For all } E_j \in \mathcal{E},\ \textbf{Cond}(E_j) = p^*(\cdot \mid E_j)$$

Quasi-Conditionalization from prior p^* given experiment $\mathcal{E}$: given experiment $\mathcal{E}$, **QC** is defined by

$$\mathcal{QC}: \qquad \text{For all } E_j \in \mathcal{E},\ \textbf{QC}(E_j) = \{q \in \mathcal{P} \mid \textbf{Cond}(E_j) \xrightarrow{R} q\}$$

Theorem. Of all acts that are available given and experiment $\mathcal{E}$, each quasi-conditionalizing updating rule is optimal. I.e.,

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

For any act **R**,

$$EU^{p^*}(\mathbf{R}) \quad = \quad \sum_{s\in S} p^*(s)U(s,\mathbf{R}(s))$$

For any act **R**,

$$\begin{aligned} EU^{p^*}(\mathbf{R}) &= \sum_{s\in S} p^*(s)U(s,\mathbf{R}(s)) \\ &= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} p^*(s)U(s,\mathbf{R}(E_j)) \end{aligned}$$

For any act **R**,

$$\begin{aligned} EU^{p^*}(\mathbf{R}) &= \textstyle\sum_{s\in S} p^*(s)U(s,\mathbf{R}(s)) \\ &= \textstyle\sum_{E_j\in\mathcal{E}}\sum_{s\in E_j} p^*(s)U(s,\mathbf{R}(E_j)) \\ &= \textstyle\sum_{E_j\in\mathcal{E}}\sum_{s\in E_j} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j)) \end{aligned}$$

For any act **R**,

$$\begin{aligned}
EU^{p^*}(\mathbf{R}) &= \sum_{s\in S} p^*(s)U(s,\mathbf{R}(s))\\
&= \sum_{E_j\in\mathcal{E}}\sum_{s\in E_j} p^*(s)U(s,\mathbf{R}(E_j))\\
&= \sum_{E_j\in\mathcal{E}}\sum_{s\in E_j} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j))\\
&= \sum_{E_j\in\mathcal{E}}\sum_{s\in E_j} \frac{p^*(E_j)}{p^*(E_j)}p^*(s\wedge E_j)U(s,\mathbf{R}(E_j))
\end{aligned}$$

For any act **R**,

$$
\begin{aligned}
EU^{p^*}(\mathbf{R}) &= \sum_{s\in S} p^*(s)U(s,\mathbf{R}(s)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} p^*(s)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} \frac{p^*(E_j)}{p^*(E_j)} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} p(E_j)(\sum_{s\in E_j} p^*(s\mid E_j)U(s,\mathbf{R}(E_j)))
\end{aligned}
$$

For any act **R**,

$$\begin{aligned}
EU^{p^*}(\mathbf{R}) &= \sum_{s\in S} p^*(s)U(s,\mathbf{R}(s)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} p^*(s)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} \sum_{s\in E_j} \frac{p^*(E_j)}{p^*(E_j)} p^*(s\wedge E_j)U(s,\mathbf{R}(E_j)) \\
&= \sum_{E_j\in\mathcal{E}} p(E_j)(\sum_{s\in E_j} p^*(s\mid E_j)U(s,\mathbf{R}(E_j))) \\
&= \sum_{E_j\in\mathcal{E}} p(E_j)EU^{p^*(\cdot\mid E_j)}(k_{\mathbf{R}(E_j)})
\end{aligned}$$

For $\mathbf{Q} \in \mathcal{QS}$, $EU^{p^*}(\mathbf{Q}) = \sum_{E_j \in \mathcal{E}} p^*(E_j) EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{Q}(E_j)})$

For all $\mathbf{Q} \in \mathcal{QS}$, for all $E_j \in \mathcal{E}$, $\mathbf{Cond}(E_j) \xrightarrow{R} \mathbf{Q}(E_j)$

For $\mathbf{Q} \in \mathcal{QS}$, $EU^{p^*}(\mathbf{Q}) = \sum_{E_j \in \mathcal{E}} p^*(E_j) EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{Q}(E_j)})$

For all $\mathbf{Q} \in \mathcal{QS}$, for all $E_j \in \mathcal{E}$, $\mathbf{Cond}(E_j) \xrightarrow{R} \mathbf{Q}(E_j)$

For all $\mathbf{Q} \in \mathcal{QC}$, for all $E_j \in \mathcal{E}$, for all $\mathbf{R} \in \mathcal{A}$,

$$EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{Q}(E_j)}) \geq EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{R}(E_j)})$$

$$\sum_j p^*(E_j) EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{Q}(E_j)}) \geq \sum_j p^*(E_j) EU^{p^*(\cdot \mid E_j)}(k_{\mathbf{R}(E_j)})$$

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

- Have we shown that every quasi-conditional act is optimal, without any assumptions on the utility functions??

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

- Have we shown that every quasi-conditional act is optimal, without any assumptions on the utility functions?? No!

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

- Have we shown that every quasi-conditional act is optimal, without any assumptions on the utility functions?? No!
- In the absences of information about the utility function, we have no idea what the recommended probabilitites $\{\mathbf{Q}(E_j)\}_{E_j \in \mathcal{E}}$ are.

$$\forall \mathbf{Q} \in \mathcal{QC}, \forall \mathbf{R} \in \mathcal{A}_{\mathcal{E}}, EU^{p^*}(\mathbf{Q}) \geq EU^{p^*}(\mathbf{R})$$

- Have we shown that every quasi-conditional act is optimal, without any assumptions on the utility functions?? No!
- In the absences of information about the utility function, we have no idea what the recommended probabilitites $\{\mathbf{Q}(E_j)\}_{E_j \in \mathcal{E}}$ are.
- We know that all quasi-conditional acts are optimal, but we do not know which acts those are.

Corollary 1. Conditionalization is optimal for a given experiment $\mathcal{E}$ if the conditional probabilities $\{p^*(\cdot \mid E_j) \mid E_j \in \mathcal{E}\}$ are self-recommending. If, in addition, at least one of the conditional probabilities is strongly self-recommending, then conditionalization is strongly optimal.

Corollary 2. If the agents epistemic utility function U is everywhere stable, then conditionalization is optimal. If U is everywhere strongly stable, then conditionalization is strongly optimal.

$$U(s,p) = -\sum_{X\subseteq S} \lambda_X(\chi_X(s) - p(X))^2$$

H. Greaves. *Epistemic decision theory*. Mind (2013).

(1) What an agent believes does not causally influence the truth of the propositions that her beliefs are about.

(2) While one generally hopes that the agent is more likely to believe that P if P is true than if P is false, still the fact that S believes that P on the basis of evidence E is not itself additional evidence in favor of, or against, P.

Promotion

Alice is up for promotion. Her boss, however, is a deeply insecure type: he is more likely to promote Alice if she comes across as lacking in confidence. Furthermore, Alice is useless at play-acting, so she will come across that way iff she really does have a low degree of belief that shes going to get the promotion. Specifically, the chance1 of her getting the promotion will be $1 - x$, where x is whatever degree of belief she chooses to have in the proposition P that she will be promoted. What credence in P is it epistemically rational for Alice to have?

Leap

Bob stands on the brink of a chasm, summoning up the courage to try and leap across it. Confidence helps him in such situations: specifically, for any value of x between 0 and 1, if Bob attempted to leap across the chasm while having degree of belief x that he would succeed, his chance of success would then be x. What credence in success is it epistemically rational for Bob to have?

Arrogance

Dennis is wondering whether or not he is arrogant. He takes a low (resp. a high) degree of belief that one is arrogant to be evidence for the proposition that one in fact is (resp. is not) arrogant: specifically, his credence in the proposition A that he is arrogant, conditional on the proposition that he will end up with credence x in A, is $1 - x$, for all $x \in [0, 1]$. What credence is it epistemically rational for Dennis to have in the proposition A that he is arrogant?

Imps

Emily is taking a walk through the Garden of Epistemic Imps. A child plays on the grass in front of her. In a nearby summerhouse are n further children, each of whom may or may not come out to play in a minute. They are able to read Emily's mind, and their algorithm for deciding whether to play outdoors is as follows. If she forms degree of belief 0 that there is now a child before her, they will come out to play. If she forms degree of belief 1 that there is a child before her, they will roll a fair die, and come out to play iff the outcome is an even number.

Imps

More generally, the summerhouse children will play with chance $1 - \frac{1}{2}q(C_0)$, where $q(C_0)$ is the degree of belief Emily adopts in the proposition C_0 that there is now a child before her. Emily's epistemic decision is the choice of credences in the proposition C_0 that there is now a child before her, and, for each $j = 1, \ldots, n$, the proposition C_j that the jth summerhouse child will be outdoors in a few minutes' time.

Imps

1. Emily has conclusive evidence that there is now a child before her, so presumably she should retain her degree of belief 1 in the proposition C_0 that indeed there is.

2. There will be a chance of 0.5 of each summerhouse child coming out to play, so she should have credence 0.5 in each C_i.

3. If Emily can just persuade herself to ignore her evidence for C_0, and adopt (at the other extreme) credence 0 in C_0, then, by adopting degree of belief 1 in each C_j ($j = 1, \dots, 10$) she can guarantee a perfect match to the remaining truths.

   Is it epistemically rational to accept this 'epistemic bribe'?

Savage Decision Theory

$$EU_{Cr}(a) = \sum_{s \in S} Cr(s)U(s \,\&\, a)$$

If the state partition is simply that over which the agent is selecting credences, since the epistemic utility function is a proper scoring rule and no new information is to be acquired, the theory will simply recommend retaining ones initial credences, whatever they happen to be. This epistemic decision theory will therefore not capture the sense in which, for example, any credence in P other than 0.5 in the Promotion case is epistemically deficient.

If the state partition is simply that over which the agent is selecting credences, since the epistemic utility function is a proper scoring rule and no new information is to be acquired, the theory will simply recommend retaining ones initial credences, whatever they happen to be. This epistemic decision theory will therefore not capture the sense in which, for example, any credence in P other than 0.5 in the Promotion case is epistemically deficient.

The predictions of 'Savage' EDT depend on the state partition. For example, in the Promotion case, if the states are propositions that specify how the chance of promotion depends on the agents choice of epistemic act, the theory does indeed recover the intuitively correct result that only credence 0.5 in P is rationally permitted

Deterrence Problem: You park your car in a dodgy neighbourhood. A hooligan approaches you, and tells you that he will probably smash your windscreen while youre gone unless you pay him \$10 now; if you do pay, he will probably leave your car alone. The acts are $\{\text{pay}, \text{donot pay}\}$. What should you do?

Evidential Decision Theory

$$EU_{Cr}(a) = \sum_{s \in S} Cr(s \mid a) U(s \,\&\, a)$$

Newcomb's Paradox

Two boxes in front of you, A and B.

Box A contains \$1,000 and box B contains either \$1,000,000 or nothing.

Newcomb's Paradox

Two boxes in front of you, A and B.

Box A contains \$1,000 and box B contains either \$1,000,000 or nothing.

Your choice: either open both boxes, or else just open B. (You can keep whatever is inside any box you open, but you may not keep what is inside a box you do not open).

Newcomb's Paradox

A very powerful being, who has been invariably accurate in his predictions about your behavior in the past, has already acted in the following way:

1. If he has predicted that you will open just box B, he has in addition put $1,000,000 in box B
2. If he has predicted you will open both boxes, he has put nothing in box B.

What should you do?

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

Newcomb's Paradox

	B = 1M	B = 0
1 Box	1M	0
2 Boxes	1M + 1000	1000

Newcomb's Paradox

	B $= 1$M	B $= 0$
1 Box	1M	0
2 Boxes	1M $+$ 1000	1000

	B $= 1$M	B $= 0$
1 Box	h	$1 - h$
2 Boxes	$1 - h$	h

Newcomb's Paradox

In this case, Jeffrey's theory predicts that you should take only the opaque box. But this, as is generally (if not universally) accepted, is the wrong answer.

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)
- Causal Decision Theory: expected utility involves probabilities of causal counterfactuals (2-box)

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)
- Causal Decision Theory: expected utility involves probabilities of causal counterfactuals (2-box)
- No Acyclic Reasons: reasoning cannot refer to the act of choice in an essential way (2-box)...plus some “mental gymnastics” (1-box)

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)
- Causal Decision Theory: expected utility involves probabilities of causal counterfactuals (2-box)
- No Acyclic Reasons: reasoning cannot refer to the act of choice in an essential way (2-box)...plus some "mental gymnastics" (1-box)
- "Tickle": $Pr($ page box contain \$0 $\mid T$ & 1-box$) =$ $Pr($ page box contain \$0 $\mid T$ & 2-box$)$ (2-box)

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)
- Causal Decision Theory: expected utility involves probabilities of causal counterfactuals (2-box)
- No Acyclic Reasons: reasoning cannot refer to the act of choice in an essential way (2-box)...plus some "mental gymnastics" (1-box)
- "Tickle": $Pr($ page box contain \$0 $\mid T$ & 1-box$) =$ $Pr($ page box contain \$0 $\mid T$ & 2-box$)$ (2-box)
- Evidential Decision Theory: decisions to act provides evidence for the consequences (1-box)

- Orthodox Bayesian: It is a problem of act-state dependence (1-box)
- Causal Decision Theory: expected utility involves probabilities of causal counterfactuals (2-box)
- No Acyclic Reasons: reasoning cannot refer to the act of choice in an essential way (2-box)...plus some "mental gymnastics" (1-box)
- "Tickle": $Pr($ page box contain \$0 $\mid T$ & 1-box$) =$ $Pr($ page box contain \$0 $\mid T$ & 2-box$)$ (2-box)
- Evidential Decision Theory: decisions to act provides evidence for the consequences (1-box)
- Ratifiability: decision makers must assess the act in light of the decision to perform it and only choose acts that are self-ratifiable (1-box)

Causal Decision Theory

A. Egan. *Some Counterexamples to Causal Decision Theory*. Philosophical Review, 116(1), pgs. 93 - 114, 2007.

The Psychopath Button: Paul is debating whether to press the ‘kill all psychopaths’ button. It would, he thinks, be much better to live in a world with no psychopaths.

The Psychopath Button: Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button.

The Psychopath Button: Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button. Paul very strongly prefers living in a world with psychopaths to dying. Should Paul press the button?

(Set aside your theoretical commitments and put yourself in Paul's situation. Would you press the button? Would you take yourself to be irrational for not doing so?)

$p(\text{press button}\text{dead}) = 0.001$

$p(\text{press button}\text{live in a world without psychopaths}) = 0.999$

This is because Paul either is or is not a psychopath, and the probability of the two possibilities does not depend on what he decides to do.

Press Button: $p(\text{press buttondead}) \cdot u(\text{dead}) +$
$p(\text{press buttonlive in a world without psychopaths}) \cdot$
$u(\text{live in a world without psychopaths}) =$
$(0.001 \cdot -100) + (0.99 \cdot 1) = 0.89$

Do Not Press Button:
$p(\text{do not press buttonlive in a world with psychopaths}) \cdot$
$u(\text{live in a world with psychopaths}) = 1 \cdot 0 = 0$

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight.

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight. As the man knows, Death is a good predictor of his whereabouts. If he stays in Damascus, he thereby has evidence that Death will look for him in Damascus. However, if he goes to Aleppo he thereby has evidence that Death will look for him in Aleppo.

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight. As the man knows, Death is a good predictor of his whereabouts. If he stays in Damascus, he thereby has evidence that Death will look for him in Damascus. However, if he goes to Aleppo he thereby has evidence that Death will look for him in Aleppo. Wherever he decides to be at midnight, he has evidence that he would be better off at the other place. No decision is stable.

A. Gibbard and W. Harper. *Counterfactuals and Two Kinds of Expected Utility*. In Ifs: Conditionals, Belief, Decision, Chance, and Time, pp. 153190, 1978.

Ratifiability

The notion of ratifiability is applicable only where, during deliberation, the agent finds it conceivable that he will not manage to perform the act he finally decides to perform, but will find himself performing one of the other available acts instead...The option in question is ratifiable or not depending on whether or not the expected desirability of actually carrying out each of the alternatives (in spite of having chosen to carry out a different option, as hypothesized) (Jeffrey, 1983, pgs. 18-20)

Deliberational Decision Theory

A rational agent should not necessarily perform the act that has highest expected utility according to her initial credences; rather, she should allow her credences to develop according to a specified dynamical rule (which rule involves the expected utilities of the acts under consideration), and she should perform the mixed act with probabilities equal to her equilibrium credences.

This theory gives intuitively reasonable judgements in all problem cases considered to date.

A problem for Epistemic Decision Theory

Causal practical decision theory would issue the correct verdict on a practical analogue of our Imps case, but we have no epistemic decision theory that deals adequately with this case.

There is no theory that recovers the obviously correct result that an agent (epistemically-) should retain credence 1 in propositions for which she has conclusive evidence, even in the face of 'epistemic bribes'.

A problem for Epistemic Decision Theory

Causal practical decision theory would issue the correct verdict on a practical analogue of our Imps case, but we have no epistemic decision theory that deals adequately with this case.

There is no theory that recovers the obviously correct result that an agent (epistemically-) should retain credence 1 in propositions for which she has conclusive evidence, even in the face of 'epistemic bribes'.

Problem: a decision-theoretic utility function always assesses epistemic utility globally, and hence will always be open to the move of increasing overall expected epistemic utility by making a sacrifice of a relatively small number of propositions; our intuitive notion of epistemic rationality, meanwhile, does not seem to exhibit this willingness.

Perfectionism The accuracy of a credence function at a world is its proximity to the ideal credence function at that world.

Let S be a set. A **distance** is a function $D : S \times S \rightarrow \mathbb{R}_0^+$ such that

- D is non-negative: $D(c, c') \geq 0$ with equality iff $c = c'$
- D is symmetric: $D(c, c') = D(c', c)$
- D satisfies triangle inequality: $D(c, c'') \leq D(c, c') + D(c', c'')$

We will assume that D satisfies non-negativity. I.e., D is a **divergence**.

Perfectionism If I is a legitimate inaccuracy measure, there is a **divergence** D such that $I(w, c) = D(i_w, c)$. Recall: i_w is the ideal or vindicated credence function at w. We say that D generates I (relative to that notion of vindication).

Brier Accuracy

Alethic Vindication The ideal credence function at world w is the omniscient credence function at w, namely, v_w.

Squared Euclidean Distance Distance between credence functions is measured by squared Euclidean distance.

Additivity

If I is a legitimate (global) measure of inaccuracy, then there is a local measure of inaccuracy $\mathfrak{s}$ such that

$$I(w, c) = \sum_{X \in \mathcal{F}} \mathfrak{s}(i_w(X), c(X))$$

where i_w is the ideal credence at w.

When we say that we represent an agent by her credence function, it can sound as if we're representing her as having a single, unified doxastic state.

When we say that we represent an agent by her credence function, it can sound as if we're representing her as having a single, unified doxastic state.

Really, we are just representing her as having an agglomeration of individual doxastic states, namely, the individual credences she assigns to the various propositions about which she has an opinion. A credence function is simply a mathematical way of representing this agglomeration; it is a way of collecting together these individual credences into a single object.

I is a legitimate inaccuracy measure, then there is a divergence D such that

- $I(w, c) = D(i_w, c)$ (in such a case, we write $I = I_D$)
- There is a function $\mathsf{s} : [0,1] \times [0,1] \to [0, \infty]$ such that
  - for all $x, y \in [0,1]$, $\mathsf{s}(x, y) \geq 0$ with equality if $x = y$.
  - $D(c, c') = \sum_{X \in \mathcal{F}} \mathsf{s}(c(X), c'(X))$

Continuity If I is a legitimate inaccuracy measure and there is a divergence D generated by s such that

$$I(w, c) = I_D(w, c) = D(i_w, c) = \sum_{X \in \mathcal{F}} \mathsf{s}(i_w(X), c(X))$$

then $\mathsf{s}(x, y)$ is continuous in both its arguments.

To demand that s is continuous in its second argument is to say that there are no 'jumps' in inaccuracy as credences change - that is, small changes in credence will give rise to small changes in inaccuracy; there can be no small shift in credence that is accompanied by a large jump in inaccuracy.

Calibration

> Granting that [an agent] is going to think always in the same way about all yellow toadstools, we can ask what degree of confidence it would be best for him to have that they are unwholesome. And the answer is that it will in general be best for his degree of belief that a yellow toadstool is unwholesome to be equal to the proportion of yellow toadstools that are unwholesome. (This follows from the meaning of degree of belief.)
>
> (Ramsey, 1931, 195)

Decomposition If I is a legitimate inaccuracy measure generated by a divergence D, then there are α, β such that

$$D(v_w, c) = \alpha D(c^w, c) + \beta D(v_w, c^w)$$

c^w is the ideally calibrated credence in w.

Theorem (Pettigrew). Suppose Alethic Vindication, Perfectionism, Divergence Additivity, Divergence Continuity and Decomposition. Then, if I is a legitimate inaccuracy measure, there is an additive **Bregman divergence** D such that $I(w, c) = D(v_w, c)$.

Symmetry If I is a legitimate inaccuracy measure generated by a divergence D, then D is symmetric: that is, $D(c, c') = D(c', c)$ for all c, c'

Theorem (Pettigrew). Suppose Alethic Vindication, Perfectionism, Divergence Additivity, Divergence Continuity, Decomposition, and Symmetry. Then, if I is a legitimate inaccuracy measure, then I is the Brier score or some linear transformation of it.