Volume 1
Knowledge and Belief
Jaakko Hintikka

Volume 2
Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr.
William Harper and Gregory Wheeler, eds

Volume 3
Monsters and Philosophy
Charles T. Wolfe, ed.

Volume 4
Computing, Philosophy and Cognition
Lorenzo Magnani and Riccardo Dossena, eds

Volume 5
Causality and Probability in the Sciences
Federica Russo and Jon Williamson, eds

Texts in Philosophy Series Editors
Vincent F. Hendricks vincent@ruc.dk
John Symons jsymons@utep.edu

Probability and Inference

Essays in Honour of Henry E. Kyburg, Jr.

edited by

William Harper
and
Gregory Wheeler

[Ellsberg, 1961] D. Ellsberg. Risk, Ambiguity and the Savage Axioms, *Quarterly Journal of Economics*, 75:643–69, 1961.
[Ellsberg, 2002] D. Ellsberg. *Risk, Ambiguity and Decision*, New York: Garland. (Publication of 1962 Ph.D. dissertation), 2002.
[Fisher, 1956] R. A. Fisher. *Statistical Methods and Scientific Inference*, New York: Hafter, 1956.
[Good, 1951] I. J. Good. Rational Decisions, *Journal of the Royal Statistical Society*, Ser.B. 14: 107–14, 1951.
[Good, 1961] I. J. Good. Subjective Probability as a Measure of a Nonmeasurable set. In *Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress*, Stanford: Stanford University Press, 319–29, 1961.
[Jeffrey, 1965] R. Jeffrey. *Logic of Decision*, New York: McGraw Hill, 1965.
[Jeffreys, 1939] H. Jeffreys. *Theory of Probability*, Oxford University Press, 1939.
[Koslow, 1992] A. Koslow. *A Structural Theory of Logic*, Cambridge: Cambridge University Press, 1992.
[Kyburg, 1961] H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief*, Middletown, Conn.: Wesleyan University Press, 1961.
[Kyburg, 1974] H. E. Kyburg, Jr. *The Logical Foundations of Statistical Inference*, Dordrecht: Reidel, 1974.
[Kyburg and Teng, 2001] H. E. Kyburg, Jr and C. M. Teng. *Uncertain Inference*, Cambridge: Cambridge University Press, 2001.
[Levi, 1980] I. Levi. *The Enterprise of Knowledge*, Cambridge, Mass: MIT Press, 1980.
[Levi, 1991] I. Levi. *The Fixation of Belief and Its Undoing*, Cambridge: Cambridge University Press, 1991.
[Levi, 1997] I. Levi. *The Covenant of Reason*, Cambridge: Cambridge University Press, 1997.
[Seidenfeld, 1979] T. Seidenfeld. *Philosophical Problems of Statistical Inference: Learning from R.A. Fisher*, Dordrecht: Reidel, 1979.
[Smith, 1961] C. A. B. Smith. Consistency in statistical inference, *Journal of the Royal Statistical Society*, Ser.B. 23: 1–25, 1961.

Forbidden Fruit: When Epistemological Probability may *not* take a bite of the Bayesian apple

TEDDY SEIDENFELD

1 Elementary Probability Theory and some of its *Bayesian* fruits

1.1

Unconditional Probability $P(\bullet)$ is governed by three axioms:

Axiom 1 $PZ\bullet)$ is real-valued function defined over an algebra

$$0 \le P(\bullet) \le 1$$

Axiom 2 For the sure event S, $P(S) = 1$

Axiom 3 *(additivity)* For disjoint events, where $A \cap B = \emptyset$ then

$$P(A \cup B) = P(A) + P(B)$$

Conditional Probability $P(\bullet|\bullet)$ is governed by two additional axioms:

Axiom 4 $P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$.

Axiom 5 For each $B \neq \emptyset$, $P(\bullet|B)$ is an unconditional probability.

(Aside) Axiom 5 is of concern primarily when the conditioning event, B, is null. That is, when $P(B) = 0$ Axiom 4 fails to insure that $P(\bullet|B)$ is an unconditional probability satisfying Axioms 1-3. For discussion of some of the controversial aspects of the received theory's solution to this problem using regular conditional distributions, see my [Seidenfeld, 2001]

Let $\{H_1, H_2, \ldots, H_n\}$ be a partition into n-many pairwise disjoint and mutually exhaustive states.

The Law of Total Probability asserts

$$P(A) = \sum_i P(A \cap H_i) \tag{1}$$

which follows by additivity from the elementary identities:

$$\begin{aligned} A &= A \cap S \\ &= A \cap [H_1 \cup H_2 \cup \ldots \cup H_n] \\ &= [A \cap H_1] \cup [A \cap H_2] \cup \ldots \cup [A \cap H_n] \end{aligned}$$

By the principal axiom governing conditional probability, (1) yields

$$P(A) = \sum_i P(A|H_i) \times P(H_i) \tag{2}$$

Here, $P(A|\bullet)$ is called (a version of) the *likelihood function.*

A familiar Bayesian formulation of this law is as:

unconditonal probability	equals	*expected likelihood*
$P(A)$	$=$	$\sum_i P(A\|H_i) \times P(H_i)$.

Then, unconditional probability $P(A)$ is constrained as a *convex* function of conditional probability $P(A|\bullet)$ over a partition for the argument ($\bullet$).

It is a short step from this result to *Bayes' Theorem.* By the principal axiom of conditional probability, when $P(A) \neq 0$.

$$P(H|A) = \frac{P(A|H) \times P(H)}{P(A)}$$

And by an application of the previous law:

$$= \frac{P(A|H) \times P(H)}{\sum_i P(A|H_i) \times P(H)_i)}$$

An easy calculation then yields:

$$\frac{P(H_1|A)}{P(H_2|A)} = \frac{P(A|H_1)}{P(A|H_2)} \times \frac{P(H_1)}{P(H_2)} \tag{3}$$

1.2 Conditionalisation and three *Bayesian* fruits of these probability laws

Levi's account of why *Bayes'* Theorem creates interest in *conditional probability* is that the conditional probability, $P(\bullet|H)$, is the answer to an important hypothetical question:

"What would your probability function be were your current knowledge augmented with (consistent) H?"

Conditionalisation then fixes *Bayesian* inference, as follows. In response to the question what your uncertainty would be regarding rival hypotheses, H_1 and H_2, were you to learn that A, Bayes' theorem provides a helpful algorithm:

$$\frac{P(H_1|A)}{P(H)_2|A)} = \frac{P(A|H_1)}{P(A|H_2)} \times \frac{P(H_1)}{P(H_2)}$$

It is summarized by the familiar Bayesian mantra

posterior odds = *likelihood ratio* × *prior odds.*

This mantra is particularly useful when the rivals H_1 and H_2 are simple statistical hypotheses so that the likelihood function $P(A|\bullet)$ is fixed by non-controversial inference rules, e.g., *Direct Inference.* Here are three important fruits of *Bayesian* conditionalisation:

1st *product:* Eliminate *nuisance* parameters by averaging the likelihood function.

Suppose that, in order to make the likelihood function simple – in order to apply *Direct Inference* – additional parameters J_i are specified beyond the composite hypothesis H that is the investigator's focus of interest. These nuisances J_i can be eliminated by an application of the conditional version (2*) of (2), given H,

$$P(A|H) = \sum_i P(A|H, J_i) \times P(J_i|H) \tag{2*}$$

2nd *product:* Composite data may be evaluated in any order computationally advantageous for the inference.

Suppose that the composite data are the pair (A, B) and that these are independent given the statistical hypothesis H, i.e.

$$P(A, B|H) = P(A|H) \times P(B|H)$$

or equivalently

$$P(A|H,B) = P(A|H) \text{ and } P(B|H,A) = P(B|H)$$

then

$$\begin{aligned} P(H|A,B) &\propto P(A|H) \times P(B|H) \times P(H) \\ &\propto P(A|H) \times P(H|B) \\ &\propto P(B|H) \times P(H|A) \end{aligned}$$

3^{rd} *product:* The likelihood ratio equals the ratio of posterior odds to prior odds.

$$\frac{P(A|H_1)}{P(A|H_2)} = \frac{P(H_1|A)}{P(H_2|A)} \div \frac{P(H_1)}{P(H_2)}$$

So, the distribution of the likelihood ratio, viewed before the data are collected, is one perspective on how informative an experiment will be in changing the prior to the posterior. Unless the distribution of the likelihood ratio is the degenerate, constant = 1, the experiment has positive probability of generating evidence that, were it learned, would change the investigator's mind.

A familiar likelihood based index of *information* that measures how much a probability distribution Q differs from a distribution P, both defined on a space Ω, is *Kullback-Leiber* Information:

$$KL(Q,P) = \sum_{\omega} \log\left[\frac{Q(\omega)}{P(\omega)}\right] Q(\omega) \geq 0$$

Set P to the "prior" and Q to the "posterior" given data $X = x$, with both distributions over the common space Θ of the parameter of interest. Then:

$$KL(posterior, prior) = \sum_{\theta} \log\left[\frac{P(x|\theta)}{P(x)}\right] P(\theta|x)$$

This change is 0, i.e., there is no information gained in going from the *prior* to the *posterior* if and only if $P(x|\bullet)$, the likelihood function with respect to the parameter θ, is constant. Thus, unless an experiment is almost sure to produce irrelevant information, as indexed by a constant likelihood, the expected information gain in going from the prior to the posterior is strictly positive.

2 Epistemological Probability [*EP*] Theory and some of its original features

2.1 Epistemological Probability Theory and Direct Inference

(*Historical Aside 1*) Henry Kyburg's original theory of *Epistemological Probability* [*EP*] dates, I believe, from the final chapter in his 1956 Columbia University doctoral thesis ([Kyburg, 1956]), which was a study of the *Keynesian School* of probability, titled *Probability and Induction in the Cambridge School*. Its first full-dress, public appearance was in his [Kyburg, 1961] book, *Probability and the Logic of Rational Belief*. Even as recently as ten years ago, at a conference on Keynes at Wake Forest, Kyburg promoted *EP* as his preferred interpretation of Keynesian probability theory [Kyburg, 1995], where interval-valued probability provides a formal treatment of Keynes' important idea that not all (rational) probability judgments are comparable.

The canonical form of an *EP* statement is: $EP(\phi(s); K) = [p,q]$, where:

- *EP* is an *interval-valued* probability, $[p,q]$
- that an individual s, bear property ϕ – written $\phi(s)$;
- given background knowledge K that includes
  - the frequency information that between p and q percent of the members of the reference set R bear ϕ
  - the knowledge that individual s is a member of R
  - and for each rival reference set R' to which s is known to belong, K contains no *stronger* frequency information about ϕ
  - and except for larger sets R' ($\supset R$) to which s is known to belong, K contains no different frequency information about ϕ.

In Levi's terms, each *EP* statement is an instance of *Direct Inference*:

from the knowledge of frequencies $[p,q]$ of ϕ in a population R and that $s \in R$ *to* an interval valued probability $[p,q]$ that $\phi(s)$.

2.2 Epistemological Probability Theory and Inverse Inference

What is entirely original to *EP* is how the interplay of the *strength* and *difference* clauses for fixing the winning reference class R yields important cases of statistical *Inverse Inference* derived from *Direct Inference*.
from an interval valued probability $[p,q]$ that $\phi(s)$.
to an interval valued probability $[p,q]$ of ϕ in a population R.

EXAMPLE 1. Suppose that we have a scale on which to weigh objects. Our scale is calibrated so that, within its functioning range, if an object of μ-units mass is weighed, the readings are distributed as $X \sim N(\mu, 1)$; a Normal distribution with unit variance and mean μ.

We weigh an 1878 Indian Head penny on our scale and observe that $X = c$. This reading is a sample of one from the population of measurements taken with scales of this calibration. Our background knowledge K is otherwise uninformative about the distribution of weights of 1878 Indian Head pennies, of which about 5.8 million were minted.

What is the *EP* for statements of *Inverse Inference* about μ? For instance, what is $EP(c-2 \le \mu \le c+2|X=c, K)$? The key to *EP*'s original treatment of this problem is to focus on special *pivotal* properties $\phi_r(\bullet)$ of readings from such scales.

- $\phi_r(X)$ obtains for X *if and only if* $|\mu - X| \le r (r \ge 0)$

The special feature of *pivotal* properties is that the percent of X-readings that satisfy them is known exactly, based solely on K.

$$EP(-r \le \mu - X \le r|K) = [\Phi^r_{-r}, \Phi^r_{-r}],$$

where Φ^r_{-r} is the probability that a $N(0,1)$ variate has its value in the interval $[-r, r]$. For instance, $EP(-2 \le \mu - X \le 2|K) \approx [.95, .95]$.

By the *Strength* rule, this yields a precise *EP Inverse* statement

$$EP(-2 \le \mu - c \le 2|X = c, K) \approx [.95, .95]$$

or

$$EP(c-2 \le \mu \le c+2|X = c, K) \approx [.95, .95]$$

EXAMPLE 2. Suppose that, as before, we have our scale with which to weigh objects. Our scale is calibrated so that, within its range, if an object of μ-units mass is weighed, the separate readings X_i of the same object are identically and independently distributed [*iid*] $X_i \sim N(\mu, \sigma^2)$, a Normal distribution with mean μ and *unspecified* variance σ^2. We take n readings $\tilde{x} = (x_1, \ldots, x_n)$ of our 1878 penny. What is the $EP(c-2 \le \mu \le c+2|\tilde{x}, K)$? This problem is importantly different from the first because, though μ remains the *parameter of interest*, in this version σ^2 is a *nuisance parameter* whose value we do not know.

Again, there is a special (*Student's t*) pivotal property to deal with the inference about μ in the absence of knowledge of σ^2.

$$\phi'_r(\bar{X}) \text{ obtains if and only if } \frac{|\mu - \bar{X}|}{S} \le r$$

where $\bar{X} = \sum_{i=1}^n X_i$ and $S^2 = \frac{\sum_{i=1}^n (X-\bar{X})^2}{n(n-1)}$. For instance, with $n = 2$ and $r = |X_1 - X_2|/2$, we have

$$EP(X_{min} \le \mu \le X_{max}|\bar{X}, K) = [.5, .5].$$

And by *EP*'s *Strength* rule for determining the reference class in a Direct Inference, we conclude the *Inverse EP* statement

$$EP(x_{min} \le \mu \le x_{max}|\bar{X} = \tilde{x}, K) = [.5, .5].$$

(*Historical Aside 2*) Though Kyburg developed this mode of Inverse reasoning to show how Keynes' 1921 theory of probability – a theory that allowed (logical) probability to take non-real values – might be interpreted inside a theory of interval-valued probability, in fact, *EP* really is a wonderful and fully principled generalization of R.A.Fisher's 1930 enigmatic proposal of *fiducial* probability.

In what I think was the last of 3 rounds of correspondence exchanged with Kyburg, during Fisher's last year of life, Fisher began his letter,

> After a long while I have now succeeded in obtaining your book on Rational Belief. So far it seems to be as good as I had hoped, which would be high praise. (14 May, 1962)

But also Kyburg was mildly criticized by Fisher in a way that I suspect no other had ever been. In a 13 January 1962 letter to the Canadian Statistician D.A.Sprott, Fisher wrote,

> Do you know the name of H. Kyburg of the Rockefeller Institute 21, N.Y.? His line seems to be abstract symbolic logic, but he has recently caught fire on the fiducial argument and indeed may be exaggerating its importance"

During his long and influential career, Fisher showed no restraint criticizing many for failing to appreciate the importance to Statistics of fiducial reasoning. (See, e.g., [Fisher, 1973], section III.3.) But, Kyburg was singled out, and is unique among Fisher's targets I believe, for having committed the other error!

3 When Epistemological Probability may not go *Bayesian*!

3.1 *EP* Theory and Statistical Inference with Nuisance Parameters

Approximately 28 years ago, in an article *Direct Inference*, I. Levi demonstrated that *EP* does not satisfy *Bayesian* conditionalisation ([Levi, 1974]).

Levi's counterexample highlighted some anti-Bayesian features of the *Strength* rule: the rule to give priority to reference sets that yield precise, i.e. narrower probability intervals.

EXAMPLE 3 (Levi, 1977). Suppose we know that Petersen (denoted s) is a Swedish resident of Malmo. We are interested in the *EP* that he is a Protestant. Our rational corpus of knowledge includes the following frequency facts about the two competing reference sets: Swedes, and residents of Malmo.

- We know that 90% of Swedes are Protestants.
- But all we know about Malmo is that either

  H_1 : 85% of Malmo's residents are Protestant
  or H_2 : 91% of Malmo's residents are Protestant
  or H_3 : 95% of Malmo's residents are Protestant

  with a resulting known frequency interval [.85, .95] of % Protestant.

So,

$$EP(\phi(s); K) = [.9, .9]$$

because the *Strength* rule allows the larger reference set (Swedes) to win over the rival reference set of Malmo's residents, whose frequency interval for the property in question is less informative [.85, .95].

However, *EP* theory also entails the following statements

$$EP(\phi(s); H_1, K) = [.85, .85]$$
$$EP(\phi(s); H_2, K) = [.91, .91]$$
$$EP(\phi(s); H_3, K) = [.95, .95]$$
$$EP(\phi(s); H_1 \vee H_3, K) = [.9, .9]$$
$$EP(\phi(s); H_1 \vee H_2, K) = [.9, .9]$$
$$EP(\phi(s); H_1 \vee H_2 \vee H_3, K) = [.9, .9]$$

Each of the last three of these six *EP* statements results by an application of the *Strength* rule, which picks the larger reference class (Swedes) for determining the *Epistemological Probability* that Petersen is a Protestant.

The contradiction that results is with our first elementary law:

$$P(\phi(s)) = \sum_i P(\phi(s)|H_i) \times P(H_i)$$

There is no prior distribution $P(H_i)$ over these three simple statistical hypotheses that satisfies all six EP values. In other words, *EP* theory does not follow the

Bayesian law that there exists a *prior*, $P(H_i)$, against which one may average the likelihood function.

The second Bayesian version of this law is that we may eliminate *nuisance parameters* J_i by an application of the rule:

$$P(A|H) = \sum_i P(A|H, J_i) \times P(J_i|H)$$

If, to the contrary, *EP* theory followed this law, then in good Bayesian style, we could eliminate nuisance parameters by averaging them with other *EP* probabilities.

In our second example of *EP* inference, where $X \sim N(\mu, \sigma^2)$, μ is the parameter of interest, and σ^2 is the nuisance parameter. A Bayesian elimination of σ^2 can go like this:

$$p(\mu|\bar{x}) = \int p(\mu|\bar{x}, \sigma^2) p(\sigma^2|\bar{x}) dp(\sigma)$$

EP theory provides precise probabilities for each of the terms on the right-hand side of this equation. But it does not take a bite of the *Bayesian* apple! This calculation is invalid. Instead, (Example 2) a direct *Student's* pivotal duplicates the conclusion of this *Bayesian* inference.

In the previous case, then, *EP* theory gets to the same place it would were it *Bayesian*. But that is not always possible, as the next example illustrates.

EXAMPLE 4 (The Behrens-Fisher problem). Let $\tilde{X}_1 = (X_{11}, \ldots, X_{1n})$ and $\tilde{X}_2 = (X_{21}, \ldots, X_{2n})$ be independent *iid* samples respectively from the two Normal distributions: $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. The parameter of interest is $\delta = \mu_1 - \mu_2$. The nuisance parameter is $\xi = \frac{\sigma_1}{\sigma_2}$, about which we have no frequency information.

A *Bayesian* elimination of the nuisance parameter is as follows. Let $\tilde{X} = (\tilde{X}_1, \tilde{X}_2)$.

$$p(\delta|\tilde{x}) = \int p(\delta|\tilde{x}, \xi) p(\xi|\tilde{x}) dp(\xi)$$

Again, there are pivotal variables available for *EP* to derive precise *Inverse* probabilities for each of the two terms on the right side of this equation. However, as *EP* theory is not *Bayesian*, the calculation from right to left is invalid.

Alas, there is no *direct* pivotal available, analogous to *Student*'s t-pivotal, to solve the left-hand side. It appears that *EP* theory here is missing the pleasures of this *Bayesian* fruit. *EP* theory could take a bite of this Bayesian apple, but it does not.

However, the conflict between *EP* theory and these Bayesian laws is not merely a case of *EP* theory missing out some Bayesian consequences of what it already entails. We cannot graft onto *EP* theory these missing Bayesian conclusions, as the next example illustrates.

EXAMPLE 5 (The Hollow Cube). We are interested in the volume V of a hollow cube. We have available two sources of experimental data. We may accurately fill the hollow cube with a liquid of density, 1-unit mass/unit volume, and weigh that on our scale of known precision, resulting in the random variable $X_L \sim N(V, 1)$ Alternatively, we may cut a rod of density 1-unit mass/unit length, to the edge of the cube and weigh that on our scale: $X_R \sim N(V^{1/3}, 1)$.

As in Example 1, with either observation taken alone, there is an *Inverse EP* statement about the unknown V: With $X_L = x_L$ then *EP* entails that $V \sim N(x_L, 1)$. With $X_R = x_R$ then *EP* entails that $V^{1/3} \sim N(x_R, 1)$.

Though it is invalid by *EP* standards we may try to use the 2nd set of Bayesian laws to combine the two observations. There are three approaches:

$$\begin{aligned} p(V|x_L, x_R) &\propto p(x_L|V) \times p(x_R|H) \times p(V) \\ &\propto p(x_L|V) \times p(V|x_R) \\ &\propto p(x_R|V) \times p(V|x_L) \end{aligned}$$

EP theory does not entail a precise *prior* $p(V)$ for use in the first line. Moreover, there is no direct pivotal method using (X_L, X_R). At bottom, this is because there is no common 1-dimensional sufficient statistic for V that summarizes the 2-dimensional data.

But by the preceding results, *EP* theory entails precise (point-valued) probabilities for each term in the 2nd and 3rd lines, above. But they may not be added to EP theory. *These yield contradictory results*! This is because the Bayes-model associated with the 2nd line carries a precise, different prior for V than does the Bayes-model associated with the 3rd line. Thus, *EP* theory *must not* take this bite of the Bayesian apple as a method for combining composite data.

I do not know the full *EP* solution to the problem of the Hollow Cube. I conjecture that, because there are so many competing pivotal variables available for inference about V each yielding a different interval *EP* solution, the resulting *EP* interval estimates about V are vacuous, or nearly so. For example, in addition to the two pivotal variables relating to the inference of Example 1, each of which uses only one of the two observations, also there is the pivotal variable $[(X_L + X_R) - (V + V^{1/3})]$, which is pivotal based on the fact that the random variable $(X_L + X_R)$ has a normal distribution $N(V + V^{1/3}, 2)$. These three pivotal variables generally result in competing, precise statistical statements about V that prevent each other, by *EP*'s *Difference* rule.

The open challenge I see to *EP* theory highlighted by the Hollow Cube problem is how to combine a variety of statistical data, data that do not admit a common sufficient statistic. It appears that with a variety of evidence, within *EP* theory, an increase in the variety of evidence available may decrease the informativeness of the resulting statistical conclusions. This fact provides transition to a discussion of the third and final *Bayesian* law in Section 1 of this essay concerning the informational value of new evidence. That law says, as measured by any one of a large family of indices of statistical *Information*:

> unless an experiment is almost sure to produce irrelevant evidence, it carries a positive expected *Information* gain comparing the *posterior Information* with the *prior Information*.

In short, that law promises that changes in expected *Information* that result from *conditionalization* on new evidence will not go down, and will go up unless the data are irrelevant, as judged by the likelihood. *EP* theory does not partake in this *Bayesian Tree of Knowledge*. Is that ignorance a state of statistical bliss for *EP*?

3.2 *EP* theory and *Dilation* of interval valued probabilities.

The final contrast I want to draw is with a rival position that, like *EP*, uses interval-valued probability rather than real-valued probability, but unlike *EP* it incorporates *Bayesian conditionalization*. I. Levi's *Indeterminate Probabilities* [*IP*] provides an ideal version of such a rival theory ([Levi, 1974]). In it a rational agent's degrees of belief are represented by a convex set of $\wp$ of probabilities. The agent obeys conditionalization in the sense that the corresponding set of conditional probabilities $\{P(\bullet|H) : P \in \wp\}$ answers the question,

> What would your probability be were your current knowledge augmented with (consistent) H?

In these two rivals, *EP* and *IP* Theories, by contrast with the original (Bayesian) theory, one entirely *new* aspect of the agent's *uncertainty* of an event E is captured by the range of the probability interval for E. For example, in this new sense there is maximal uncertainty about E when the probability interval is the vacuous $[0, 1]$ range, and in this same sense that uncertainty is reduced when the probability interval for E shrinks to, say, $[.4, .7]$.

The anomalous phenomenon concerning this sense of *uncertainty*, on which I close this essay is called *dilation* (See [Seidenfeld, 1993] and [Herron, 1997]). Let experiment E carry possible outcomes $\{e_1, \ldots, e_n\}$. Let $\wp$ be a non-empty convex set of probabilities. And let B be some event of interest.

DEFINITION 6. *E dilates* the set of probabilities for *B* just in case, for $i = 1, \ldots n$,

$$inf_{\wp}P(B|e_i) < inf_{\wp}P(B) \leq sup_{\wp}P(B) < sup_{\wp}P(B|e_i)$$

In words, when *dilation* occurs, under conditionalization the hypothetical new evidence is sure to *increase* the *uncertainty* of *B*, in the sense just described.

EXAMPLE 7 (A *Heuristic Example of Dilation*). Let $P^*(\bullet)$ denote the upper probability and let $P_*(\bullet)$ denote the lower probability with respect to the set $\wp$. Suppose that *A* is a highly uncertain event. That is $P^*(A) - P_*(A) \approx 1$. Let $\{H, T\}$ indicate the flip of a fair coin with outcomes independent of *A*. That is, $P(A, H) = P(A)/2$ for each $P \in \wp$. Define event *B* by, $B = \{(A, H), (A^c, T)\}$. The situation is depicted by the familiar 2*x*2 table:

	H	H
A	B	B^c
A^c	B^c	B

Note that *B* is pivotal-like! That is, it follows, simply, that $P(B) = .5$ for each $P \in \wp$. B carries no *uncertainty* in the novel sense of uncertainty common to *EP* and *IP*.

But

$$0 \approx P_*(B|H) < P_*(B) = P^*(B) < P^*(B|H) \approx 1$$

and

$$0 \approx P_*(B|T) < P_*(B) = P^*(B) < P^*(B|T) \approx 1$$

Thus, regardless how the coin lands, the conditional probability for event *B* dilates to a large interval, from a precise value of .5. In the novel sense of uncertainty relevant to *IP*, the uncertainty for *B* increases *for certain* by conditionalizing on the outcome of the $\{H, T\}$ experiment. Thus, within *Indeterminate Probability* theory, where conditionalization obtains, new evidence may increase uncertainty for sure.

In [Seidenfeld, 1993] Theorem 4.1, we show that only the density-ratio model for statistical uncertainty among neighborhood models is immune to dilation. In that sense, dilation is not rare within *IP* theory.

Though I have not here reported the decision theory that goes together with the theory of *Indeterminate Probabilities*, it should not be surprising that a decision maker will try to avoid learning evidence that dilates probabilities. I am ready to argue that such a decision maker will pay to avoid dilation! Then, for such a decision maker, the new evidence carries negative value.

By contrast, dilation is an impossibility within *EP* theory, and for the very same reason that it resists conditionalization! The *Strength* rule, which is the culprit that prevents *EP* from being *Bayesian*, also is the reason that *EP* is immune to dilation! Within *EP* theory, the evidence that causes dilation for *Indeterminate Probability* theory, is made innocuous by *strength*. Simply put, those problematic data are treated as irrelevant! This raises the question whether ignorance of certain Bayesian methods may indeed result in a state of bliss concerning statistical inference!

BIBLIOGRAPHY

[Bennett, 1990] J.H. Bennett ed., *Statistical Inference and Analysis: selected correspondence of R. A. Fisher*. Oxford U. Press, 1990.

[Fisher, 1973] R.A. Fisher, *Statistical Methods and Scientific Inference*, (3rd ed.), Hafner Press, 1973.

[Herron, 1997] T. Herron *et. al.*, Divisive Conditioning: further results on dilation, *Philosophy of Science*. 64, pp. 411-444, 1997.

[Kyburg, 1956] H.E. Kyburg, *Probability and Induction in the Cambridge School*. Ph.D. Thesis. Columbia University: NYC, 1956.

[Kyburg, 1961] H.E. Kyburg, *Probability and the Logic of Rational Belief*. Wesleyan U Press, 1961.

[Kyburg, 1995] H.E. Kyburg, Keynes as a Philosopher, in *New Perspectives on Keynes, History of Political Economy*, Annual Supplement to Volume 27, pp. 7–32, 1995.

[Levi, 1974] I. Levi, On Indeterminate Probabilities, *J.Phil* 71, pp. 391–418, 1974.

[Levi, 1977] I. Levi, Direct Inference, *J.Phil* 74, pp. 5–29, 1977.

[Seidenfeld, 1992] T. Seidenfeld, R.A.Fisher's Fiducial Argument and Bayes' Theorem, *Statistical Science* 7: 358-368, 1992.

[Seidenfeld, 2001] T. Seidenfeld, Remarks on the Theory of Conditional Probability, in *Probability Theory: Philosophy Recent History and Relations to Science*, V.F.Hendricks, S.A.Pedersen, and K.F.Jorgensen, eds, Kluwer Academic, pp. 167–178, 2001.

[Seidenfeld, 1993] T. Seidenfeld and L. Wasserman, Dilation for Sets of Probabilities", *Annals of Statistics*, 21, pp. 1139-1154, 1993.