Synopsis

Bayesianism, Fundamentally

Center for Philosophy of Science
University of Pittsburgh
Friday October 13-Saturday October 14, 2006

"Rational Belief and Reasonable Belief, a Ramseyian Distinction,"
James Joyce, Department of Philosophy, University of Michigan
(Annual Lecture Series, Center for Philosophy of Science)

"Objective Bayes: Old and New Questions, Some Answers"
Jayanta K. Ghosh, Department of Statistics, Purdue University

"Probability is like a language used to express personal degrees of uncertainty,"
Jay Kadane, Department of Statistics, Carnegie Mellon University

"A price too great for unbounded utilities?"
Teddy Seidenfeld, Department of Statistics and Philosophy, Carnegie Mellon University

"Comparative Probability, Comparative Confirmation, and the 'Conjunction Fallacy'"
Branden Fitelson, Department of Philosophy, University of California at Berkeley

"Ignorance and Indifference"
John D. Norton, Department of History & Philosophy of Science
and Center for Philosophy of Science, University of Pittsburgh

A Brief Report on the Conference: *Bayesianism, Fundamentally*

Jonah N. Schupbach
Department of History and Philosophy of Science
University of Pittsburgh

James Joyce – "Rational Belief and Reasonable Belief"
Dr. Joyce intended this talk to be both an introduction to work that has been done on the problem of the priors and a summary of recent work that Dr. Joyce has been pursuing on this problem. A good deal of time was spent talking about the attempts of objective Bayesians to single out certain subsets of credal states as evidentially appropriate by appealing to symmetries (e.g., Jeffreys priors or Haar measures). Dr. Joyce admitted that such appeals to symmetry considerations do lead to prior assignments that are not susceptible to Venn's problem (i.e., the standard objection that prior assignments depend on how we describe our possibilities); however, Dr. Joyce proceeded to raise a different objection to such methods. The "crazy Henri" example was intended to show that the feasibility of such approaches depends upon substantive empirical assumptions. Insofar as objective Bayesians are seeking an objective, logical method to assigning priors, this is a problem.

Dr. Joyce later presented his ideas on methods for evaluating priors on the basis of epistemically desirable characteristics. He focused on evaluating the *accuracy* of credal states. A believer's degree of belief is that subject's estimate of its truth value. A scoring rule measures the inaccuracy of a subject's belief. Dr. Joyce proposed that there is compelling epistemological motivation for requiring a truth-directed, strictly proper, and convex scoring rule.

Jayanta K. Ghosh - "Objective Priors: Old and New Questions, Some Answers"
Dr. Ghosh discussed various statistical methods for deriving objective priors including Jeffreys Priors and the L-B entropy measure. His talk also touched on various common criticisms of these methods, including invariance, uniqueness, and model dependence objections. Dr. Ghosh argued that the standard objections to objective priors are met by current methods for deriving objective priors. However, Dr. Ghosh also wanted to clarify that one would only want to use objective priors in certain contexts.

Jay Kadane - "Probability is Like a Language Used to Express Personal Degrees of Uncertainty"
To begin this talk, Dr. Kadane reflected on the question why we use these axioms rather than others to model our degrees of belief. His answer was that these are the axioms that ensure that we avoid sure loss. Dr. Kadane proceeded to argue that these axioms don't do any more than that for us though. Ridiculous beliefs – e.g., the moon is made out of green cheese – can be expressed in the probability calculus. Doing so ensures that you will not be susceptible to a Dutch book, but of course, it does not turn your ridiculous belief into a sensible belief. The claim that one's prior is objective is strong and misleads people into thinking that there is no need to defend that assignment; however, there is no basis for such a claim. Such claims (to objectivity, consensus, etc) amount to a lazy surrender to the need to explain yourself. All we really have in such cases is informed opinion and persuasion. Dr. Kadane argued for this take on objective Bayesianism from the non-uniqueness of probability assignments derivable from the host of available objective analyses

In the question and answer time, the question was asked whether we couldn't have a coherent prior distribution that we find, after a series of failures, to be simply wrong or inaccurate. In response, Dr. Kadane stated that in such cases, a subject sees certain consequences of his assignments that he was previously unaware of and decides that those assignments didn't actually reflect his beliefs. The problem becomes one of internal rather than external accuracy. In response to another question, Dr. Kadane expressed his belief that, given what he had said in his talk, it would be appropriate to say that induction holds no objective power between people. For example, when we make inductions in physics, we are really talking about our opinions.

Teddy Seidenfeld, "A Price Too Great for Unbounded Utilities?"
Dr. Seidenfeld was the bearer of bad news at this conference. He began by allowing that there may be contexts for which we would want to allow for unbounded utilities. However, if we allow for unbounded utilities, he showed that coherence mandates a strict preference among some equivalent random variables. Thus, there can be no function of probability and utility that will preserve the equality of equal random variables.

Branden Fitelson, "Comparative Probability, Comparative Confirmation, and the 'Conjunction Fallacy' "
Dr. Fitelson began by discussing a historical interaction between Hempel, Carnap, and Popper. Popper pointed out that Carnap was confusing two different qualitative ideas of confirmation and in fact that he was holding to each one at various points in his work. The two notions that Popper was confusing were "firmness" - according to which E confirms H to degree $P(H|E)$ – and "increase in firmness" - according to which E confirms H to degree measured by some $f[P(H|E), P(H)]$.

Dr. Fitelson pointed out that contemporary Bayesians hold a formally similar notion of confirmation to Carnap's "increase in firmness" notion, and $f[P(H|E), P(H)]$ is spelled out in various ways in the Bayesian literature according to a variety of comparatively distinct measures.

Dr. Fitelson proceeded to apply the distinction between firmness and increase in firmness to Tversky and Kahneman's Linda example. The results of the tests done with this example can be explained to some degree if the test-takers are drawing from intuitions about increase in firmness as opposed to firmness (essentially if they are having the same confusion among terms that Popper was shown to have). Although the test-takers are still making probabilistic mistakes in Dr. Fitelson's opinion (as can be seen by the fact that the test-takers still make the same decisions generally when the tests are run with betting odds rather than probabilities), increase in firmness constitutes a concept that is in the neighborhood of firmness via which the test-takers would be thinking accurately. That is, in the Linda example, it *is* the case that $C(H1\&H2,E) > C(H1,E)$ – where C is some measure of the degree to which E confirms H in the "increase in firmness" sense – so long as (i)$C(H2,E|H1) > 0$ and (ii) $C(H1,E) <= 0$. Dr. Fitelson argued that both (i) and (ii) are clearly the case in the Linda example. (i) and (ii) can be logically weakened to (i) $Pr(E|H1\&{\sim}H2) < Pr(E|H1\&H2)$ and (ii*) $Pr(E|H1\&{\sim}H2) <= Pr(E|{\sim}H1)$.

Lastly, Dr. Fitelson showed that there exist non-traditional conjunction fallacy examples for which (i) seems plausible but (ii) / (ii*) do not; thus, there is more work to be done in developing more general confirmation-theoretic models to account for all known conjunction fallacies.

John Norton, "Ignorance and Indifference"

Dr. Norton began by stating his belief that it is a platitude that beliefs are grounded in reason; if there is no difference in reason between beliefs, then there ought not be a difference in belief at all. Hence, the principle of indifference is accurate and it couldn't be wrong given this platitude. Dr. Norton proceeded to review some of the challenges that have been put forward to the principle of indifference and to symmetry or invariance conditions. These problems don't reflect any failings of such principles; rather, they reflect the fact that we ought not always model ignorance with probabilities. His main point was that if we drop the assumption that ignorance states must always also be probability distributions, then any principle of invariance of ignorance will pick out an ignorance state that is incompatible with axioms of probability theory. Dr. Norton clarified that in his view, probabilistic methods represent a fine way of doing induction in a broad range of contexts; however, there are also contexts in which it is problematic to attempt to fit states of ignorance into probability distributions. In concluding, Dr. Norton offered an example (the "dome") of an indeterministic system, beliefs about which could not be represented by any probability distribution.

Is the Probability Calculus the Universal Logic of Induction? A Dissenting View
John D. Norton

One of the persistent themes of the workshop, perhaps its strongest theme, has been the problem of deciding how to handle prior probabilities. The standard view seems to be that these problems can be accommodated one way or another within the framework of Bayesian thought, although we do not have agreement on what that accommodation should be. I've found myself taking a very different view of just what those problems signify

The basis of this view is a deep sense that we have compromised our original, possibly naive aspirations. Was there not a time at which we each hoped that Bayesianism might realize a simple inductive ideal? It would give us a formal tool for representing the transition from ignorance to knowledge. In the beginning we have a truly vacuous belief distribution. Then, as new evidence is learned, we update our belief distribution, which becomes richer and better informed.

We all know that hopes for such a dynamics in a Bayesian system were dashed by many problems. Perhaps the most prominent is the impossibility of forming a probability distribution that can properly be described as vacuous. So we needed to retract.

A lot of the discussion at the workshop was devoted to determining how that retraction should happen. The objectivist approach, which I personally find most appealing, seeks conditions that pick out unique prior probability distributions. That these methods fail in general is a widely known result. So objectivists make the major concession that these methods can only work if we are already in a context in which we have sufficient knowledge already to fix an objective prior. That lets slip away the original idea of a dynamics of belief that brought us from ignorance to knowledge. Bayesian analysis is restricted to isolated islands of analysis in which a unique prior can reasonably be selected.

The subjectivist response is to give up the idea that there is one correct prior. Instead everyone is allowed to pick whatever prior they like. These priors are still rich in differing content but those differences are demoted to opinion, with the expectation that those opinions will be forced to merge into knowledge under the weight of accumulating evidence, eventually.

What disturbs me in this approach is that we are giving up the idea that evidence can have a definite import. For now our probability distributions are mixtures of opinion and warranted belief. We can no longer say what the total burden of our evidence is. We can at best say what it is *for you* or *for you*. The original idea that we would have a logic of induction is drifting towards the idea that what we really have is some kind of psychology of induction. By that I mean that the dynamics of belief is no longer to be dictated purely by logical principles; those dynamics have become a mix of principle and opinion. The latter cannot be right or wrong but at best are to be elicited from each individual subject by some operational technique, rather like an experiment in psychology.

It has taken us a long time to get to this situation. My proposal is that very early on we took a wrong turn when we decided that the probability calculus was to be preserved as the correct logic of induction, even when we were in domains in which we know very little. If we give up this idea, I hold out the hope of restoring the objectivists' goal of there being a single, correct distribution of belief in every evidential situation.

To illustrate this hope, my paper was devoted to the extreme, idealized case in which we are supposed to be in complete ignorance. I showed that the standard techniques of the objectivists will now pick out a unique, ignorance belief distribution, although it is not probabilistic. Of the methods I discussed, the most effective makes use of the idea that ignorance with regard to contingent propositions is invariant under negation.

Let me mention two alternatives that I did not discuss in my presentation.

First, a very popular technique for representing ignorance in Bayesian analysis is to use sets of probability measures. My real concern with this technique is that it is using an additive measure to simulate belief distributions that are inherently non-probabilistic. So, while one is adhering to the letter of Bayesian lore in using probability measures, at the surface level, the effective belief distributions themselves are no longer additive measures. They have a different logic on the surface. My project is to discover what can be said of that surface logic. There is also a technical problem with the use of these sets to represent complete ignorance. If we require any representation of ignorance to be invariant under negation, these sets fail the requirement. For under a negation map, additive measures do not remain additive measures but turn into measures with a different algebra. Elsewhere I have called them "dual additive measures."

Second, there has been an extensive literature whose goal is to find a good measure of incremental support within the context of a Bayesian analysis. For my purposes, what is important is that all these measures are defined on a background probability distribution. The presumption that such a probability distribution exists and that all our inductive quantities must be defined in terms of it places a powerful restriction on our analysis. (Or if it doesn't, why have it?) My project is to drop this presumption and see what can be said of inductive inference.

Remarks
Jay Kadane

I welcome this opportunity to compare perspectives with John Norton. Taking the risk of being too pessimistic, I state my views in stark terms.

1. What is ignorance ? While of course I have no objection to people thinking about whatever they choose, I do not see this question as a fruitful one. I am quite willing to rephrase the overall issue as one of having a principled way of going from one state of partial knowledge to another as a result of the acquisition of data, rather than worrying about a privileged starting place.

2. Is there hope for objectivism ? The concentration on the status of prior distributions (is there a basis for an objectivist account) leads the discussion away from the equally problematic status of likelihood functions for real data. Just as in the case of priors, there is not (and, I think, cannot be) a satisfactory objectivist account of likelihoods. How can you prove that my favorite statistical model for some real data is objectively wrong and yours is right ?

My conclusion from these questions is that the only position I can defend is the subjective one. It at least offers an alternative to the useless fiction that claims of objectivity sometimes support.

Contributions of Jayanta Ghosh

Part 1. My presentation on Objective priors (for low dimensional parameters) went along the following lines.

My position is that of a Bayesian statistician who became a Bayesian because of the paradoxes in classical statistics. My goal is to summarise the data through a posterior. In particular, I always assume we have an experiment and data to start with. The posterior is more important to me than the prior from this point of view.

I started with the assumption that if there are a finite number of states of Nature, then uniform prior on these states is the unique objective prior. At least a partial justification is that it maximizes Shannon entropy.

Given this assumption I presented two justifications of the Jeffreys prior. The first justification is as a weak limit of discrete uniforms spread on points that approximate compact subsets of the parameter space up to small Hellinger distance ε with $\varepsilon \downarrow 0$. This is a theorem of Ghosal, Ghosh and Ramamoorthi. The second justification is due to Bernardo. Bernardo shows the Jeffreys prior maximizes the Lindley-Bernardo divergence between the prior and posterior.

More details are available from Ghosh and Ramamoorthi (Bayesian Nonparametrics, Ch 1) and Ghosh, Delampady and Samanta (Bayesian Analysis- Theory and Methods, Ch 5). The two books were published in 2003 and 2006 (publisher- Springer).

I then replied to the following three common criticisms.

A. Possible lack of invariance of an objective prior under smooth 1-1 transformations on the parameter space.
B. Dependence of an objective prior on the experiment.
C. Lack of uniqueness.

Briefly, my answer was as follows. The Jeffreys prior is invariant under 1-1 differentiable transformations.

The dependence on the experiment comes from the dependence of the Lindley-Bernardo measure of divergence. Its philosophical or information theoretic basis may be the fact that it is impossible to define the information in a prior except in the context of an experiment. In this connection, see also Part 3 of this write up.

Finally, even if objective priors, Jeffreys and other objective priors, are not unique, it seems from examples that the corresponding posteriors are very similar.

Part 2. Discussion of comments on my presentation

Jay Kadane asked why should these priors be called objective. This may affect negatively others with different priors.

John Norton felt choice of even the uniform discrete isn't compelling unless there is enough invariance to lead to the same prior for all possible re-labeling of different states of Nature.

Someone (was it James Joyce ?) asked why the uniform discrete is a good representation of ignorance. A uniform discrete also seems like a representation of some knowledge or belief.

I myself raised the question of lack of any connection with the Jeffreys prior and Robust Bayesian Analysis, which I find puzzling.

Some tentative thoughts on the above issues if offered below.

The reason for calling the Jeffreys prior is partly historical and partly the fact that, as often explained by Bernardo and my discussion of the third criticism in Part 1, the posterior seems to be determined mostly by the data. I can't call it subjective since it arises from an algorithm unconnected with any subjective belief or knowledge. Would "a non-subjective" prior seem better ?

I appreciate John's point but if one assumes no knowledge or context, surely permutation of label names would not matter and the only invariant prior would be the discrete uniform. Perhaps I'm missing something.

I find the third point very interesting but seem to have seen it before. If all priors are possible and equally plausible one might want to put a prior on priors and integrate out. I guess we would work out such that the final result would again be a discrete uniform. But this could also lead to an infinite regress.

Finally, about my own point on lack of connection with Bayesian robustness, I am toying with the idea of trying to connect with Herman Rubin's old idea of achieving robustness by minimizing the prior Bayes risk approximately so that the same minimizer works for many similar priors. Herman does not believe in posteriors.

Part 3. Miscellaneous

While thinking about information (or lack of it), I realize information may have many meanings, just as credence may have many meanings.

I also note that invariance should apply to measures of information. The Lindley-Bernardo measure is invariant under all 1-1 differentiable transformations on the parameter space. The same is true if we construct a similar measure based on the Hellinger distance. The invariance fails for the variation distance between prior and posterior. This may explain why some priors maximizing the variation norm are unintuitive candidates for an objective prior. Also, the usual practice of judging the informativeness of a prior by its peakedness becomes suspect since peakedness isn't invariant.

In my talk I mentioned reference priors but didn't discuss them. A monograph by Bernardo, Berger and Sun is expected to be published soon.

The October meeting at Pittsburgh was great fun and for me, an excellent opportunity to learn not only about objective and subjective Bayes but new resolutions of old cognitive puzzles (a la Branden).

Jayanta

Comments by Teddy Seidenfeld

I thank the organizers and fellow participants for a most enjoyable workshop. These few comments are offered with the intent of continuing the conversation about Objective Bayesian methods initiated by Professor Ghosh's stimulating presentation.

Professor Ghosh gives us a Tempered Objective Bayesian Theory. His justifications of so-called "objective" methods are judiciously situated in constrained contexts of inference. Nonetheless, I am skeptical. My concern is that even in such well defined settings there remain, nonetheless, conflicts between the "objective" methods and core Bayesian principles. Here, I suggest two lines of conflict, (1) and (2). I conclude with a reminder (3) that there is an additional price to pay for the using reference priors.

(1) Where the prior for the parameter of interest in Objective Bayesian inference depends upon the sample space, then that approach leaves us without coherent guidance about how to choose the experiment that determines the sample space. Suppose, for instance, that we design our experiments by comparing expected informational gain about the parameter of interest and choose the experiment that maximizes this expected gain. (The expected informational gain is judged, pre-data, by calculating the expected difference in information between the "objective" prior and resulting posterior for the parameter of interest.) But since the "objective" prior changes with the experiment being evaluated, this decision rule is not admissible in Wald's sense. It is incoherent in de Finetti's sense.

(2) Where the prior for Objective Bayesian inference depends upon the sample space, then there are conflicting patterns of inference when the observed data are composite. Here is a schematic version of the objection when the composite data $D = (d_1, d_2)$ arise from two different experiments, respectively, E_1 and E_2. Also, denote by E_3 the composite experiment with components E_1 and E_2. Suppose that with data from E_i, the Objective Prior for the parameter of interest, θ, is P_i ($i = 1, 2, 3$). There are then three rival Bayesian inference patterns for computing an Objective posterior for θ, $P_i(\theta \mid D)$:

$$P_1(\theta \mid D) \propto P_1(\theta)\, P(d_1 \mid \theta)\, P(d_2 \mid \theta, d_1)$$
$$P_2(\theta \mid D) \propto P_2(\theta)\, P(d_2 \mid \theta)\, P(d_1 \mid \theta, d_2)$$
$$P_3(\theta \mid D) \propto P_3(\theta)\, P(d_1, d_2 \mid \theta)$$

Because, generally, $P_1 \neq P_2 \neq P_3 \neq P_1$, then $P_1(\theta \mid D) \neq P_2(\theta \mid D) \neq P_3(\theta \mid D) \neq P_1(\theta \mid D)$.
What, then, is "objective" about these rival inferences?

(3) Where the objective prior is "improper," which is not uncommon with reference priors if the parameter space is unbounded, the resulting inferences are valid for finitely but not countably additive probability theory. For example, with Lebesgue measure as the reference prior for a real-valued parameter θ, e.g., with a flat "improper" prior for the mean of Normally distributed data – "improper" because the parameter space is given infinite measure – the resulting inferences correspond to a purely finitely additive prior probability for θ. Under Lebesgue measure, each unit interval for θ is given equal prior weight. But the parameter space, the real line, admits a countable partition by unit intervals. Hence, each of these unit intervals gets 0 prior probability, etc. The Bayesian theory that results from using a purely finitely additive prior probability is qualitatively different from the more familiar countably additive version. In Kadane et al (1986) we discuss a variety of these differences.

Additional Reference

Kadane, J.B., Schervish, M.J., and Seidenfeld (1986), "Statistical Implications Finitely Additive Probability," in *Bayesian Inference and Decision Techniques*, P.K. Goel and A. Zellner (eds).

Amsterdam: Elsevier Science, 59-76.