Direct Inference and Inverse Inference

Teddy Seidenfeld

The Journal of Philosophy, Vol. 75, No. 12. (Dec., 1978), pp. 709-730.

Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197812%2975%3A12%3C709%3ADIAII%3E2.0.CO%3B2-R

The Journal of Philosophy is currently published by Journal of Philosophy, Inc..

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/jphil.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Tue Mar 4 10:39:15 2008

THE JOURNAL OF PHILOSOPHY

VOLUME LXXV, NO. 12, DECEMBER 1978

DIRECT INFERENCE AND INVERSE INFERENCE *

THERE is no surer test of the character of an inductive logic than problems of inverse inference: inference from "sample" to "population." Bayesian programs, like Carnap's, need no special principles (either substantive or methodological) for solving inverse inference.[1] After an initial (prior) probability is fixed for some hypothesis (and bayesian programs differ on whether the logic stipulates this prior), then confirmational conditionalization determines the subsequent (posterior) probability for that hypothesis, assuming new evidence is accepted. Moreover, this machinery works whether the hypothesis is general, as in inverse inference, or particular, as in predictive inference.

In opposition to bayesian theory, frequentists usually object to the principle of confirmational conditionalization not because its formal basis (Bayes' theorem) is less than valid, but because the presupposition of a precise, prior credal probability for a hypothesis (given an arbitrary knowledge base) is thought to be indefensible.[2]

* I wish to thank Henry Kyburg for his unbounded patience with me so that I might better understand his position. He has helped me correct several misunderstandings I had about epistemological probability. Isaac Levi has read earlier drafts of this paper and has suggested important revisions, which I have acted on. Also, I thank Carl Posy for his assistance in clarifying some obscure passages.

[1] Throughout this paper I will call an inductive logic *bayesian* when (i) an agent's beliefs, at a time, are represented by a single coherent probability function, (ii) an agent's confirmational commitments, at a time, to changes in belief that follow a growth in knowledge are governed by Bayes' theorem (this is called confirmational conditionalization), and (iii) total evidence is respected.

[2] I call an inductive logic *frequentist* when it seeks to by-pass confirmational conditionalization in inverse inference and, instead, relies upon (what Ian Hacking calls) a "frequency principle." That is, as described in the text, frequentists solve inverse inference by reduction to direct inference, and problems of direct inference (the "single case") are solved with a frequency principle.

The name 'frequentist' is *not* to suggest a frequency interpretation of probability. It is a separate question whether or not some nonconfirmational probability

0022-362X/78/7512/0709$02.20 © 1978 The Journal of Philosophy, Inc.

Especially in problems of inverse inference, where the hypothesis is general and the initial knowledge base is often weak, frequentists typically reject the bayesian assumption that there exists a precise prior probability for the hypothesis. Hence, frequentists are wont to deny a bayesian solution to inverse inference because confirmational conditionalization assumes too much.

Whereas there is discord between the two schools on matters of inverse inference, there is relative harmony about simple cases of direct inference: inference from population to sample. For example, given merely the information that an urn contains equal numbers of red and black chips and that a chip is to be drawn by some suitable (random) process, there is unanimity that the probability is .5 that the chip to be drawn is black. This agreement is significant because, if one characteristic of a frequentist is opposition to bayesian conditionalization in inverse inference, a second is support for the constructive strategy of solving inverse inference problems by reducing them to cases of direct inference.[3]

In this paper I attempt a refutation of frequentism by arguing that the program for reducing inverse to direct inference is not viable. Moreover, I claim my arguments also show that varieties of bayesian theory which try to resuscitate Laplacean Insufficient Reason (so that precise "ignorance" prior probability might be identified) fail in the same way the frequentist idea does.[4] In section I I consider Henry Kyburg's original frequentist theory, *epistemological probability*, and argue against its solution to inverse inference. (I am forced to treat his position separately because of the unusual stand that Kyburg takes in rejecting conditionalization.) In section II I consider Ian Hacking's novel reconstruction of Fisher's fiducial probability and argue that the inconsistencies present are representative of the generic weakness in the frequentist solution to inverse inference. I limit the discussion to problems of statistical inference, i.e., where populations and samples are of interest for their statistical properties only. Besides having practical significance, problems of statistical inference highlight the philosophical issues at stake here.

---

is even recognized by a frequentist. For example, both Hacking's fiducial inference and Kyburg's epistemological probability are frequentist positions; yet Hacking acknowledges chances but Kyburg rejects any such notion.

[3] For example, advocates of Neyman's confidence intervals, or Fisher's fiducial probability, or Fraser's structural inference, or Kyburg's epistemological probability, all subscribe to this policy. In this paper I will not address confidence-interval theory. I believe adequate criticism of this program is already common in the literature.

[4] I have in mind particularly Sir Harold Jeffreys' inductive logic, especially his theory of *invariants*. A recent defense of Jeffreys' position is found in Roger Rosenkrantz, *Inference, Method and Decision* (Boston: Reidel, 1977).

I

Within the last year readers of this JOURNAL have witnessed aspects of an ongoing debate between Isaac Levi and Henry Kyburg over fundamental issues in statistical inference. In "Direct Inference,"[5] Levi reports to us an incompatibility he uncovers between Kyburg's original theory, epistemological probability, and a conjunction of familiar inductive principles: principles all of which Levi finds compelling. (Levi attacks with a counterexample focused at Kyburg's analysis of direct inference; hence his title.) Kyburg, who has developed epistemological probability over a twenty-year span, is not caught napping. In his recent book, *The Logical Foundations of Statistical Inference*, he devotes a chapter (chapter 11) to an evaluation of one of the inductive principles essential to Levi's counter argument, namely, confirmational conditionalization. On the basis of several numerical examples worked out in epistemological probability theory, Kyburg concludes that conditionalization is worse than suspect.

In "Randomness and the Right Reference Class,"[6] Kyburg reiterates his rejection of conditionalization while challenging Levi to defend his own version of direct inference, a version obeying the controversial principle. Kyburg charges that, whereas all violations of conditionalization imposed by epistemological probability are either banal or otherwise intuitively satisfying (Levi's counterexample being of the first kind), it is Levi's position that is inadequate despite its adherence to conditionalization. Specifically, Kyburg sees Levi's program as overburdened with a bevy of unwarranted empirical assumptions, all needed to ensure conditionalization in even the simplest cases of direct inference. Thus, a central theme in their debate is the clash of arguments for and against the principle of conditionalization.

It is not my intent in this article to critically review Kyburg's evaluation of Levi's approach to direct inference.[7] I do, however, strongly disagree with Kyburg on his assessment of the status of epistemological probability theory. Specifically, I find serious difficulties in the epistemological-probability solution to inverse-inference problems. Kyburg's theory is unusual in part in that all probability statements have the *direct* form. Since, as we expect of a frequentist inductive logic, epistemological probability solves inverse inference by reduction to direct inference, it is only natural to wonder whether the violation of conditionalization (identified by Levi) affects inverse inference too. In what follows in this section I

[5] This JOURNAL, LXXIV, 1 (January 1977): 5–29.
[6] This JOURNAL, LXXIV, 9 (September 1977): 501–521.
[7] See Levi's rebuttal, this JOURNAL, this issue, 730–737.

shall demonstrate that the sufficiency principle (a logical *consequence* of conditionalization) is violated, with the result that many basic inverse-inference problems are left unsolved. Also, once we understand the severity of Kyburg's rejection of conditionalization, we can identify three garden-variety statistical inferences (each related to inverse inference) that cannot be reproduced in epistemological probability theory.

My concern with Kyburg's theory grows from a respect for and an appreciation of what we can learn from him about the limits of the frequentist program. Where epistemological theory encounters difficulties in inverse inference there is good reason to worry whether other frequentist approaches commit subtle, but serious errors in their haste to solve these problems. (We shall return to answer this question in section II.)

To understand better the controversy surrounding conditionalization, let us review the role played by this principle within a familiar logic of induction, to wit: a Carnapian framework. One of the secure principles in the Kyburg/Levi debate, i.e., one accepted with only minor revisions, is *coherence*. In a Carnapian system, coherence postulates the existence of a single probability function $Q_K(\ )$, defined over sentences or propositions, which represents an agent's degree of belief; where K (a consistent, deductively closed set of sentences) corresponds to his corpus of knowledge. Now, *conditionalization* fixes the commitments (the agent bears), arising from the epistemic state represented by $Q_K(\ )$, to other epistemic states that result merely by enlarging K through acceptance of new evidence. That is, conditionalization identifies $Q_K(\ /d)$ as the extension of the credal state $Q_K(\ )$, when K is augmented by datum d (and all its consequences). The multiplication axiom then yields the important result that, if $Q_K(d) \neq 0$ and h is an hypothesis,

$$Q_K(h/d) = Q_K(d/h) \cdot Q_K(h) \div Q_K(d) \quad \text{(Bayes' theorem)}$$

Calling K' the new corpus of knowledge that follows upon accepting evidence d [and assuming, as always, that $Q_K(d) \neq 0$], we have the familiar bayesian formulation of conditionalization:

$$\textit{Conditionalization: } Q_{K'}(h) = Q_K(d/h) \cdot Q_K(h) \div Q_K(d)$$

Returning to Kyburg's theory, we see that reformulation of these principles is no simple matter. Epistemological probability, unlike the Carnapian alternative, is not a real-valued but an *interval*-valued function. $EP_K(h) = [p_L, p_U]$. The epistemological probability of h, given knowledge base K,[8] is the interval $[p_L, p_U]$,

[8] The knowledge base K need not be deductively closed, or even strictly consistent, in epistemological theory. However, for the arguments considered here nothing rests on this fact.

$0 \leq p_L \leq p_U \leq 1$. Thus, Kyburg accommodates the usual frequentist objection to conditionalization by generalizing the representation of a rational belief state to include intervals of probability. His plausible idea is that, in cases where the knowledge base is empirically too impoverished to support precise probability values, intervals of probabilities are adequate.

The analogue to *coherence* in Kyburg's theory is his result that, for any finite set of sentences, there is some coherent probability function $Q_K(\ )$ that agrees with all the epistemological probability intervals $EP_K(\ )$ over the set.[9] That is, there is at least one $Q_K(\ )$ whose values fall within the corresponding intervals $EP_K(\ )$ for sentences in the set. But the difficulty in reformulating *conditionalization* is that the very simple try:

$$EP_{K'}(h) = EP_K(d/h) \cdot EP_K(h) \div EP_K(d)$$

will not work, since the product and division of intervals is not meaningful.

In "On Indeterminate Probabilities,"[10] Levi develops an inductive logic that both uses (convex) sets of coherent probability functions to represent degrees of belief and satisfies a natural extension of *conditionalization*. However, Levi's important criticism of epistemological theory (in "Direct Inference") does *not* presume this generalization of conditionalization; for, quite cleverly, Levi chooses a counterexample in which all epistemological probabilities are precise, point-valued ($[p, p]$), even with changes in evidence. Were Kyburg's objections to conditionalization limited to the familiar frequentist plea: that conditionalization is to be restricted to problems where well-defined, i.e., precise, probability exists, then he would be obligated to satisfy conditionalization in Levi's example. What we learn from Levi's argument is that Kyburg's reservations on conditionalization are more extensive than is usual for a frequentist.[11]

[9] See MT-12.1 of his *Logical Foundations of Statistical Inference* (Boston: Reidel, 1974).

[10] This JOURNAL, LXXI, 13 (July 18, 1974): 391–418. Although Levi's program utilizes both confirmational conditionalization and a frequency principle, it is not bayesian (as that title is used here) since beliefs are not represented by a unique probability function, nor is it frequentist since Levi does not try to solve inverse inference by reduction to direct inference. Hence, his position is an interesting alternative to those discussed in the text.

[11] I summarize Levi's counterexample as follows: There exists a corpus of knowledge K such that, for a hypothesis h, $EP_K(h) = [.9, .9] = .9$. Also, with three possible evidential reports d_i ($i = 1, 2, 3$) (d_1 or d_2 or d_3) ϵK, $EP_{K_1}(h) = .85$, $EP_{K_2}(h) = .91$, and $EP_{K_3}(h) = .95$; where K_i is the result of accepting d_i into K. Finally, it is the case that $EP_{K_{1\text{or}2}}(h) = EP_{K_{1\text{or}3}}(h) = .9$; where $K_{i\text{ or }j}$ is the result of accepting into K the weaker evidence d_i or d_j. But an elementary calculation shows that no coherent probability function can satisfy these six equations if conditionalization holds, even though all probabilities are precise.

As I reported at the beginning of this section, Kyburg reacts to Levi's charge by claiming that whenever conditionalization fails in epistemological theory it is either because the failure is banal—i.e., What do we lose in Levi's example?[12]—or else the failure is satisfactory, as when we ignore conditionalization because, true to frequentism, our knowledge cannot be accurately represented by a precise, coherent probability function—i.e., conditionalization does not apply. I maintain that Kyburg is wrong and that we can clearly recognize the severity of the loss of conditionalization in terms of the collapse of his program for inverse inference.

Let me begin, then, with a brief rehearsal of Kyburg's ingenious scheme for reducing inverse to direct inference. Epistemological probability statements all have a direct-probability form: the EP_K is $[p_L, p_U]$ that entity a bears property F, since a is known to be an R and it is known only that between p_L and p_U of all Rs are Fs. Most of the important philosophical issues in inductive inference are treated through Kyburg's regulations governing *randomness*. Randomness conditions fix the reference class R and frequencies p_L, p_U of the epistemological probability statement. That is, $EP_K(Fa) = [p_L, p_U]$, because randomness conditions yield the claim: a is a *random* member of the (reference) set of Rs, with respect to property F, and by appeal to the known frequency bounds $[p_L, p_U]$ for Fs among Rs. Thus, epistemological randomness solves the vexing problem of direct inference: how to choose the reference set. The trade-off that must be evaluated pits size of the reference set (the narrower the better) against width of the frequency bounds (the narrower the better), for with decreasing reference-set size comes (usually) a broadening of the frequency interval. For example, in the extreme case of a unit reference set $\{a\}$, the frequency of Fs may be known to be bounded only by the trivial interval $[0, 1]$.

The trick in converting inverse to direct inference rests on the use of a special property, called *rational representativeness*. For instance, suppose the inverse inference is from an observed sample of n flips with a newly bent coin to hypotheses about the bias of the coin. We assume, initially, no nontrivial frequency information about the bias of the coin, but we agree that the flips are (1) statistically independent, and (2) modeled by a binomial distribution with the binomial parameter θ equal to the bias for landing heads up. Then

[12] In "Subjective Probability: Criticisms, Reflections, and Problems," *Journal of Philosophical Logic*, VII, 2 (April 1978): 157–180, Kyburg argues against using betting arguments to show the penalties to be imposed on violators of coherence and conditionalization.

a sample of n flips is *rationally representative at the* .1 *level* (RR$_{.1}$) just in case the difference between the observed frequency of heads and the bias for heads is not more than .1. If the coin is flipped 100 times and 67 trials land heads up, the sample is RR$_{.1}$ if and only if $|\theta - .67| \leq .1$. If the coin is "fair" ($\theta = .5$) then the sample is not RR$_{.1}$.

Although the RR property is not observable (since we are ignorant about the bias of the coin), it is very useful for inference about the bias, since (using a normal distribution approximation for the binomial frequencies) between 95% and 100% of all samples of 100-fold independent trials from a binomial process are RR$_{.1}$. Furthermore, because we are relatively ignorant about the bias of this coin, upon learning that 67 of the 100 flips landed heads up, randomness conditions yield the claim that our observed sample is a *random* member of the (reference) set of all such 100-fold trials, with respect to the RR$_{.1}$ property and subject to the known frequency bounds [.95, 1]. Hence, EP$_K$(RR$_{.1}$ (our sample)) = [.95, 1]. In words, the epistemological probability equals the interval [.95, 1] that $|\theta - .67| \leq .1$, or, equivalently, the epistemological probability is no less than .95 that $.57 \leq \theta \leq .77$. This last statement is a nontrivial solution to the inverse-inference problem, obtained by reduction to direct inference through the RR predicate.[13]

Obviously, the critical steps in this argument involve the randomness conditions; in particular, why is it that ours is a *random* sample with respect to the RR property and frequencies bounded by .95 and 1.0? The answer lies in the character of the randomness regulations which support claims of irrelevance *unless* relevance can be demonstrated. For example, if we concede that the next flip of a fair coin is random with respect to landing heads up and (precise) frequency .5, then *unless* we know enough about flips of fair coins that stay in the air for more than two seconds to defeat the randomness claim just made, the next flip *will* be a *random* member of the set of flips with fair coins that stay up for more than two seconds, with respect to landing heads up and (precise) frequency .5. That is, Kyburg's randomness conditions treat evidence as irrelevant unless relevance is demonstrable.

Before observing the sample of 100 flips, a simple direct inference leads to the conclusion that it is *random* with respect to RR$_{.1}$ and frequencies bounded by .95 and 1.0. *After* observing the flips and noting that 67 fell heads up, unless this extra information can be

[13] This argument was first presented in Kyburg, *Probability and the Logic of Rational Belief* (Middletown: Wesleyan UP, 1961).

shown relevant to the initial randomness claim (which cannot happen if we are ignorant about θ), the sample is labeled *random* as in the argument above.[14] It is this attitude toward relevance that permits Levi's counterexample. But, as we see next, this treatment of relevance also leads to a rejection of the sufficiency principle (a consequence of conditionalization) with dire consequences for Kyburg's solution to inverse inference.

A question of both practical and theoretical importance for problems of inverse inference is how to recognize summaries of large bodies of data that preserve all the relevant information contained in the evidence as a whole. A powerful and widely accepted answer is given by the sufficiency principle.[15]

Definition: With respect to inverse inference about an unknown θ, a simplification (condensation) of data d to t, a function of d alone, i.e., $f(d) = t$, is *sufficient* just in case:

$$Q_K(d/t \,\&\, H_\theta) = Q_K(d/t)$$

where H_θ is a hypothesis (of inverse inference) about θ. That is, t is sufficient for d whenever the conditional probability of d, given t, is independent of the unknown θ.

Sufficiency principle: If and only if t is sufficient for d, with respect to θ, then inference from t (alone) preserves all the relevant evidence contained in d concerning θ.

For instance, in the problem discussed above, the epistemological-probability solution to inverse inference about the bias of a newly bent coin depends on knowing the frequency of flips landing heads up and the total number of trials. Kyburg's reduction to direct inference using RR rests upon the irrelevance (subject to randomness) of other aspects of the total evidence, such as the order of trials landing heads up among those landing tails up, the outcome of the first trial, the proportion of heads to tails among the odd-numbered flips, etc. Now it is not difficult to show that, for the binomial distribution with statistically independent trials, a sufficient set of statistics for sample data (with respect to inference about the binomial parameter) is the pair: number of trials and frequency of outcome.[16] Thus, Kyburg's ingenious use of the property of being rationally

[14] I thank Professor Levi for suggesting this distinction to me and for pointing out its significance in separating Kyburg's from Hacking's solutions to inverse inference.

[15] See Appendix 1 for a derivation of the sufficiency principle from conditionalization.

[16] This and other purely statistical claims made in the text can be referred to any basic statistics book, e.g., D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint* (New York: Cambridge, 1965).

representative does not violate the total-evidence requirement (for *this* problem), if we assume sufficiency.

Nonetheless, we should have doubts about the viability of Kyburg's program for using the RR predicate to solve inverse inference through direct inference because *representativeness* is always limited to a few aspects of the total evidence and we have no guarantee that these aspects form a sufficient summary of all the data. Of course, if conditionalization were valid epistemologically, then sufficiency would be, too, and our doubts would be assuaged. But this answer is of no consequence for Kyburg, since he is adamant in his denial of conditionalization (as Levi's analysis clearly shows). What we discover, from the example sketched below, is that sufficiency fails in epistemological theory, with the upshot that inverse inference fails because the randomness conditions do *not* pick out the *sufficient* aspects of the evidence for characterizing representativeness.

Imagine a problem where we are to determine the volume V of a hollow cube about which we are currently "ignorant." In epistemological terms, given our present knowledge, only trivial $[\epsilon, 1 - \epsilon]$ probability intervals attach to hypotheses about V. For instance, $\mathrm{EP}_K(6 \leq V \leq 12) = [\epsilon, 1 - \epsilon]$.[17] We have at our disposal a simple experimental procedure for collecting evidence. A liquid of known density, say 1 unit weight/unit volume, can be poured into the cube to fill it exactly, and then we can weigh the quantity of liquid on a scale. We assume considerable knowledge about this scale; namely, that it is unbiased with an error component that is closely approximated by a normal distribution with known variance σ^2. That is, on separate (independent) weighings of an object with true weight W, observed weights w_i are distributed in accord with the familiar normal curve with mean (center) at W and variance (spread) σ^2.

A single measurement w provides a datum for inverse inference about W. In epistemological theory a solution to this inference problem exists by reduction to a direct inference about the rational representativeness of the observed weight. Call an observation w rationally representative at the qth level just in case w does not differ from W by more than q standard deviations $(=q\sigma)$, i.e., so long as $|W - w| \leq q\sigma$. From several interesting properties of the normal distribution (see Appendix 2), we know the exact frequency of RR samples, at any level. The randomness rules convert this information into a claim that, before measurement, the observed

[17] I suppress the units, for simplicity. In all cases the magnitude of the measurement is all that matters.

weight will be a *random* member of the set of weighings with respect to representativeness (and known frequency). As in the examples discussed above, after w is observed (since the new evidence is not demonstrably relevant) the randomness claim remains valid, and direct inference about the representativeness of the sample results in a solution to the inverse-inference problem. Moreover, since $W = V$ (the liquid has unit weight/volume), the inverse inference about W is equivalent to inverse inference about V.

Instead of a single weighing of the liquid, we may take (statistically independent) repeated measurements, leading to the data $\{w_1, w_2, \ldots, w_n\}$, n large and odd. For such a data set there are alternative summaries, e.g., a condensation of information to the sample average $\bar{w} = \sum_i w_i \div n$, or to the sample median w_m [=that reading for which $(n - 1) \div 2$ observations are smaller, and the same number larger, than it]. With respect to each of these summaries we can define a rationally representative property. But the frequency of representative samples with respect to these aspects of the data conflict. The sample cannot be *random* with respect to both kinds of representativeness. (See Appendix 2.)

It is not uncommon, in direct inference, for there to be a surfeit of frequency information. The frequency of Fs among Rs can differ from the frequency of Fs among R^*s. If it is known that entity a is both an R and an R^*, then we must look to the *randomness* regulations for a proper reference class (and frequency bounds). If we are suitably ignorant of the frequency of Fs among the (narrower) class of R-and-R^*s, then each candidate for randomness may be defeated by the rival (because of the incompatible frequencies), with the upshot that only a trivial $[\epsilon, 1 - \epsilon]$ epistemological probability is valid for the hypothesis that a is F.

This is the circumstance we find ourselves in with the n-fold sample of weighings. Having seen $\bar{w}$ and w_m, we are forced to reject each claim that the sample is representative (with respect to the corresponding aspect) because information about the other aspect of the sample defeats the contention of randomness. Thus, there is no epistemological solution to the inverse-inference problem in this case, since no direct inference about representativeness is available. Only trivial $[\epsilon, 1 - \epsilon]$ probability intervals attach to hypotheses about V after the data are observed. But these probability values were valid before any experimentation. So the new evidence is irrelevant to the inverse inference, since there is no change in probabilities.

What is unsatisfactory about this result is that, in this problem $\bar{w}$ (and not w_m) is sufficient for the data, with respect to W. That is,

all the relevant features of the evidence (about V) are summarized in $\bar{w}$, if sufficiency is accepted. Then, assuming sufficiency, inference about V should not be defeated by learning w_m, if $\bar{w}$ is already known. Just as, for the binomial problem, information about the first flip does not defeat the randomness claim about the sample's representativeness, so too information about the median value of the sample should *not* defeat the randomness claim about the representativeness of the sample with respect to its average. But epistemological probability theory does not recognize this fact and sufficiency is violated. [See Appendix 2 for details.]

In addition to the losses to inverse inference that follow from the failure of sufficiency, epistemological probability suffers additional defects because of the very general rejection of conditionalization. To the best of my knowledge no other violation of the conditionalization principle is the result of a failure of some inductive principle that is entailed by conditionalization. Instead, I can categorize the losses only by listing the sorts of problems that are left unsolved. In each case, epistemological theory is peculiar because precise probabilities are licensed which can serve as initial probabilities in Bayes' theorem, yet conditionalization is not satisfied.[18]

Setting aside the difficulties with sufficiency, the first of three kinds of problems that go unanswered in Kyburg's theory is exemplified by the historically important Behrens-Fisher problem. Kyburg's objection to conditionalization in this problem repeats Neyman's grounds for rejecting Fisher's solution.[19] To put matters as simply and generally as I can, the inverse problem about unknown ϕ is intractable because the attempted reduction to direct inference is blocked by ignorance of some other nuisance factor ζ. However, inverse inference about ζ is possible (by reduction to direct inference). In summation, we have a precise epistemological probability defined for hypotheses H_ζ, given data d: denoted $Q_K(H_\zeta/d)$. Also, precise epistemological probability exists for hypotheses H_ϕ,

[18] I use this opportunity to correct an error I made in my review of Kyburg's recent book [see this JOURNAL, LXXIV, 1 (January 1977), 47–62]. On page 60 of my review I mistakenly charged Kyburg's theory with a violation of *order invariance* (see sec. II below). In fact his system will satisfy order invariance by resorting to broad probability intervals when all the evidence is known. Also, I take this opportunity to explain that Kyburg's "compromise" between bayesian/nonbayesian methods agrees with the former *only* when the prior epistemological probability is directly supported by known frequencies, and not when the epistemological probability is obtained by using randomness conditions to suppress evidence (as irrelevant) just because it cannot be shown to be relevant.

[19] See Jerzy Neyman, "Fiducial Argument and the Theory of Confidence Intervals," *Biometrika*, XXXII (1941): 128–150.

given d and knowledge of ζ: denoted $Q_K(H_\phi/d \,\&^\circ\, H_\zeta)$. Conditionalization solves the problem, since (assuming conditionalization):[20]

$$Q_K(H_\phi/d) = \sum_\zeta Q_K(H_\phi/d \,\&^\circ\, H_\zeta)\cdot Q_K(H_\zeta/d)$$

Even though each of the epistemological probabilities needed to solve the right-hand side of this equation is precisely defined, conditionalization does not hold, and this equation is epistemologically invalid.[21] Thus, Kyburg's system does not generally permit the statistically important technique of "integrating out" nuisance parameters, even when there is a well-defined probability for the nuisance factor.

The second of the three varieties of inverse inference which suffer under epistemological theory can be described under the heading, "combining data of two kinds." Here, data from one kind of experiment support inverse inference about some unknown ϕ; that is, epistemologically, a precise inverse solution exists: $Q_K(H_\phi/d)$. A second experimental report, d', is also available, but d and d' represent different kinds of information which *cannot* be combined into a single report for one inverse inference about ϕ. (It does not matter here whether d' supports an inverse inference by itself.) For example, suppose d is the reported sample average $\bar{w}$ of one set of weighings, and suppose d' is the reported sample median w_m of a *different* set of weighings (all weighings of the same liquid). Suppressing the problem discussed earlier, let us grant a precise epistemological solution to the inference problem, from data d to hypotheses about ϕ: denoted $Q_K(H_\phi/d)$. Also, a simple direct inference about possible values of d', given knowledge of ϕ, exists: denoted $Q_K(d'/H_\phi)$. Moreover, the assumption that d and d' are statistically independent yields the equality:

$$Q_K(d'/H_\phi \,\&^\circ\, d) = Q_K(d'/H_\phi)$$

If conditionalization were epistemologically valid, these *precise* epistemological probabilities would be sufficient to determine pre-

[20] Except for the change to continuous distributions, this is Fisher's method of solution for the Behrens-Fisher problem, with ϕ the difference in population means, and ζ (the nuisance factor) the population variance ratio. Other examples of this "step-by-step" method are found in R. A. Fisher, *Statistical Methods and Scientific Inference*, 3rd ed. (New York: Hafner, 1973). Fisher's commitment to conditionalization is demonstrated in his later works, e.g., "Some examples of Bayes' method of the experimental determination of probabilities *a priori*," *Journal of the Royal Statistical Society*, B, 24 (1962): 118–124.

[21] Kyburg's solution to the bayesian/nonbayesian compromise does not apply here, since the epistemological probabilities are not directly supported by frequency distributions. See fn 18.

cise probabilities for hypotheses about ϕ, given all the data, $d \mathbin{\&^{\circ}} d'$. That is, assuming conditionalization:

$$Q_K(H\phi/d \mathbin{\&^{\circ}} d') \propto Q_K(d'/H_\phi \mathbin{\&^{\circ}} d) \cdot Q_K(H_\phi/d)$$

Once again conditionalization fails, and the only solutions to inverse questions (once all the evidence is known) are trivial $[\epsilon, 1 - \epsilon]$ intervals.[22]

The last of the triad of statistical problems that are left unsolved can be described as a composite of the other two. Here we are given one experimental report, say d, and, in the absence of knowledge of ϕ, we must hypothesize about the other experiment d'. That is, to use Carnap's account, we are faced with a problem of "singular predictive inference" with data of two kinds. Assuming conditionalization:

$$Q_K(d'/d) = \sum_{\phi} Q_K(d'/H_\phi \mathbin{\&^{\circ}} d) \cdot Q_K(H_\phi/d)$$

Once again, each of the credal probabilities in the right-hand side of this equality is epistemologically well defined and given a precise value. But the equation is invalid in Kyburg's theory and the inverse problem is left with only trivial $[\epsilon, 1 - \epsilon]$ answers.[23]

Let us tally up the scorecard on epistemological probability. Kyburg claims that conditionalization is not a sound inductive principle, and defends this view by developing a theory that does not obey the suspect axiom. I contend that Kyburg's defense is inadequate because his theory survives without conditionalization only by failing to resolve basic problems of inverse inference. The catalogue of unsolved problems has the peculiarities that: (i) sufficiency (hence, conditionalization) is invalid; and (ii) each of three other kinds of inference is invalid: inverse inference with nuisance factors, inverse inference with data of different kinds, and singular predictive inference with data of different kinds. What is common to these diverse failures is that, were conditionalization valid, they would not exist.

I conclude that Kyburg has *not* given us a satisfactory frequentist position because the violations of conditionalization (and sufficiency) are anything but banal.

II

Kyburg's position represents an extreme frequentism, due to the extensive rejection of conditionalization. We find a more moderate

[22] Evidence of Fisher's use of conditionalization to solve this kind of problem is found in chapter 5 of his last book. See fn 20.

[23] Again, see chapter 5 of Fisher's last book for numerical examples.

frequentism in Ian Hacking's elegant reconstruction of Fisher's fiducial probability. Hacking offers us a program that solves inverse inference by reduction to direct inference, a program that denies the existence of precise prior probabilities in inverse inference, yet a program that (unlike Kyburg's) recognizes conditionalization once the appropriate probabilities are defined.[24] Hence, we can appreciate Hacking's concern for logical consistency (and completeness) of his reconstruction, as evaluated against the family of alternative (coherent) bayesian inferences. That is, loosely put, Hacking's frequentism takes the existence of a coherent bayesian model as a condition of adequacy for inferences licensed by his reconstruction of fiducial inference (150/1).

In contrast with Kyburg's theory, Hacking considers a reduction of inverse to direct inference which is limited to problems admitting a precise (point-valued) probability solution. The trick in his approach is to identify a privileged (hypothetical) random variable, called the *pivotal* variable, which is the object of a simple direct inference before the data are observed, and then by appeal to a special irrelevance principle to retain this direct inference after learning the experimental outcomes, thereby inducing a solution to the inverse-inference problem.

We may rehearse Hacking's argument with the example of the hollow cube. Remember that we are investigating the volume V of a hollow cube. Our experiment consists in filling the cube with a liquid of known density, one unit weight/unit volume, and then weighing the volume of liquid on a scale whose readings w are normally distributed with mean W (the weight of the liquid) and known variance σ^2. The random quantity $(W - w)$ is a pivotal variable whose distribution is known precisely, to wit: a normal distribution with mean 0 and variance σ^2. Thus, before the liquid is weighed (w unobserved) there is a simple direct inference to hypotheses about the next value of the pivotal variable, the value it will have after w is fixed. For example, there is a credal probability of about .67 that $-\sigma \leq (W - w) \leq \sigma$. After the measurement we know the magnitude w (the observed weight), and this direct inference about the pivotal variable is no longer valid unless the extra information is irrelevant to hypotheses about $(W - w)$.

Hacking presents a novel "principle of irrelevance" which provides a formal criterion for assessing irrelevance claims regarding pivotal variables (149/50). Whereas Kyburg's position grants

[24] Ian Hacking, *Logic of Statistical Inference* (New York: Cambridge, 1965), especially chapter 9.

irrelevance unless relevance is demonstrated, the effect of Hacking's irrelevance principle is to deny irrelevance unless it is demonstrated. In the example above, the additional information w passes the test for irrelevance and the direct inference survives intact. Suppose we observe $w = 15$, then the direct inference about the pivotal variable becomes a credal probability of about .67 that $-\sigma \leq (W - 15) \leq \sigma$, or equivalently, that $15 - \sigma \leq W \leq 15 + \sigma$. This last statement is a solution to inverse inference about W. Hence, by appeal to an irrelevance principle, inverse inference about the volume of the cube is reduced to direct inference about the pivotal variable.

During the past half-century, since Fisher first introduced fiducial inference through pivotal variables, the statistical literature has been replete with paradoxes showing fiducial probability to be incoherent. A common theme in these counterexamples is the construction of alternative pivotal variables from the same data and parameter(s) for inverse inference, such that the fiducial probability obtained through one pivotal variable is inconsistent with that obtained through a different pivotal variable. For instance, in the example discussed on page 717 (repeated weighings of the liquid), both random quantities $(W - \bar{w})$ and $(W - w_m)$ are pivotal variables, yet the solutions to inverse inference about W that follow from these are inconsistent. (Remember that there is a conflict between the frequencies for rational representativeness of the two sample aspects, page 718.)

It is a tribute to Hacking's reconstruction that his principle of irrelevance solves many of these "paradoxes" by eliminating all but one family of pivotals for any given problem, and that family yields a common, coherent solution to the inverse-inference problem. Moreover, Hacking claims, the solution obeys confirmational conditionalization as well (154/5). For example, only the pivotal using the *sufficient* statistic $\bar{w}$ is legitimated by the irrelevance principle, which is what we would expect since *sufficiency* is a consequence of conditionalization. Finally, if Hacking is correct about the validity of conditionalization in his theory, then each of the three varieties of statistical problems considered unsolvable in Kyburg's epistemological-probability theory is given a precise, coherent answer. In the discussion that follows, I will show that Hacking's program is unsatisfactory because an important inductive principle, *order invariance* (a consequence of conditionalization), fails. Hence, Hacking's reconstructed fiducial argument violates conditionalization. Lastly, I will argue that no correction exists for this deficiency short of reverting to Kyburg's account where only trivial $[\epsilon, 1 - \epsilon]$ probabilities are valid.

The inductive principle I choose to investigate next postulates that no effect can result from the order of acceptance of evidence. Subject to conditionalization, the degree of belief in a hypothesis h, upon learning the compound evidence d_1 & d_2, does not depend either upon the sequence in which the data are accepted or upon whether the data are accepted in one fell swoop. Hence,

$$\textit{Order invariance: } Q_{K_{1\&2}}(h) = Q_{K_{1_2}}(h) = Q_{K_{2_1}}(h)$$

where K_{i_j} is the corpus of knowledge that results from corpus K by first accepting datum d_i and then enlarging this new body of knowledge by accepting datum d_j; and where $K_{i\&j}$ is the corpus of knowledge that follows from accepting the conjunction d_i & d_j into K. [See Appendix 3 for the derivation of order invariance from conditionalization.]

What is the significance of this rule for problems of statistical inference? Suppose we flip the newly bent coin n_1 times and, on the basis of this evidence, assess alternative hypotheses about the magnitude of the bias. Next, if the experiment is continued so that another n_2 flips are made, shall the evaluation of the competing hypotheses be changed by the order in which we consider the two sets of outcomes? Moreover, shall our evaluations change once again when we lump together the two sets of trials into one sequence of $n_1 + n_2$ flips? Order invariance secures the identity of solutions to the inverse-inference problem, regardless of how we choose to accept the total evidence available.[25]

[25] It is interesting to note that traditional significance testing fails order invariance. For example, the commonplace practice of significance testing for a multinomial distribution (of which the text's binomial distribution is a special case) by the use of a χ^2 test obeys the right-hand, but fails the left-hand equality of order invariance.

For a simple case in point, the χ^2 test of the null hypothesis that the coin is "fair," based on n_i flips of which h_i land heads up, determines χ^2 (with one degree of freedom) by:

$$\chi^2 = 4(h_i - (n_i \div 2))^2 \div n_i$$

For the composite sequence of $n_1 + n_2$ flips, $n_i = n_1 + n_2$ and $h_i = h_1 + h_2$. However, the recommended policy for combining significance tests from independent trials is to use the rule that the sum of (independent) χ^2 has a χ^2 distribution with the sum of the degrees of freedom of the independent tests. [See, for instance, Fisher, *Statistical Methods for Research Workers* (New York: Hafner, 13th ed., 1973).] If, for instance, the two runs (n_1 and n_2) are of equal length, with 40% heads in the first string and 60% heads in the second, then $\chi^2 = 0$ for the combined $n_1 + n_2$ sequence (with one degree of freedom); yet each χ^2 (hence, also the sum which is χ^2 with two degrees of freedom) is positive.

An alternative analysis for this failure of order invariance is that the rule for combining independent significance tests relies only on the χ^2 values of the separate tests. It is easy enough to show that, within a binomial model, the χ^2 value of a test is *not* sufficient for the original data used in the test.

Reconsider, for the final time, the problem of the hollow cube. For simplicity, we assume only a single weighing of the liquid, with the observation w_L. Imagine, also, that we have at our disposal a different experiment in which we cut a rigid rod of known density, say one unit weight/unit length, to the length of an edge of the cube. As with the liquid, we may weigh the segment of rod on our scale and observe the weight w_R. Now we have an inverse-inference problem about V which involves data of two kinds (see the discussion of the second variety of inverse inference, page 720).[26] But there is a pivotal variable connecting the parameter of interest V to each datum. $(W_L - w_L)$ is a pivotal variable that can serve for inverse inference about V (since $W_L = V$, where W_L is the true weight of the liquid), if w_L is measured. Also, $(W_R - w_R)$ is a pivotal variable that can serve for inverse inference about V (since $W_R = \sqrt[3]{V}$, where W_R is the true weight of the rod), if w_R is measured. Most importantly, each pivotal satisfies the special irrelevance principle, with respect to its counterpart datum.

Let us rehearse the different solutions to the inverse-inference problem which follow by conditionalization and by Hacking's fiducial inference. First, observe w_L and solve the fiducial inference about V using the pivotal variable $(W_L - w_L)$. The result is a fiducial probability for values of V, which is precisely described as a normal probability distribution with mean w_L and variance σ^2. The second datum, w_R, can be combined with this credal probability by conditionalization (as described on page 721). An alternative solution begins with the fiducial inference about V using the other pivotal variable $(W_R - w_R)$ and observation w_R. The result is a fiducial probability for values of $\sqrt[3]{V}$ which is precisely described as a normal probability distribution with mean w_R and variance σ^2. Again, the other datum, w_L, can be combined with this credal probability by conditionalization. Unfortunately, the results are not the same for the two procedures!

We can trace the failure of order invariance (and thus of conditionalization too) by examining the bayesian models that support the fiducial inferences. With pivotal variables, like these, which have normal distributions, the bayesian argument that duplicates the fiducial inference is one whose prior probability for the parameter appearing in the pivotal is a *uniform* prior probability. Thus, the fiducial inference about W_L [using pivotal $(W_L - w_L)$] can be modeled by a bayesian argument where the initial credal probability

[26] The data (w_L, w_R) do not admit of a single sufficient statistic, hence the description as data of two kinds.

for values of W_L is a flat, uniform distribution. Similarly, the fiducial inference about W_R [using pivotal $(W_R - w_R)$] can be modeled by a bayesian argument where the initial credal probability for values of W_R is a flat, uniform distribution.[27]

Uniform distributions are commonly used in bayesian theory to represent *ignorance*. The oft-cited Laplacean principle of Insufficient Reason can be relied upon to generate uniform "ignorance" distributions. However, it is common knowledge that indiscriminate use of Insufficient Reason quickly leads to contradictions. For example, with an unknown whose possible values form a continuum, representation of those values by a parameter ϕ leads to a uniform distribution of ϕ values under Insufficient Reason. But transform the parameterization to ϕ^3 and the resulting uniform distribution for ϕ^3 values is *inconsistent* with the uniform distribution for ϕ values.

Hacking is fully aware that his fiducial argument has bayesian models that require uniform "ignorance" probabilities. He finds this result acceptable and argues that it provides a satisfactory formulation of Insufficient Reason, satisfactory because (allegedly) the theory is consistent.[28] He is not alone among those who would try to find a viable construal of Laplace's intuitively pleasing dictum.[29] However, the failure of order invariance is tied to the consistency of Insufficient Reason. A uniform distribution for values of the weight of the liquid, W_L, is equivalent to a uniform distribution for values of the volume V of the cube. A uniform distribution for values of the weight of the rod, W_R, is equivalent to a uniform distribution for values of the length of an edge of the cube, which is

[27] See D. V. Lindley, "Fiducial Distributions and Bayes' Theorem," *Journal of the Royal Statistical Society*, B, xx (1958): 102–107.

[28] Hacking's error arises from an apparent misunderstanding of Lindley's central theorem (Lindley, *op. cit.*). Lindley's result is limited to data all of one kind! See, Hacking, *op. cit.*, pp. 146–155.

Levi has shown (in an unpublished manuscript) that Hacking's principle of irrelevance (as needed for fiducial inference) follows from his law of likelihood. Hence the law of likelihood is inconsistent with direct inference and conditionalization. Since the more familiar likelihood principle [see A. Birnbaum, "Concepts of Statistical Evidence," in S. Morgenbesser, P. Suppes, and M. White, eds., *Philosophy, Science, and Method* (New York: St. Martin's, 1969)] both entails order invariance and follows from conditionalization, we see that Hacking's law of likelihood is inconsistent with direct inference and the likelihood principle.

Finally, Hacking's remarks about "initial" support are somewhat unclear. He points out, correctly, that unbounded uniform "ignorance" distributions are improper. However, he concludes (and here I disagree) that, therefore, they are not suitable as initial measures of inductive belief. He also suggests that, as a formal postulate, "ignorance" distributions are validated when his irrelevance principle leads to identical results. I interpret Hacking to mean that "ignorance" distributions are justifiable only when fiducial inference is possible.

[29] See, for instance, Rosenkrantz, *op. cit.*

equivalent to a uniform distribution for values of the *cube root* of the volume V. One inference procedure assumes ignorance modeled by a uniform distribution for values of V, and the other inference procedure assumes ignorance modeled by a uniform distribution for values of $\sqrt[3]{V}$.

The violation of order invariance (hence of conditionalization) within Hacking's exceptionally lucid reconstruction points to the basic flaw that pervades the frequentist strategy for reducing inverse to direct inference. If conditionalization is to be valid once credal probabilities are well defined (and how else is the theory to provide answers to the varieties of problems discussed in the previous section?), then a precise solution to inverse inference will be modeled by some precise "ignorance" probability. But just which prior probability is chosen to model ignorance will depend upon the kind of evidence to be acquired. No one "ignorance" probability will serve to model the reduction of inverse to direct inference for all the kinds of data that might be accepted. Thus the parallel between the choice of a privileged parameterization for applying Insufficient Reason and the choice of a privileged random variable for direct inference is both clear and instructive. Just as frequentists are prepared to reject the bayesian solution to inverse inference because the assumption of a precise prior probability is found to be untenable, so too they should not expect a reduction of inverse to direct inference because the assumption of a special variable to support the direct inference is incompatible with even a weakened commitment to conditionalization.

TEDDY SEIDENFELD

University of Pittsburgh

Appendix 1

A demonstration that conditionalization entails sufficiency.

Assumptions and notation: Let K be a consistent, deductively closed set of sentences that represent a body of knowledge. Let $Q_K(\ )$ be a coherent probability function, based on the corpus K. Let K_i be the corpus that results from K after data i are accepted. Let d be the new evidence available, and let t $(f(d) = t)$ be some simplification of this evidence.

With respect to hypotheses H_θ about some unknown θ,

To show: if $Q_K(d) > 0$, then $Q_{K_d}(H_\theta) = Q_{K_t}(H_\theta)$ whenever t is sufficient for d with respect to θ.

Proof: If $Q_K(d) > 0$ then $Q_K(t) > 0$. By conditionalization,

$$Q_{K_t}(H_\theta) = Q_K(t/H_\theta) \cdot Q_K(H_\theta) \div Q_K(t)$$

and

$$Q_{Kd}(H_\theta) = Q_K(d/H_\theta)\cdot Q_K(H_\theta) \div Q_K(d) \qquad (*)$$

and, since t is a simplification (function) of d,

$$Q_K(d \,\&\, t) = Q_K(d), \text{ i.e., } Q_K(t/d) = 1 \qquad (**)$$

Substituting (**) into (*),

$$Q_{Kd}(H_\theta) = Q_K(d \,\&\, t/H_\theta)\cdot Q_K(H_\theta) \div Q_K(d \,\&\, t)$$

and, by the multiplication axiom,

$$Q_{Kd}(H_\theta) = Q_K(d/t \,\&\, H_\theta)\cdot Q_K(t/H_\theta)\cdot Q_K(H_\theta) \div [Q_K(d/t)\cdot Q_K(t)]$$

and, by conditionalization,

$$Q_{Kd}(H_\theta) = Q_K(d/t \,\&\, H_\theta)\cdot Q_{Kt}(H_\theta) \div Q_K(d/t)$$

Thus, $Q_{Kd}(H_\theta) = Q_{Kt}(H_\theta)$ just in case $Q_K(d/t \,\&\, H_\theta) = Q_K(d/t)$.

Appendix 2

In this appendix we examine, in some detail, the epistemological-probability solution to the problem discussed in section I of the paper. Following conventional notation, let $N(\mu, \sigma^2)$ stand for the normal distribution with mean μ and variance σ^2. In the example considered, the sequence $(w_1, \ldots, w_n)$ form a data set of n (n moderately large and odd) mutually independent readings from a $N(V, 1)$ distribution. We investigate two statistics (reductions) of the data: to $\bar{w}$ (sample average) and m (sample median).

The relevant distributions for the sample statistics are:

(i) $\bar{w}$ has a $N(V, 1/n)$ distribution.
(ii) m has, approximately, a $N(V, \pi/2n)$ distribution.
(iii) $\bar{w}$ is sufficient; so the conditional distribution of m, given $\bar{w}$, is approximately $N(\bar{w}, .57/n)$.

From the distribution of $\bar{w}$, it follows that the percentage of n-fold samples with $\bar{w}$ within $\sqrt{j/n}$ units of V, i.e., the percentage of n-fold samples satisfying $RR_{\sqrt{j/n}}(\bar{w})$, is just the percentage of a normal distribution within $\sqrt{j}$ standard deviations of the mean. That is, approximately 95% of all n-fold samples satisfy $RR_{2/\sqrt{n}}(\bar{w})$.

Similarly, RR applies to sample medians; so that the percentage of n-fold samples with a median within $(\sqrt{\pi/2})(\sqrt{j/n})$ units of V, i.e., the percentage of n-fold samples satisfying $RR_{\sqrt{\pi j/2n}}(m)$, is just the percentage of a normal distribution within $\sqrt{j}$ standard deviations of the mean. That is, approximately 95% of all n-fold samples satisfy $RR_{(\sqrt{\pi/2})(2/\sqrt{n})}(m)$.

The distributions that would follow for V, if both $RR(\bar{w})$ and $RR(m)$ support appropriate randomness claims, are precise and mutually inconsistent. Hence, unless trivial epistemological probability statements about V are to result from the data $(\bar{w}, m)$, one of the randomness claims

[randomness of either RR(m) or RR($\bar{w}$)] must be defeated. Moreover, if the sufficiency principle is to hold, then only the randomness claim regarding RR($\bar{w}$) is valid.

Thus, to show that sufficiency *fails* in epistemological probability theory it is enough to demonstrate that either that

(a) knowledge of $\bar{w}$ does *not* defeat the purported randomness claim: the n-fold sample is a random member of all n-fold samples with respect to RR(m)—relying on frequencies generated by (*ii*) (above); or that

(b) knowledge of m defeats the purported randomness claim: the n-fold sample is a random member of all n-fold samples with respect to RR($\bar{w}$)—relying on frequencies generated by (*i*) (above).

Kyburg's regulations for determining valid randomness claims are extremely complicated. There are several prevention and counterprevention clauses to each basic strategy for showing (or defeating) a randomness claim. At the time of this writing (and following much correspondence with Kyburg on the point) it appears that (b) is true. However, to clinch the matter, let me show how (a) holds as well.

Suppose the evidence consists of the reported m value only. Then, since $\bar{w}$ is unknown, there is no challenge to the randomness claim of (*a*) (above). If $\bar{w}$ becomes known, does it defeat this randomness claim? Because $\bar{w}$ (and not $\bar{m}$) is sufficient for V, by the sufficiency principle $\bar{w}$ is relevant to inference about V if only m is known. But Kyburg's randomness rules do not agree with this result of sufficiency.

The conditional distribution of m, for known $\bar{w}$, is found in (iii) (above). If we consider the reference set of all n-fold samples with a fixed sample mean (fixed at $\bar{w}$) then the distribution of m values will be $N(\bar{w}, .57/n)$, approximately. That is, in this conditional reference set the m values tend to be distributed about $\bar{w}$, with a distribution independent of V (since $\bar{w}$ is sufficient for V). The randomness regulations for testing the effect of this conditional distribution on the claims in (a) are found in Kyburg's randomness principles I and II (p. 227 of his book), which can be summarized as: narrower reference sets get priority over competing frequency distributions, unless the narrower set also carries a broader (containing) interval of frequencies than the intervals associated with the wider reference set.

When $\bar{w}$ is close to V, the conditional reference set has frequencies for RR(m) which exceed the corresponding frequencies for RR(m) unconditionally. When $\bar{w}$ is far from V, the conditional reference set has frequencies for RR(m) which fall far below the corresponding frequencies for RR(m) unconditionally. Thus, in the narrower reference set, obtained by fixing the sample average, the distribution for RR(m) ranges from frequencies below, to frequencies above the frequencies in the unconditioned reference set of all n-fold samples. Hence, by randomness conditions I and II, $\bar{w}$ does *not* prevent the randomness claim in (1)—contrary to the dictates of sufficiency.

Appendix 3

A demonstration that conditionalization entails order invariance.

Assume: $Q_K(d_1 \,\&\, d_2) > 0$.
To show: $Q_{K_{1\&2}}(h) = Q_{K_{1_2}}(h) = Q_{K_{2_1}}(h)$.
Proof:

$$\text{If } Q_K(d_1 \,\&\, d_2) > 0, \text{ then } Q_K(d_i) > 0 \qquad (i = 1, 2)$$

since

$$Q_K(d_1 \,\&\, d_2) \leq Q_K(d_i) \qquad (i = 1, 2)$$

By conditionalization,

$$Q_{K_{1\&2}}(h) = Q_K(d_1 \,\&\, d_2/h) \cdot Q_K(h) \div Q_K(d_1 \,\&\, d_2)$$

and, by the multiplication axiom,

$$= Q_K(d_1/d_2 \,\&\, h) \cdot Q_K(d_2/h) \cdot Q_K(h) \div [Q_K(d_1/d_2) \cdot Q_K(d_2)] \qquad (*)$$

and

$$= Q_K(d_2/d_1 \,\&\, h) \cdot Q_K(d_1/h) \cdot Q_K(h) \div [Q_K(d_2/d_1) \cdot Q_K(d_1)] \qquad (\dagger)$$

That is, (*) = (†).
By conditionalization,

$$(*) = Q_{K_2}(d_1/h) \cdot Q_{K_2}(h) \div Q_{K_2}(d_1) \qquad (**)$$

and

$$(\dagger) = Q_{K_1}(d_2/h) \cdot Q_{K_1}(h) \div Q_{K_1}(d_2) \qquad (\dagger\dagger)$$

Finally, by conditionalization,

$$(**) = Q_{K_{2_1}}(h) \qquad \text{and} \qquad (\dagger\dagger) = Q_{K_{1_2}}(h)$$

TS

COMMENTS AND CRITICISM

CONFIRMATIONAL CONDITIONALIZATION

HENRY KYBURG and I agree that direct inference plays an important role in scientific inquiry and practical deliberation. We disagree as to whether knowledge of chances is indispensable as an ingredient of such inference or whether, as Kyburg insists,* such knowledge may be replaced by knowledge of relative frequencies.

In "Direct Inference,"† I *demonstrated* that Kyburg's account of

* "Randomness and the Right Reference Class," this JOURNAL, LXXIV, 9 (September 1977): 501–521, hereafter RRRC; parenthetical page references to Kyburg are to this paper.

† "Direct Inference," *ibid.*, LXXIV, 1 (January 1977): 5–29, hereafter DI; parenthetical page references to my work are to this paper.

0022-362X/78/7512/0730$00.70 © 1978 The Journal of Philosophy, Inc.

http://www.jstor.org

LINKED CITATIONS

You have printed the following article:

Direct Inference and Inverse Inference
Teddy Seidenfeld
The Journal of Philosophy, Vol. 75, No. 12. (Dec., 1978), pp. 709-730.
Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197812%2975%3A12%3C709%3ADIAII%3E2.0.CO%3B2-R

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

[5] **Direct Inference**
Isaac Levi
The Journal of Philosophy, Vol. 74, No. 1. (Jan., 1977), pp. 5-29.
Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197701%2974%3A1%3C5%3ADI%3E2.0.CO%3B2-W

[6] **Randomness and the Right Reference Class**
Henry E. Kyburg, Jr.
The Journal of Philosophy, Vol. 74, No. 9. (Sep., 1977), pp. 501-521.
Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197709%2974%3A9%3C501%3ARATRRC%3E2.0.CO%3B2-L

[19] **Fiducial Argument and the Theory of Confidence Intervals**
J. Neyman
Biometrika, Vol. 32, No. 2. (Oct., 1941), pp. 128-150.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28194110%2932%3A2%3C128%3AFAATTO%3E2.0.CO%3B2-8

NOTE: *The reference numbering from the original has been maintained in this citation list.*