THE REWARDS AND PERILS OF USING IMPRECISE PROBABILITIES TO REPRESENT UNCERTAINTY

Rutgers University
November 10, 2009

Jim Joyce
Department of Philosophy
The University of Michigan
jjoyce@umich.edu

In *Bayesian Rationality* (Oxford, 2007) M. Oaksford and N. Chater argue that rationality involves the ability to reason correctly about uncertainty, and that "cognition in general, and human everyday reasoning in particular, is best viewed as solving probabilistic, rather than logical, inference problems." (BBS 2009, p. 69) While I agree with the first point, and would like the second to be true, both ideas leave open the question of how probabilistic judgments and reasoning should be modeled. One issue, in particular, concerns whether to model beliefs, inferences, and information processing using "sharp" probabilities, thereby portraying agents as having a single subjective probability function, or whether to employ "imprecise" probabilistic models that use structured sets of probability functions to represent uncertainty? Many philosophers have championed imprecise probabilities as the right response to incomplete or ambiguous evidence, and some psychologists (e.g., N. Pfeifer and G. D. Kleiter) have suggested that imprecise models might serve better than sharp probabilities as "normative reference systems" that can be used to explain human cognition. I will sympathetically assess the normative prospects for one version of the theory of imprecise probabilities, arguing that it does a better job than "sharp" theories do at reflecting incomplete or ambiguous evidence. Moreover, since the imprecise view has the capacity to express far more types of attitudes and judgments than can be encoded in sharp probabilities; it allows us more freedom in representing mental states. This is both a benefit and a pitfall. Imprecise probabilities offer us more nuanced ways of representing beliefs, but also more nuanced ways of misrepresenting them. I will discuss cases of each type and draw some tentative morals.

BASIC IDEAS OF BAYESIAN EPISTEMOLOGY

- Believing is not an all-or-nothing matter. Opinions come in varying gradations of strength ranging from full certainty of truth to complete certainty of falsehood.

- Gradational belief is governed by the laws of probability, which codify the minimum standards of consistency (or "coherence") to which rational opinions must confirm.

- Learning involves *Bayesian conditioning*: a person who acquires data *D* should modify her opinions in such a way that her "posterior" views about the relative odds of events consistent with *D* agree with her "prior" views about these relative odds.

- Gradational beliefs are often revealed in decisions. Rational agents choose options that they *estimate* will produce desirable outcomes, and these (gradational) estimates are a function of their (gradational) beliefs.

	X & Y	X & ~Y	~X & Y	~X & ~Y
Option - O_X	**prize**	**prize**	**penalty**	**penalty**
Option - O_Y	**prize**	**penalty**	**prize**	**penalty**

You should (determinately) prefer O_X to O_Y if and only if you are more confident of X than of Y.

RATIONALITY AND "RATIONAL ANALYSIS"

Bayesianism is a *normative* theory. The relationship between normative and descriptive theories is complicated, but one option is the **rational analysis model** of Anderson (1990) or Oaksford & Chater (2007), which seeks to understand cognitive systems by portraying them as producing behavior that is largely rational given the agent's evidence and goals.

- *Rationality Principle* (Anderson, 1990). The cognitive system optimizes the adaptation of the behavior of the organism. (Compare Davidsonian "interpretivism" about mental states.)

    1. Precisely specify what the goals of the cognitive system are.
    2. Develop a formal model of the environment that the system is adapted to.
    3. Make the minimal assumptions about computational costs.
    4. Derive the optimal behavioral function given 1-3. (THEORY OF RATIONALITY USED HERE.)
    5. Examine the empirical literatures to see if the predictions of the behavioral function are confirmed.

- Rational analysis is *not* a theory of cognitive *mechanisms*.
- It is, rather, a theory of those general features of thought and behavior that obtain irrespective of the details of the underlying cognitive processes.

Key Question: What theory of rationality should we employ in "rational analysis"?

AN ANSWER (Oaksford & Chater, 2007 & 2009): **BAYESIAN!**

"Everyday thought involves astonishingly rich and subtle probabilistic reasoning – but probabilistic reasoning that is primarily qualitative rather than numerical" (2009, p. 69)

> Note: Commitment to a using a Bayesian account of rationality does not require one to think that mistakes in probabilistic reasoning are either rare or unsystematic.

Even if we grant that Bayesian accounts of rationality are well-suited to the purposes of modeling and explaining human thought and behavior, there remains the question of which version of Bayesianism provides the best account of rationality for purposes of "rational analysis".

- I. J. Good once wrote a paper claiming that there are (at least) $2^4 \cdot 3^6 \cdot 4 = 46656$ kinds of Bayesians!

Concern Today: Is a "precise" or "imprecise" version of Bayesianism better for modeling and explaining human thought and action?

👎 "PRECISE" BAYESIANISM 👎

- Graded beliefs come in sharp numerical degrees: in any context, we can think of an agent's epistemic state as given by a single *credence function* $\boldsymbol{c}$ that assigns a degree of belief $\boldsymbol{c}(X) \in [0, 1]$ to each proposition X (in some Boolean algebra).

  Note: "Objective" and "subjective" Bayesians differ about whether credence functions reflect objective constraints on beliefs or (also) matters of personal opinion.

- Rational credences are *additive*: $\boldsymbol{c}(X \vee Y) = \boldsymbol{c}(X) + \boldsymbol{c}(Y)$ when X and Y are contraries.

- Learning is governed by *Bayes' Theorem*: a person who acquires data D should modify her opinions so that she believes each proposition X to degree

  $$\boldsymbol{c}^*(X) = \boldsymbol{c}(X)\cdot[\boldsymbol{c}(D \mid X)/\boldsymbol{c}(D)]$$

  where $\boldsymbol{c}(D \mid X) = \boldsymbol{c}(D \,\&\, X) / \boldsymbol{c}(X)$ is the prior conditional probability of D given X.

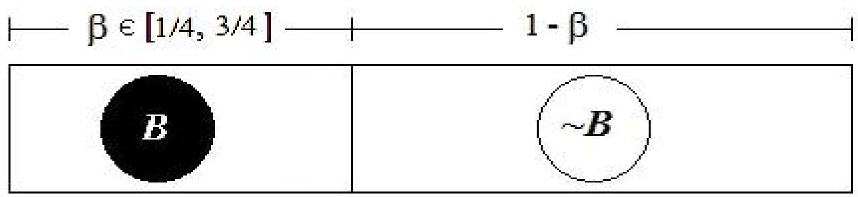
- Rational decision making is a matter of choosing options that maximize *expected utility* computed relative to one's credences.

PROBLEMS WITH PRECISE BAYESIANISM

- It is psychologically unrealistic to suppose that people have attitudes that are precise enough to be represented by real numbers. What could $\boldsymbol{c}(X) = 1/\pi^2$ mean?

- It misrepresents decision making, e.g., Ellsberg's paradox, ambiguity aversion.

- Since evidence is often incomplete, imprecise or equivocal, the *rational* response is often to have beliefs that are incomplete, imprecise or equivocal.



A black/white coin is chosen randomly from an urn containing coins of every possible bias $\frac{1}{4} < \beta < \frac{3}{4}$.
You have *no information* about the proportions with which coins of various biases appear in the urn.

How confident should you be that the coin comes up black when next tossed?

- "Objective" Precise Bayesian: $\boldsymbol{c}(B) = \frac{1}{2}$ because this choice uniquely minimizes the amount of *extra information* one needs to add to get a sharp degree of belief.

- "Subjective" Precise Bayesian: Any $\boldsymbol{c}(B)$ between ¼ and ¾ can be coherently held.

- "Imprecise" Bayesian: It is determinate that $\frac{1}{4} < \boldsymbol{c}(B) < \frac{3}{4}$, but $\boldsymbol{c}(B)$ lacks a determinate value because the evidence does not discriminate among values $\boldsymbol{c}(B) = p$ with $\frac{1}{4} < p < \frac{3}{4}$.

👍 IMPRECISE BAYESIANISM 👍

- IMPRECISION. Graded beliefs do not come in sharp degrees. A person's *credal state* is best represented by a set $\boldsymbol{C}$ of credence functions.

- COHERENCE. For a rational agent, $\boldsymbol{C}$ can be identified with a *set* of probability functions.

- CONDITIONING. If a person with credal state $\boldsymbol{C}$ learns that some proposition D is certainly true, then her post-learning credal state will be $\boldsymbol{C}_D = \{\boldsymbol{c}(\bullet|D) : \boldsymbol{c} \in \boldsymbol{C} \text{ with } \boldsymbol{c}(D) > 0\}$.

- COMPLETENESS. If my credal state is $\boldsymbol{C}$ and yours is $\boldsymbol{C}^*$, then we have the same total system of beliefs if and only if $\boldsymbol{C} = \boldsymbol{C}^*$.

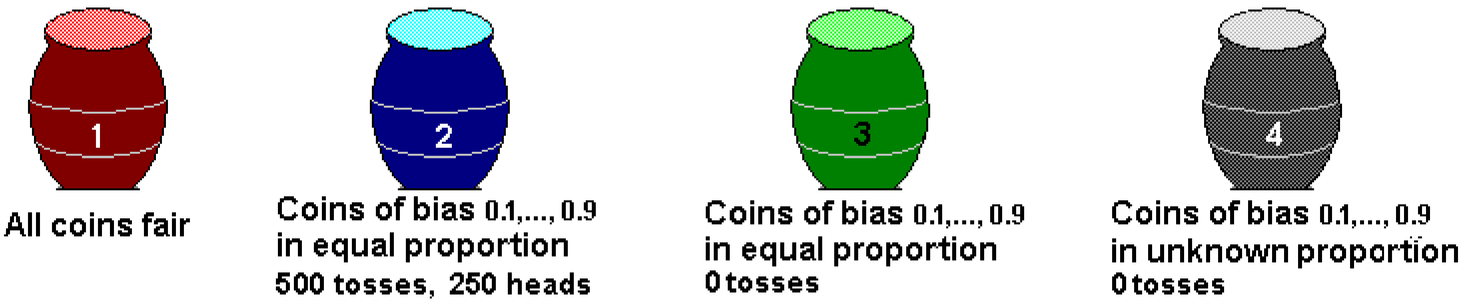
- SUPERVALUATION. If some claim about probabilities $\varphi(\boldsymbol{c})$ holds for every $\boldsymbol{c} \in \boldsymbol{C}$, then φ is a determinate truth about what the person believes.

Important qualifications: imprecise ≠ vague and imprecise ≠ unknown.

ILLUSTRATION: THE FOUR URNS

It's natural to think that a rational person's beliefs "reflect her evidence," but beliefs reflect evidence in a variety of ways, not all of which are captured by sharp probabilities.

Imagine that coin drawn from the following urns will be tossed.



Popper's Objection: Bayesianism treats all four cases as identical by assigning ***C***(*Heads*) = ½, but these are entirely different evidential situations.

Jeffrey's Reply: same credence ≠ same epistemic state.

- The Urn_1 and Urn_2 probabilities are *resilient*. They remain fixed (exactly for Urn_1, roughly for Urn_2) given future evidence: ***c***(*Heads* | 24H, 1T) = ½ and ***c***(*Heads* | 24H, 1T) ≈ 0.50023.
- The Urn_3 probabilities are *unstable* in the face of evidence: ***c***(*Heads* | 24H, 1T) = 0.880.

What about Urn_4? What value should we assign ***c***(*Heads* | 24H, 1T)?

TWO COMMON ANSWERS

Objective Bayesian: Urn_4 and Urn_3 are equivalent. ***c***(*Heads* | *Data*) exists and equals ***c***(*Heads* | *Data*) for all data.

- Key Claim: In both cases, no evidence distinguishes any one bias from any other. Principles of sound epistemology (Insufficient Reason, MaxEnt) require that we treat symmetrical cases symmetrically *by assigning them the same probability*.

  The choice of ***c***(*Heads* | *data*) = ***c***(*Heads* | *Data*) is often justified by appeal to the requirement that the prior probability ***c*** should encode the minimum amount of information consistent with the evidence, so that $c(\beta = 0.i) = p_i$ maximizes $\text{Entropy}(p_i) = -\sum_i p_i \cdot ln(p_i)$.

  Problem(s): ‘‘Apart from evolving a vitally important piece of knowledge, that of the exact form of the distribution, out of an assumption of complete ignorance, it is not even a unique solution.’’ R.A. Fisher, 1922, pp. 324-325.

Subjective Bayesian: Urn_4 and Urn_3 are not equivalent. ***c***(*Heads* | *Data*) exists but it can consistently have any value in [0, 1], whatever the value of ***c***(*Heads* | *Data*).

- Key Claim: A rational agent can have any probabilistically coherent set of credences over the possible biases. So, any credence for heads in light of data can be rationally entertained.
- Problem: The choice of one sharp probability over any other seems arbitrary. In particular, in the face of symmetrical evidence there is no more reason to choose a prior with $c(\beta = 0.i) = p_i$ than the symmetrical prior with $c(\beta = 1 - 0.i) = p_i$.

A THIRD WAY

"The problem is not that Bayesians have yet to discover the truly noninformative priors, but rather that no precise probability distribution can adequately represent ignorance."
(P. Walley, *Statistical Reasoning with Imprecise Probabilities*, 1991)

Imprecise Bayesian: Urn_4 and Urn_3 are not equivalent. ***c***(*Heads* | *Data*) does not exist! Instead modeling our beliefs about possible biases using a single credence function, we should use a ***set*** of credence functions that best reflects our state of uncertainty.

- *Objective Imprecise*: There is a single imprecise credal state that is appropriate for any given body of evidence. For symmetric evidence this state is symmetric.

- *Subjective Imprecise*: There are often many sets of credence functions consistent with the data, and a believer is free to adopt any of these sets as her credal state.

Judgments in light of evidence are seen as imposing (typically qualitative) "constraints" on credal states**.**

X is more likely that *Y*	$\boldsymbol{c}(X) > \boldsymbol{c}(Y)$ for all $\boldsymbol{c} \in \boldsymbol{C}$
X is at least twice as likely as not	$\boldsymbol{c}(X) > {}^{2}/_{3}$ for all $\boldsymbol{c} \in \boldsymbol{C}$
X is pretty likely	$\boldsymbol{c}(X) > t$ for all $\boldsymbol{c} \in \boldsymbol{C}$ (threshold *t* contextually determined)
X and *Y* are independent	$\boldsymbol{c}(X \,\&\, Y) = \boldsymbol{c}(X)\cdot\boldsymbol{c}(Y)$ for all $\boldsymbol{c} \in \boldsymbol{C}$
X and *Y* are (+, –) correlated	$\boldsymbol{c}(X \,\&\, Y) >,< \boldsymbol{c}(X)\cdot\boldsymbol{c}(Y)$ for all $\boldsymbol{c} \in \boldsymbol{C}$
X has a binomial distribution	$\boldsymbol{c}(X^n \,\&\, {\sim}X^m \mid \theta) = \theta^n\cdot(1-\theta)^m$ for all $\boldsymbol{c} \in \boldsymbol{C}$

SOME USEFUL DOXASTIC ATTITUDES THAT CAN BE REPRESENTED USING IMPRECISE PROBABILITIES

- *Independence/Relevance*: One can see X and Y as independent even when one lacks probabilities for X, Y, X & Y or X given Y.

  Example: X = "St. Kilda will win the Australian Football League Grand Final in 2015." Y = "India and Pakistan will go to war sometime between 2025 and 2050."

  Note: This attitude has no "definition" in terms of betting behavior, i.e., there is no pattern of preferences that is a perfect indicator of a bare independence judgment.

  That's OK: Bayesianism should jettison its operationist past! Sometimes we must rely on indirect evidence and partial tests in belief attribution.

  - If $y \approx$ [1 if Y, 0 if $\sim Y$], then $\boldsymbol{c}(X \,\&\, Y) = \boldsymbol{c}(X)\cdot\boldsymbol{c}(Y)$ is determinate when [1 if X & Y, 0 if X & $\sim Y$, y if $\sim X$] $\approx$ [1 if $\sim X$ & Y, 0 if $\sim X$ & $\sim Y$, y if X].
  - If [1 if X & Y, 0 if X & $\sim Y$, y if $\sim X$] > [1 if $\sim X$ & Y, 0 if $\sim X$ & $\sim Y$, y if X] for all $y \in [0, 1]$, then X and Y are not independent (and Y is positively relevant to X).

- *Unknown Interaction*: One can be clueless about the dependence between X and Y, in which case C contains all functions with $\boldsymbol{c}(Y \mid X) = y$ and $\boldsymbol{c}(Y \mid \sim X) = y^*$ for $y, y^* \in [0, 1]$.

  Example: X = "A cricket team from Hyderabad will win the *The Irani Trophy* in 2015." Y = "A Hyderabad cricket team will win the *Ranji Trophy* in 2015."

- *Equiprobability*: One can judge that *X* and *Y* are determinately equally likely even though one lacks sharp probabilities for both.

- *Dominance*: One can judge that *X* is determinately more likely than *Y* even though one lacks sharp probabilities for both.

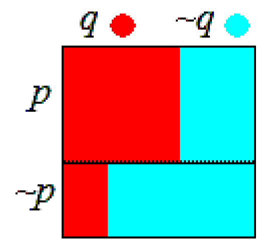
  Example: I have no idea how easy it is to gain admission to State U., but I know State cares only about high-school GPA and athletic ability. I know Jane has better grades and is better at sports than Joe. I don't know how likely either is to get in, but I know that Jane is more likely to get in than Joe.

- *Complementarity*: One can judge that *X* is as likely given *Z* as ~*X* is given ~*Z* even though one lacks a sharp probability for *X*.

  Example: Jekyll and Hyde never show up at the same party, but one of them always comes to any party to which Jekyll has been invited. I have no idea whether Jekyll was invited to tonight's party. But, I know that Hyde is exactly as likely to attend given that Jekyll drank the potion as Jekyll is to attend if he did not drink the potion.

APPLICATION: PROBABILISTIC MODUS PONENS

Oaksford and Chater argue that simple inferences involving conditionals are best understood probabilistically.

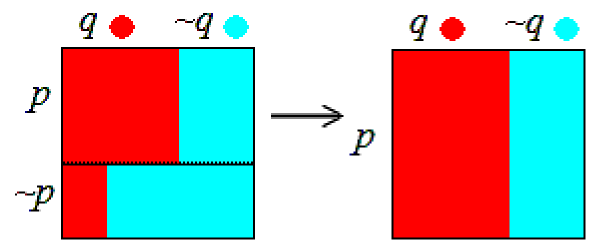


MP If *p* then *q* and *p*. Thus, *q*. VALID

$\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = 1$. Therefore, $\boldsymbol{c}_1(q) = x$.

Application: Probabilistic Modus Ponens

Oaksford and Chater argue that simple inferences involving conditionals are best understood probabilistically.



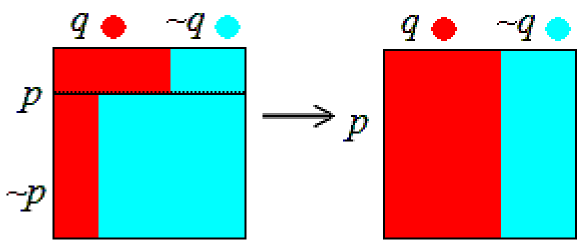
MP If p then q and p. Thus, q. VALID

$\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = 1$. Therefore, $\boldsymbol{c}_1(q) = x$.

So, reasoning by MP is effectively equivalent to Bayesian conditioning on the antecedent.

APPLICATION: PROBABILISTIC MODUS PONENS

Oaksford and Chater argue that simple inferences involving conditionals are best understood probabilistically.



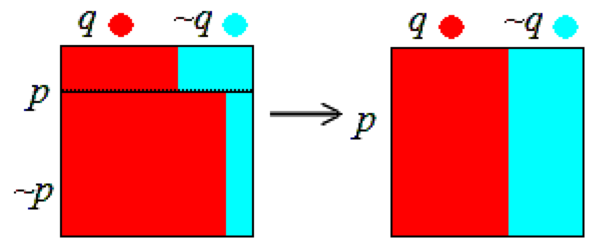
MP If *p* then *q* and *p*. Thus, *q*. VALID

$\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = 1$. Therefore, $\boldsymbol{c}_1(q) = x$.

Note: The results of the inference do not depend on prior probability of *p*.

APPLICATION: PROBABILISTIC MODUS PONENS

Oaksford and Chater argue that simple inferences involving conditionals are best understood probabilistically.



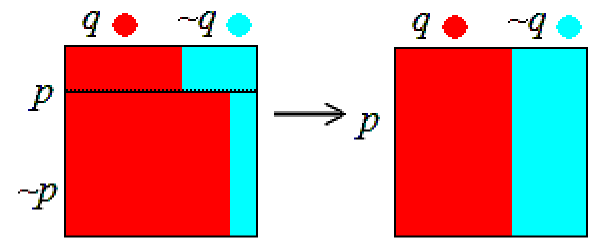
MP If *p* then *q* and *p*. Thus, *q*. VALID

$\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = 1$. Therefore, $\boldsymbol{c}_1(q) = x$.

Note: The results of the inference do not depend on prior probability of *p* or on the probability of *q* or the probability of *q* given ~*p*.

APPLICATION: PROBABILISTIC MODUS PONENS

Oaksford and Chater argue that simple inferences involving conditionals are best understood probabilistically.



MP If p then q and p. Thus, q. VALID

$\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = 1$. Therefore, $\boldsymbol{c}_1(q) = x$.

Note: The results of the inference do not depend on prior probability of p or on the probability of q or the probability of q given $\sim p$.

- Agents drawing **MP** inferences are portrayed as probabilistically reasonable.
- *Rigidity*. It is assumed that one learns the unconditional premise at $t = 1$ in a way that does not undermine the conditional premise known at $t = 0$. (If Bond tries to kill me, I'll never know. Bond tries to kill me. So, I'll never know.)

Problem (?). This suffers from the requirement that the categorical premise is learned with certainty.

APPLICATION: PROBABILISTIC AFFIRMING THE CONSEQUENT

AC If *p* then *q* and *q*. Thus, *p*. INVAID, OFTEN PLAUSIBLE

$\boldsymbol{c}_0(q \mid p) = x$, $\boldsymbol{c}_1(q) = 1$. Thus, $\boldsymbol{c}_1(p) = x{\cdot}y/z$, on the assumption that $\boldsymbol{c}_0(q) = z$ and $\boldsymbol{c}_0(p) = y$ are both known.

Again, subjects drawing **AC** inferences are portrayed as probabilistically reasonable, though when the subject has no priors for *p* and *q*, the inference is unreasonable.

General Principle (**MP**, **AC**, **MT**, **DA**): As result of the inference, the subject assigns a conditional probably to the conclusion given the categorical premise.

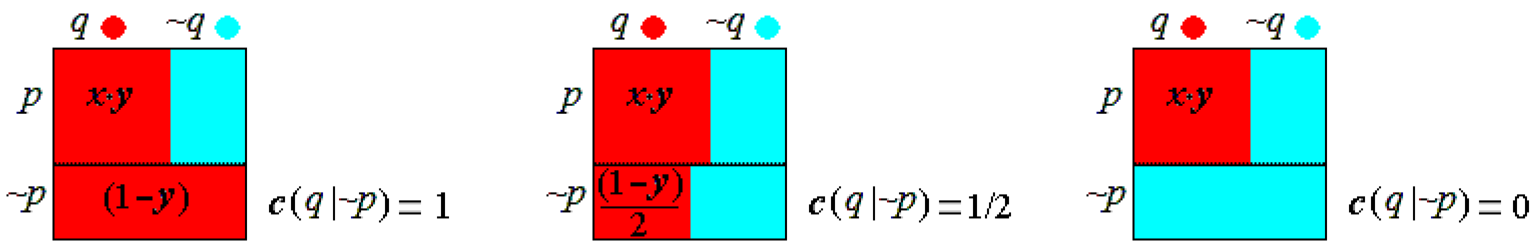
This assumes (a) rigidity holds (often a problem for **MT** inferences), and (b) the subject has sufficient additional knowledge to determine the relevant conditional probability given the information in the premises (needed for all but **MP**).

Two problems:

- This suffers from the requirement that the condition is learned with certainty.
- **AC** only applies when the agent has determinate priors for *p* and *q*.

Precise or Imprecise MP?

O & C note that combining $\boldsymbol{c}_0(q \mid p) = x$ and $\boldsymbol{c}_1(p) = y$ in **MP** *bounds* q's probability (via Jeffrey conditioning): $x{\cdot}y \leq \boldsymbol{c}_1(q) \leq x{\cdot}y + (1 - y)$.



This can be interpreted as either:

Precise **MP**. The conclusion is $\boldsymbol{c}_1(q) = z$ for some definite $z \in [x{\cdot}y, x{\cdot}y + (1 - y)]$.

Here, those who draw **MP** inferences are portrayed as supplementing the premises with prior beliefs in the form of a sharp probability for $\boldsymbol{c}_0(q \mid {\sim}p)$.

Imprecise **MP**. The conclusion has the form $\boldsymbol{c}_1(q) \in [x{\cdot}y, x{\cdot}y + (1 - y)]$.

Here, those who draw **MP** inferences are portrayed as concluding only what can be deduced from the premises, without bringing in information about $\boldsymbol{c}_0(q \mid {\sim}p)$.

O & C seem inclined toward Precise **MP**.

AN ALTERNATIVE: PFEIFER AND KLEITER ON "MENTAL PROBABILITY LOGIC"

Pfeifer and Kleiter (2007): "If a person is uncertain, probability logic supposes that human subjects make coherent imprecise probabilistic assessments."

P & K think of this reasoning as a matter of *deducing* probability intervals from premises.

Imprecise **MP** (and **AC**, **DA**, **MT**) are portrayed as (often) involving deductive inferences from probabilistic premises, which may be imprecise, to conclusions of the form "the probability of event E is no less than l_E and no greater than u_E."

Uncertainty explicitly represented by probability *intervals*.

OAKSFORD AND CHATER ON A "FUNDAMENTAL COMPUTATIONAL BIAS"

Response to criticism by P & K that suggests that people do reason about imprecise probabilities, O & C invoke a "fundamental computational bias" (Stanovich & West, 2000)

> FCB: People automatically import prior beliefs into inferences, and cannot avoid doing so even when they try.
>
> > *Logic Teachers' Lament*. If Jim runs three miles every day for a month then he will look like a fit 20-year-old. Jim will run three miles every day for a month. So, Jim will look like a fit 20-year-old.

O & C conclude (2009, p. 109): "It is unlikely that people reason deductively about probability intervals."

Quite an inferential leap!

A Charitable Reading: People tend to draw probabilistic inferences only when their background beliefs, when conjoined with the premises, permit them to assign fairly precise probabilities to the conclusion.

AN ILLUSTRATIVE EXPERIMENT (P & K, 2007)

Subjects were posed this inference problem:

80% of the red cars on the lot are two door-cars.	$c(\textit{2-door} \mid \textit{Red}) = 0.8$
90% of the cars on the lot are red cars.	$c(\textit{Red}) = 0.9$
How many cars are two-door cars?	
	$c(\textit{2-door}) \in [0.72\ ,\ 0.82]$

Subjects could choose to answer either

- Exactly ___% of the cars are two-door cars.
- At least ___% and at most ___% of the cars are two-door cars.

Some findings, which remained fairly consistent when the probabilities were varied (but small sample, $n = 15$, and an easy problem):

- About $^2/_3$ of subjects *preferred* to state an interval.
- Almost everybody stated a lower bound (or point value) above 0.72, and about 2/3 stated an upper bound below 0.82.

It's hard to know what to make of this, but it does seem that some people think in terms of imprecise probabilities when evidence is insufficient to warrant precise assignments.

Middle Ground?

The FCB is undeniable, but it is also true that we almost never think or act like precise Bayesians (except, perhaps, in casino's).

Objective Bayesian Insight. Evidence often imposes direct constraints on credal states.

E.g., symmetry requirements: $\boldsymbol{c} \in \mathcal{C}$, then $\boldsymbol{c}(\bullet) = \boldsymbol{c}(\sim X)\cdot\boldsymbol{c}(\bullet \mid X) + \boldsymbol{c}(X)\cdot\boldsymbol{c}(\bullet \mid \sim X) \in \mathcal{C}$.

It is irrational to have a creedal states that violates these objective constraints, but they are typically far from sufficient to pick out a unique credence function (and appeals to MaxEnt or Insufficient Reason principles are illegitimate).

Subjective Bayesian Insight. People typically, and legitimately, augment the data they receive with subjective background beliefs, which further restricts their credal states.

Subjective beliefs are especially important in determining inductive policies. But, they are usually far from sufficient to pick out a unique credence function.

EXAMPLE: *TOTAL* IGNORANCE AND LEARNING FROM EXPERIENCE

We know *only* that a coin might have a bias β toward heads strictly between ¼ or ¾, and that coin tossing coin is an IID process.

- Objective constraints:

Symmetry. For each $\boldsymbol{c} \in \mathcal{C}$ with $\boldsymbol{c}(\beta = {}^1/_4) = p$, there is $\boldsymbol{c}^* \in \mathcal{C}$ with $\boldsymbol{c}^*(\beta = {}^1/_4) = 1 - p$.

Note: This is weaker than asking each $\boldsymbol{c} \in \mathcal{C}$ to be symmetric, which would require $\mathcal{C}$ to be the singleton of $\boldsymbol{c}(\beta = {}^1/_4) = {}^1/_2$. Analogy: the impartial judge vs. the impartial committee.

Binomial Likelihood. For all $\boldsymbol{c} \in \mathcal{C}$, $\boldsymbol{c}(\langle h, t\rangle \mid \beta) = \binom{h+t}{h}\, \beta^h \cdot (1 - \beta)^t$ where $\langle h, t\rangle$ is the event of getting a sequence of h heads and t tails (in any order).

Based only on these constraints, we might hope to model our uncertainty by letting $\mathcal{C}$ contain all probability functions that assign a positive probability to each possibility:

$$\mathcal{C} = \{\boldsymbol{c}(\bullet): 0 < \boldsymbol{c}(\beta = {}^1/_4) < 1\}$$

This choice ignores a deeply held inductive tendency that I suspect all of us have.

ADDING DATA

We toss the coin 500 times and observe 385 heads.

- In general,

$$\boldsymbol{c}(\beta = {}^{1}\!/_{4} \mid \langle h, t\rangle) = \frac{\boldsymbol{c}(\beta = {}^{1}\!/_{4})\cdot 3^{t}}{\boldsymbol{c}(\beta = {}^{1}\!/_{4})\cdot 3^{t} + \boldsymbol{c}(\beta = {}^{3}\!/_{4})\cdot 3^{h}}$$

- We seem to have *overwhelming* evidence for $\beta = {}^{3}\!/_{4}$.

  Even if we start with $\boldsymbol{c}(\beta = {}^{3}\!/_{4})$ as small as one in a quadrillion, the value of $\boldsymbol{c}(\beta = {}^{3}\!/_{4})$ given this data will be 0.999…..9 with the '9's extending past 100^{th} decimal place!

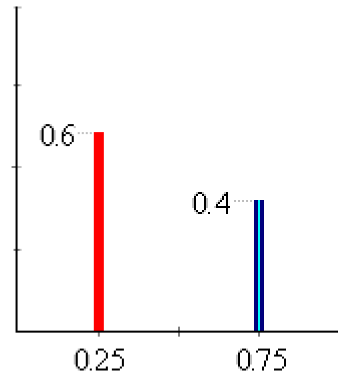
  This concentration of probability will be even more extreme if we start with a larger $\boldsymbol{c}(\beta = {}^{3}\!/_{4})$.

So, we should expect a great deal of inductive learning to take place here.

TOTAL IGNORANCE PRECLUDES LEARNING FROM EXPERIENCE!

FACT: If $\mathcal{C} = \{\boldsymbol{c}(\bullet) : 0 < \boldsymbol{c}(\beta = {}^{1}/_{4}) < 1\}$, then $\mathcal{C}_{\langle h,\, t\rangle} = \{\boldsymbol{c}(\bullet \mid \langle h, t\rangle): \boldsymbol{c} \in \mathcal{C}\}$ is *identical* to $\mathcal{C}$.

Pf: For any $p \in (0, 1)$ if we assume the prior is $\boldsymbol{c}(\beta = {}^{1}/_{4}) = p{\cdot}3^{h} / (p{\cdot}3^{h} + (1 - p){\cdot}3^{t})$, then the posterior is $\boldsymbol{c}(\beta = {}^{1}/_{4} \mid \langle h, t\rangle) = p$.

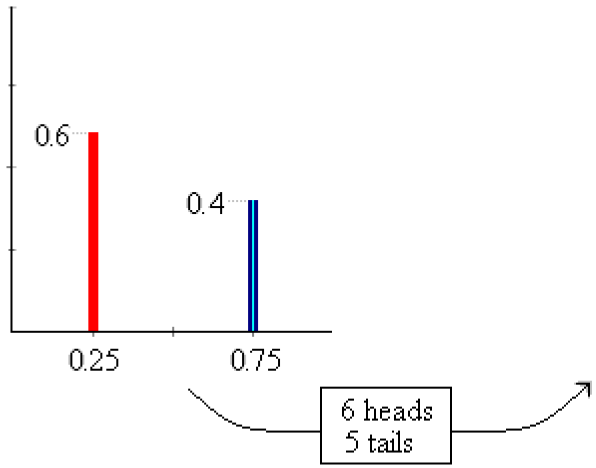


Prior Probability

TOTAL IGNORANCE PRECLUDES LEARNING FROM EXPERIENCE!

FACT: If $\mathcal{C} = \{\boldsymbol{c}(\bullet) : 0 < \boldsymbol{c}(\beta = {}^1/_4) < 1\}$, then $\mathcal{C}_{\langle h,\, t\rangle} = \{\boldsymbol{c}(\bullet \mid \langle h, t\rangle): \boldsymbol{c} \in \mathcal{C}\}$ is *identical* to $\mathcal{C}$.

Pf: For any $p \in (0, 1)$ if we assume the prior is $\boldsymbol{c}(\beta = {}^1/_4) = p\cdot 3^h / (p\cdot 3^h + (1 - p)\cdot 3^t)$, then the posterior is $\boldsymbol{c}(\beta = {}^1/_4 \mid \langle h, t\rangle) = p$.

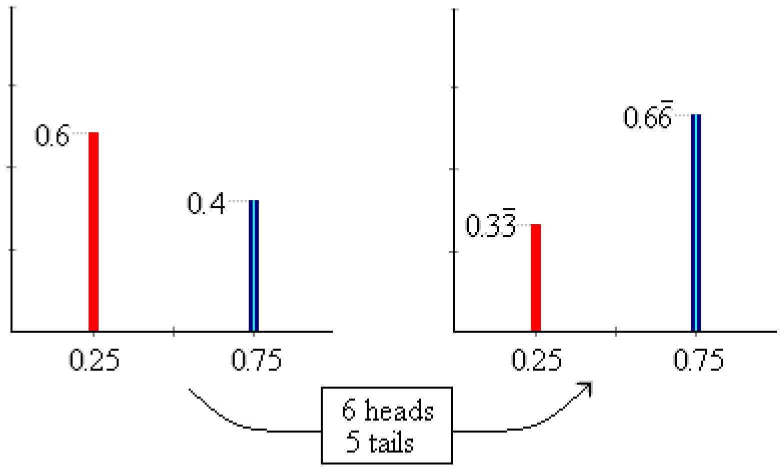


Prior Probability + Data

TOTAL IGNORANCE PRECLUDES LEARNING FROM EXPERIENCE!

FACT: If $\mathcal{C} = \{\boldsymbol{c}(\bullet) : 0 < \boldsymbol{c}(\beta = {}^1/_4) < 1\}$, then $\mathcal{C}_{\langle h,\, t\rangle} = \{\boldsymbol{c}(\bullet \mid \langle h, t\rangle) : \boldsymbol{c} \in \mathcal{C}\}$ is *identical* to $\mathcal{C}$.

Pf: For any $p \in (0, 1)$ if we assume the prior is $\boldsymbol{c}(\beta = {}^1/_4) = p{\cdot}3^h / (p{\cdot}3^h + (1 - p){\cdot}3^t)$, then the posterior is $\boldsymbol{c}(\beta = {}^1/_4 \mid \langle h, t\rangle) = p$.



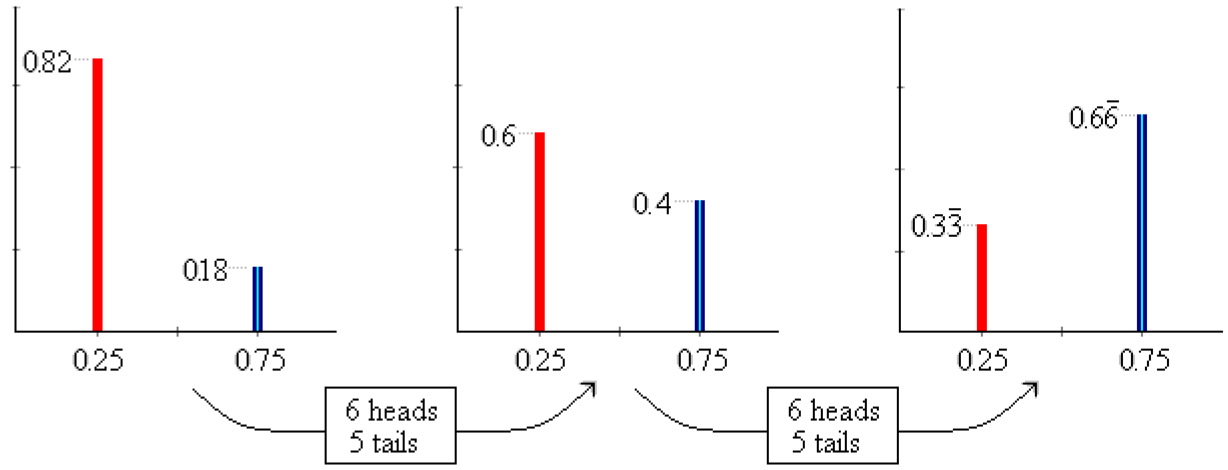
Prior Probability + Data = Posterior

Inductive Learning?

TOTAL IGNORANCE PRECLUDES LEARNING FROM EXPERIENCE!

FACT: If $\mathcal{C} = \{\boldsymbol{c}(\bullet) : 0 < \boldsymbol{c}(\beta = {}^1\!/_4) < 1\}$, then $\mathcal{C}_{\langle h,\, t\rangle} = \{\boldsymbol{c}(\bullet \mid \langle h, t\rangle) : \boldsymbol{c} \in \mathcal{C}\}$ is *identical* to $\mathcal{C}$.

Pf: For any $p \in (0, 1)$ if we assume the prior is $\boldsymbol{c}(\beta = {}^1\!/_4) = p \cdot 3^h / (p \cdot 3^h + (1 - p) \cdot 3^t)$, then the posterior is $\boldsymbol{c}(\beta = {}^1\!/_4 \mid \langle h, t\rangle) = p$.



More Extreme Prior + Data = Prior Probability + Data = Posterior

So, *nothing* is learned even in the simple case of IID coin tossing.

This looks terrible for the imprecise view! But is it really?

TWO WAYS TO THINK ABOUT LEARNING FROM EXPERIENCE

1. There is a sense in which inductive learning does occur. *From the perspective of the imprecise prior* all of the following are determinately true:

- 385 heads and 115 tails is strong evidence for $\beta = {}^3/_4$. As measured, e.g., by a likelihood ratio of 3^{270}.
- The probability of $\beta = {}^3/_4$ increases as a result of conditioning on the data, albeit not by any determinate amount. (For each $\boldsymbol{c} \in C$, the probability of $\beta = {}^3/_4$ increases by a factor of $(\boldsymbol{c}(\beta = {}^1/_4)\cdot 3^{-270} + \boldsymbol{c}(\beta = {}^3/_4))^{-1} > 1$.)
- The distributions become more concentrated around the probability of *heads* (the variance shrinks, though not by any determinate amount).
- There is *convergence* of opinion, of a sort. Given any $\boldsymbol{c}, \boldsymbol{d} \in C$, if we let $||\boldsymbol{c} - \boldsymbol{d}|| = (\boldsymbol{c}(\beta = {}^1/_4) - \boldsymbol{d}(\beta = {}^1/_4))^2 + (\boldsymbol{c}(\beta = {}^3/_4) - \boldsymbol{d}(\beta = {}^3/_4))^2$ measure the distance between $\boldsymbol{c}$ and $\boldsymbol{d}$, then $||\boldsymbol{c} - \boldsymbol{d}|| > ||\boldsymbol{c}_{\langle h,t\rangle} - \boldsymbol{d}_{\langle h,t\rangle}||$.

All great stuff, but…

2. In another sense, we learn nothing. If we have no reason to exclude any prior over $\{\beta = ¼, \beta = ¾\}$, then we have no reason to exclude any posterior either.

The argument from "the prior $\boldsymbol{c}(\beta = {}^1/_4) = p{\cdot}3^h / (p{\cdot}3^h + (1 - p){\cdot}3^t)$ is consistent with the old data" to "the posterior $\boldsymbol{c}(\beta = {}^1/_4 \mid \langle h, t\rangle) = p$ is consistent with the new data" seems unassailable.

So, when we consider things from a third-person standpoint, rather than from the perspective of the prior, we are in no better epistemic position with respect to the coin's bias than we were before.

Both **1** and **2** say something true about our evidential situation, but each leaves something important out. Analogy: The Hilbert Hotel with angry guests.

- It's hard to be happy with this situation. While we can say up front that the data has all the right evidential properties, acquiring this data it has no affect at all on our beliefs because we started out from such an extreme position.

- The problem is that $\mathcal{C}$ contains probabilities that misrepresent central aspects of our doxastic situation.

A FRUITFUL WAY TO THINK

We can write each $\boldsymbol{c} \in \boldsymbol{C}$ as a *beta distribution*

$$\boldsymbol{c}(\beta) = \frac{\beta^{s\cdot m - 1}\cdot(1-\beta)^{s\cdot(1-m)-1}}{\beta^{s\cdot m - 1}\cdot(1-\beta)^{s\cdot(1-m)-1} + \beta^{s\cdot(1-m)-1}\cdot(1-\beta)^{s\cdot m - 1}}$$

where $m = \boldsymbol{c}(heads)$ and $s = \log_B({}^{m-\beta}/_{1-\beta-m})\cdot(2m-1)^{-1}$ for $B = {}^{1-\beta}/_{\beta}$.

- This assumes that $m \neq ½$. When $m = ½$, s is determined by taking a limit
- In our example, $\boldsymbol{c}(\beta = {}^1/_4) = (1 + 3^{\,s\cdot(2\cdot m-1)})^{-1}$ and $s = \log_3({}^{4m-1}/_{3-4m})\cdot(2m-1)^{-1}$, and $s \approx 3.64096$ at $m = ½$.

☆ The parameter s reflects $\boldsymbol{c}$'s tendency to give weight the prior as opposed to the data (Walley, 1996).

$$\boldsymbol{c}(heads \mid \langle h, t\rangle) = (h + s\cdot m)/(h + t + s)$$

Compare: The Johnson/Carnap "continuum of inductive methods."

IMPORTANT FACTS ABOUT s

- When $m = \frac{1}{2}$, s is at its minimum and the prior carries the least weight.

- For $\frac{1}{4} < m < \frac{1}{2}$, each value of s *uniquely* fixes m and $\boldsymbol{c}$.

- $\boldsymbol{c}(heads \mid \langle h, t\rangle) = m$ in the limit as $s \to \infty$, and so the data carries no weight.

  s increases as one moves toward the extreme points $\boldsymbol{c}(\beta = {}^{1}/_{4}) = 1/0$.

- The value of s is the same for complementary credence functions.

  $s = s^{*}$ when $\boldsymbol{c}(\beta = {}^{1}/_{4}) = 1 - \boldsymbol{c}^{*}(\beta = {}^{1}/_{4})$.

- Evidence always increases s.

$$\text{Posterior:}\quad \boldsymbol{c}(\beta \mid \langle h, t\rangle) = \frac{\beta^{h + s\cdot m - 1}\cdot(1-\beta)^{t + s\cdot(1-m) - 1}}{\beta^{h + s\cdot m - 1}\cdot(1-\beta)^{t + s\cdot(1-m) - 1} + \beta^{h + s\cdot(1-m) - 1}\cdot(1-\beta)^{t + s\cdot m - 1}}$$

so that $s_{c(\bullet \mid \langle h, t\rangle)} = h + t + s$.

Interpretation of *s*

s measures the *resilience* of the probability assignments in ***c*** in a way that increases with the *weight* of the evidence but does not depend on the evidence's *valence*.

- To specify any probability in *C* only need specify its resilience level and its valence ($m >,< {}^{1}/_{2}$).
- Large values of *s*, which correspond to extreme credences, indicate a high level of dogmatic inflexibility in the face of evidence.
- The possibility of inductive learning depends on having attitudes that initially exclude this kind of inflexibility, so that *s* is restricted to ranges in which $\boldsymbol{c}(\beta = {}^{1}/_{4})$ is bounded away from zero/one.
- As evidence accumulates, even subjects who start out flexible (those whose credal states contain probabilities with low *s*-values) become increasingly inflexible.
- The range of *s*-values that a subject countenances at any point reflects her willingness to be lead by evidence or, alternatively, her level of self-assurance about the correctness of her own beliefs.

Note: This way of thinking, does *not* require us to appeal to any probability measure over *C*. We do not say "she believes her beliefs are probably right." Rather, "she updates her beliefs as if they are probably right."

THE SUBJECTIVE AND OBJECTIVE DETERMINANTS OF *C*

Our subject's credal state is determined by both

i. the objective constraints of Symmetry and Binomial Likelihood

ii. her own subjective tendency to respond to evidence in more or less dogmatic ways.

- The insistence on Symmetry rules out any form of *Radical Subjective Bayesianism* that permits the subject to have an imprecise credal state, or a sharp credence function, in which it is determinate that the probability of heads differs from $^1/_2$.

Three Possibilities Consistent with the Objective Constraints

- *Precise Objective Bayesianism*: Rationality requires believers to have maximum doxastic flexibility, so that *s* is minimized. There is a uniquely rational initial inductive policy – MaxFlex or MinResil – and it requires and $\boldsymbol{c}(\beta = {}^1/_4) = {}^1/_2$.

  This is a different way of arguing for MaxEnt, so as to get $\boldsymbol{c}(\beta = {}^1/_4) = {}^1/_2$ *a priori*.

- *Imprecise Objective Bayesianism*: Rationality prohibits believers from having any determinate level of doxastic flexibility since nothing in the objective constraints justifies such a stance. So, any consistent value of *s* should be allowed (inductive learning be damned).

- *Moderate Subjective Bayesianism*: Objective constraints must be respected, but beyond that any level of doxastic flexibility is consistent with the demands of rationality.

  A believer can have a credal state $C = \{\boldsymbol{c}(\bullet): \varepsilon < \boldsymbol{c}(\beta = {}^1/_4) < 1 - \varepsilon\}$, where of $\varepsilon > 0$ is fixed by personal inductive policies that bound *s* below infinity.

I like the last view, but there are some issues.

Empirical Question: Does MSB capture the way people actually think?

I suspect it correctly describes the attitudes of many people.

- We are uncomfortable saying that, definitely, the probability of heads is ½.
- We are uncomfortable with the idea that prior beliefs that are largely uninformed by evidence should be weighted heavily relative to future data.
- We also have no precise views about strong or weak the effect of priors should be.

Philosophical Question: Is MSB the right way to think about rational belief, and does it work within the "rational analysis" framework?

This is less clear.

- MSB is a kind of half-way house between the two objective views, whose proponents will argue that there is no principled place to stop anywhere between minimizing S, so that $S = \min_S$, and leaving S entirely unconstrained, so that Range(S) = $[\min_S, \infty)$.

THE CASE FOR MINIMIZING *s*

$s = \min_s$ is the most modest choice one can make since it assumes, to the greatest degree possible, that the believer's (pre data) judgments are fallible.

Response: Not really. The claim is true if you compare $s = \min_s$ to any other choice of a *sharp* value for s. But, anyone with a sharp s makes highly *immodest* probability assignments. E.g., setting $s = \min_s$ *commits* one to believing (and acting as if) the probability of getting 20 heads and 10 tails is 0.000172067, where *all* those digits are significant. Given what you know, would you take a bet that hangs on the value of that last "7"?

From the perspective of probability assignments, epistemic modesty involves allowing s to have a range of values.

THE CASE FOR LEAVING s UNCONSTRAINED

Range(s) = [$\min_s$, ∞) is the only policy that does not add information to the objective constraints. Our feeling that inductive learning should be possible here is based on a confusion. We typically face finite problems in which an objective lower bound for s *is* fixed. This puts us in the habit of thinking that inductive learning should always be possible. But, when we don't know *anything* about s, we should not capriciously assume it has a lower bound. To do so is to introduce an unwarranted *inductive bias*.

> Response: There is something to this. We *are* accustomed to finite problems – the urn surely can't contain more than 10^{50} coins! – and it might be that scenarios in which we know *nothing* are so rare that people misunderstand them. This might not be too great a cost if our natural inductive tendencies agree with objective constrains in realistic cases.
>
> On the other hand, it does seem that some level of epistemic modesty is warranted even in situations of extreme ignorance. In the same way that adopt the reasonable policy of thinking that events of *very* small probability do not occur (see (Shafer 2008) on "Cournot's Principle"), it might be that we rightly assume that highly extreme distributions, short of 0/1, do not accurately reflect reality.

Some Conclusions

- The theory of imprecise probabilities might be the right tool for use in a broadly Bayesian project of "rational analysis."

- The theory has the advantage of being able to express a range of epistemic attitudes that cannot be described using precise probabilities (e.g., bare judgments of independence or disagreement).

- This extra expressive power carries some risks, however, because it lets us to express attitudes that real subjects might well be incapable of having.

- An especially fruitful area for the theory is the connection between beliefs and policies of inductive learning, since the theory can capture subtle differences in attitudes of doxastic flexibility in face of evidence.

- Within this an imprecise framework a Bayesian can make room for the joint effects of objective evidential constraints and subjective judgments on beliefs.

- But, the precise way to understand these effects remains a matter of dispute.

Some Useful References

Anderson, J. R. 1991. “The Place of Cognitive Architectures in a Rational Analysis,” pp. 1-24 in K. Van Len, Ed., *Architectures for Intelligence*. Hillsdale, NJ: Erlbaum.

Couso, I., Moral, S. and Walley, P. 2000. “A survey of concepts of independence for imprecise probabilities,” *Risk, Decision and Policy*: 165-181.

Fisher, R. A. 1922. “On the Mathematical Foundations of Theoretical Statistics,” *Philosphical Transactions of the Royal Society on London* **Ser. A 222**: 309-368.

Grünwald, P. and Halpern, J. 2004. “When Ignorance is Bliss,” *ACM International Conference Proceeding Series; Vol. 70: Proceedings of the 20th conference on Uncertainty in artificial intelligence***.**

Joyce. J. “How Degrees of Belief Reflect Evidence,” *Philosophical Perspectives* **19**, 2005: 153-179.

Levi, I. 1974. “On Indeterminate Probabilities", *Journal of Philosophy* **71**: 397-418.

Levi, I. 1980. *The Enterprise of Knowledge*. Cambridge: M.I.T. Press.

Pfeifer N. and Kleiter G. D. 2007. “Human Reasoning with Imprecise Probabilities: Modus Ponens and Denying the Antecedent,” 5th International Symposium on Imprecise Probability: Theories and Applications, Prague, Czech Republic, 2007

Seidenfeld, T. and Wasserman, L. 1993. “Dilation for Convex Sets of Probabilities,” *Annals of Statistics* **21**: 1139–1154.

Shafer, G. 2008. “A Betting Interpretation for Probabilities and Dempster-Shafer Degrees of Belief http://www.glennshafer.com/assets/downloads/dsbetting.pdf

van Fraassen, B. 2006. ‘Vague Expectation Loss’ *Philosophical Studies* **127**: 483-491.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.

------- 1996. “Inferences from Multinomial Data: Learning about a Bag of Marbles,” *Journal of the Royal Statistical Society, **Series B,*** **58**:3-57.

White, R. “Evidential Symmetry and Mushy Credence,” forthcoming in Oxford Studies in Epistemology.