Janina Hosiasson and the Value of Evidence

Christian Torsell

May 16, 2024

Abstract

I.J. Good's "On the Principle of Total Evidence" (1967) looms large in decision theory and Bayesian epistemology. Good proves that in Savage's (1954) decision theory, a coherent agent always prefers to collect, rather than ignore, free evidence. It is now well known that Good's result was prefigured in an unpublished note by Frank Ramsey (Skyrms 2006). The present paper highlights another early forerunner to Good's argument, appearing in Janina Hosiasson's "Why do We Prefer Probabilities Relative to Many Data?" (1931), that has been neglected in the literature. Section 1 reviews Good's argument and the problem it was meant to resolve; call this the *value of evidence problem*. Section 2 offers a brief history of the value of evidence problem and provides biographical background to contextualize Hosiasson's contribution. Section 3 explicates the central argument of Hosiasson's paper and considers its relationship to Good's (1967).

Keywords: Janina Hosiasson-Lindenbaum, formal epistemology, probability, induction

1

In 1957, A.J. Ayer's "The Logical Conception of Probability" was published in the proceedings of a conference on the philosophy of physics held by the Colston Research Society. According to the view referenced in its title, there is an important sense of "probability" in which "what is being asserted when it is said that a statement is probable, in this sense, is that it bears a certain relation to another statement, or set of statements, which may also be described as confirming, or supporting, or providing evidence for it" (Ayer, 12). On this view, paradigmatic probability claims take the following form: "h is probable to degree p given e as evidence" (we'll follow Ayer, who follows Keynes (1921), in abbreviating this with "$h/e = p$"), where h and e

are propositions and p is a real number in the unit interval. On the logical interpretation, the meanings of e and h uniquely determine the value of p. Since the meanings of the relevant propositions are sufficient to determine the probability of the one given the other, probabilities do not depend on the attitudes of any particular agent toward the propositions in question. In this sense, probabilities are *objective* on the logical conception.

Ayer raises a challenge for this view. Suppose we are considering betting on a horse named "Eclipse" in an upcoming race (Ayer, 13). Let h be the statement "Eclipse will win the race". Let e_1 be the statement "Eclipse will be ridden by the champion jockey," and let e_2 be the conjunction of e_1 with many other statements pertinent to predicting Eclipse's performance. Suppose that $h/e_1 = p_1$ and $h/e_2 = p_2$. It seems clear that, given that e_1 and e_2 are among our available evidence (whatever this might mean), we ought to take p_2, and not p_1, as *the* probability of h. At least, we'd certainly prefer to place our bet on h on the basis of p_2 rather than p_1.

What makes probabilities based on more evidence better than, or preferable as guides to action to, probabilities based on less evidence? According to the logical conception, there is an important sense in which each probability is as good as the other: the probability of h *really is* p_1 relative to e_1, just as it *really is* p_2 relative to e_2. It is true, just in virtue of the meanings of e_1 and h, that $e_1/h = p_1$ *and* it is true, just in virtue of the meanings of e_2 and h, that $h/e_2 = p_2$, and it is unclear what grounds we might have for privileging one of these probabilities over the other. Ayer's objection is that the logical conception of probability lacks the resources to explain why we should prefer probabilities based on more evidence to probabilities based on less evidence. Call the challenge of rationalizing this preference the *value of evidence problem*.

I.J. Good took on the value of evidence problem in a three-page note titled "On the Principle of Total Evidence," published in *The British Journal for the Philosophy of Science* in 1967. The centerpiece of that paper is a short proof of a theorem in Savage's (1954) decision theory.

In Savage's decision theory, states and outcomes are taken as primitive, where the set of states S represents ways the world might be that are outside an agent's control and about which she is uncertain (e.g., whether it will rain this afternoon) and the set of outcomes O represents states of affairs the agent ultimately cares about (e.g., whether she gets wet on her afternoon walk). A set A of acts is defined as the collection of all functions from S to O. Savage proved that, if A is sufficiently rich and an agent has preferences over acts, represented by a binary relation $\succeq$ on A, that obey a few coherence constraints taken as requirements of rationality, then her preferences can be represented with a pair consisting of a unique probability function $P : S \to [0,1]$ and a utility function $U : O \to \mathbb{R}$ unique up to positive affine transformation such that, for all $A_0, A_1 \in A$,

$$A_0 \succeq A_1 \Leftrightarrow \mathbb{E}\left[U(A_0)\right] \geqslant \mathbb{E}\left[U(A_1)\right],$$

where $\mathbb{E}$ denotes expectation relative to P.

P represents the agent's degrees of belief about how the world is with respect to S, and U represents something like the overall desirability of outcomes. This result establishes that (assuming Savage's constraints on preferences are genuine requirements of rationality) a rational agent's preferences go by expected utility relative to her probabilistic degrees of belief. So, a rational agent can be represented as an expected utility maximizer.

Notice that in this setting probability is interpreted as a measure of a particular agent's subjective degrees of belief, not as a logical relation. Probabilities reflect certain of the agent's *attitudes* toward the propositions whose probabilities are being considered; they aren't fixed by the meanings of the propositions alone.

Good has us consider a rational agent facing a decision problem in which she considers a set of acts $A_1, A_2, ..., A_s$ and a set of mutually exclusive and exhaustive hypotheses $H_1, H_2, ..., H_r$. Good's resolution of the value of evidence problem consists in a proof that if the agent has an opportunity to learn new evidence by making a costless observation, the expected utility of first making the observation and then choosing (from A) on the basis of her expanded evidence is always at least as great as, and possibly greater than, the expected utility of passing up on the new evidence a choosing on the basis of her prior information. This establishes that, if an agent's preferences are coherent, she *must* prefer (at least weakly) acting on the basis of probabilities based on more rather than less evidence, assuming the cost of acquiring more evidence is negligible.

Here is a sketch of the proof, following Good's presentation. Let $U(A_j \mid H_i) = u_{ij}$ denote the utility of choosing A_j given that the true state is an element of H_i. Suppose the agent has some evidence $E \subseteq S$ (i.e., she knows the true state lies in E) so that her prior probabilities are given by $P(H_i \mid E)$ for $i = 1, 2, ..., r$. Going forward, we will drop reference to the background evidence E, writing the agent's prior probability for H_i simply as $P(H_i)$. Since she has coherent preferences, our agent will choose an act which maximizes expected utility relative to these probabilities. Suppose A_0 is an expectation-maximizing act. With just E as her evidence, then, the expected utility of our agent's act is equal to

$$\max_j \left(\sum_{i=1}^{r} P\left(H_i\right) u_{ij} \right) = \sum_{i=1}^{r} P\left(H_i\right) u_{i0}.$$

Now, suppose the agent has the opportunity to perform a costless experiment with possible mutually exclusive and exhaustive outcomes $E_1, E_2, ..., E_t$. These outcomes define posterior probabilities over the H_i,

$$P\left(H_i \mid E_k\right),$$

which describe, for each hypothesis H_i, how her probability for H_i will change upon adding the outcome E_k to her evidence. We might think of these as defining a *plan* that specifies how the agent will change her belief in each H_i contingent on each possible outcome of the experiment.

Our agent knows that, if she performs the experiment, she will subsequently choose the expected-utility-maximizing act relative to her updated probabilities: that is, given that the observed outcome is E_k, she will choose an act A_j that maximizes the value of $\sum_{i=1}^{r} P(H_i \mid E_k) u_{ij}$. Since the experiment is costless and the prior probability for each E_k is equal to $\sum_{i=1}^{r} P(H_i) P(E_k \mid H_i)$, the expected utility of performing it and acting on her expanded evidence is given by

$$\sum_{k=1}^{t} \left[\left(\sum_{i=1}^{r} P\left(H_i\right) P\left(E_k \mid H_i\right) \right) \max_j \left(\sum_{i=1}^{r} P\left(H_i \mid E_k\right) u_{ij} \right) \right],$$

which is equal to

$$\sum_{k=1}^{t} \max_j \left(\sum_{i=1}^{r} P\left(H_i\right) P\left(E_k \mid H_i\right) u_{ij} \right).$$

And since $E_1, E_2, ..., E_t$ form a partition over E, we can rewrite the expected utility of acting without learning the outcome of the experiment (i.e., choosing the act which maximizes expected utility relative to our prior probabilities) as

$$\max_j \left[\sum_{k=1}^{t}\left(\sum_{i=1}^{r} P\left(H_i\right) P\left(E_k \mid H_i\right) u_{ij}\right)\right].$$

Note that, for any t and any real-valued function, f, of j and k,

$$\sum_{k=1}^{t} \max_j \left[f(j,k)\right] \geqslant \max_j \left[\sum_{k=1}^{t} f(j,k)\right].$$[1]

Letting $f = \sum_{i=1}^{r} P(H_i)P(E_k \mid H_i)u_{ij}$, it follows that

$$\sum_{k=1}^{t} \max_j \left[\sum_{i=1}^{r} P\left(H_i\right) P\left(E_k \mid H_i\right) u_{ij}\right] \geqslant \max_j \left[\sum_{k=1}^{t}\left(\sum_{i=1}^{r} P\left(H_i\right) P\left(E_k \mid H_i\right) u_{ij}\right)\right],$$

with strict inequality unless the set of expected-utility-maximizing acts is identical for each possible experimental outcome (and so the additional evidence from performing the experiment makes no difference to the agent's choice, relative to what she would have chosen on the basis of E alone). So, the expected utility of acting on the basis of the more-informed probabilities is always at least as great, and sometimes greater than, the expected utility of acting on the basis of the less-informed probabilities.

Good notes that his result may be taken as establishing only that, given the opportunity, one should always choose to acquire *additional* evidence by means of cost-free observation, whereas Ayer raised the value of evidence problem in terms of Carnap's "Principle of Total Evidence," which is the injunction to take into account all of one's *currently available* evidence in calculating probabilities. The objection is that Good has not resolved the value of evidence problem as Ayer posed it unless he has motivated Carnap's principle.

Good answers that we can consider our currently available evidence as constituting a kind of record, where consulting the record is itself an observation—one that can be modeled in the same way as the experiment in the sketched proof above. So understood, it is clear that Good's result has "justified the decision to make this observation and to use it, provided that the cost is negligible" (Good, 320). As long as consulting the record is practically costless, it pays to consult it until our present stock of evidence is exhausted.

[1] Good treats this statement as a lemma, offering a very short proof on p. 320 of his (1967).

2

Interest in Good's landmark paper is by no means merely historical. "On the Principle of Total Evidence" has spawned a literature concerning value of evidence results for generalizations of conditionalization. Notable contributions include Graves (1989), Skyrms (1990), and Huttegger (2014). These fruits of Good's project are well known. Less well known are its seeds.

Neither Ayer's problem nor Good's solution were without precedent. Five pages of C.D. Broad's *Perception, Physics, and Reality* (1914)—adapted from his 1911 doctoral dissertation—are devoted to difficulties arising in connection with the principle that "we ought to prefer a probability calculated on a wider to one calculated on a narrower basis, even though the man who only had the narrower basis of knowledge had made his calculations properly" (Broad (1914), 151). Seven years after the publication of Broad's book, John Maynard Keynes' landmark *Treatise on Probability* (1921) was published, including a chapter on "The Application of Probability to Conduct" in which Keynes prefigures Ayer's challenge:

> [I]f two probabilities are equal in degree, ought we, in choosing our course of action, to prefer that one that is based on the greater body of knowledge? The question appears to me to be highly perplexing, and it is difficult to say much that is useful about it. But the degree of completeness of the information upon which a probability is based does seem to be relevant, as well as the actual magnitude of the probability, in making practical decisions. Bernoulli's maxim that in reckoning a probability we must take into account all the information which we have, even when reinforced by Locke's maxim that we must get all the information that we can, does not seem completely to meet the case. (345-6)

Ayer, then, was preceded by at least Broad and Keynes in highlighting the value of evidence problem. In a 1986 visit to the Frank Ramsey archives at Cambridge, Brian Skyrms discovered that Good's strategy for resolving the problem has a similarly long history. There, Skyrms found a two-page note titled "Weight, or the Value of Knowledge" in which Ramsey proves a result analogous to Good's (Skyrms 2006), apparently intended as a resolution to the value of evidence problem as it appeared in Keynes (1921).

Between Ramsey's note and Good's paper, there is Savage's independent proof of the value of information theorem in chapter 7 of *The Foun-*

dations of Statistics (1954, 125-7). Good himself acknowledges the influence of Raiffa and Schlaifer's treatment of "The Value of Sampling Information" in their *Applied Statistical Decision Theory* (1961, 89) and a statement of part of his (1967) proof in Lindley's *Introduction to Probability and Statistics* (1965). And a footnote in the 1957 Colston proceedings reports that astronomer Ernst Öpik had "produced a purely formal mathematical argument by which he claims to show that if we increase the amount of information on which we calculate probabilities to guide our actions, then the expectation of gain resulting from these actions will increase" (Ayer, 23). (No proof appears in the Proceedings or, as far as I have found, in any other published material).

A less well known forerunner to Good's argument appears in Janina Hosiasson's[2] "Why do We Prefer Probabilities Relative to Many Data?" (1931) (henceforth "Probabilities Relative to Many Data"). Born in Warsaw in 1899, Hosiasson was a logician and philosopher closely associated with the Lwów-Warsaw School. She received her doctorate from the University of Warsaw in 1926, where she wrote a dissertation on the "Justification of Inductive Reasoning" under logician Tadeusz Kotarbínski. Hosiasson would spend the next fifteen years writing extensively (in four languages) on issues related to probability and induction. Little is known about the details of Hosiasson's professional life after earning her doctorate, though Anna Jedynak (2001) reports that Hosiasson "combined her scientific research with work in a secondary school as a teacher of philosophy" (Jedynak, 97). In 1940, Hosiasson published her best-known work, "On Confirmation," notable for including the first published discussion of Carl Hempel's "raven paradox." [3] Two years later, Hosiasson would be murdered by the Gestapo in Vilnius, where she had fled in the wake of the Nazi invasion of Warsaw in 1939.

"Probabilities Relative to Many Data" was published in January 1931, shortly after a visit to Cambridge spanning the 1929/30 academic year. Hosiasson's primary interests were well represented at her host university: in addition to Keynes and Broad, Richard Braithwaite, Harold Jeffreys, and Frank Ramsey (until his death in January 1930) were employed by Cambridge at the time. It was likely during this visit that Braithwaite, a lecturer in moral sciences and close friend of Ramsey, shared Ramsey's then-unpublished

[2] Hosiasson published under multiple names, including "Lindenbaum" and "Hosiasson-Lindenbaum" following her marriage to Adolf Lindenbaum in 1935. For simplicity, I follow Marta Sznajder in using "Hosiasson" throughout this paper.

[3] Although Hempel would not publish on the problem until 1945, he had shared a version of it with Hosiasson in conversation when the two met in 1937 (Niiniluoto, 332).

"Truth and Probability" (written in 1926) with Hosiasson. "Truth and Probability" is significant for its defense of a subjectivistic conception of probability—according to which probabilities are interpreted as the degrees of belief of particular agents, in contrast to the Keynesian logical interpretation—and for including a very early example of a representation theorem deriving an expected utility representation of an agent's choice behavior from her preferences alone. Hosiasson was impressed by Ramsey's paper, seeing it as developing views about probability close to those she had independently arrived at, and acknowledges its influence on her approach to the value of evidence problem in a footnote in the 1931 paper.[4]

Hosiasson opens "Probabilities Relative to Many Data" by claiming that the probability of a given event depends on the evidence relative to which we consider that event. Hosiasson discusses the evidence-relativity of probabilities in terms of the descriptions under which we consider the relevant events. Different descriptions of some event may include different bits of evidence relevant to assessing how likely it is to occur. So a given event may have different probabilities relative to different descriptions. Hosiasson offers an example:

> If we take into account the probability that this card lying face downwards on the table is a court-card [i.e., a jack, queen, or king], we may have regard to the fact that a minute ago somebody has drawn it from a pack of fifty-two playing cards and reckon the probability as 3/13; but we may also, by a nearer examination of the back of the card, find that there is a mark on it, and we may know that amongst the marked cards only 1/5 are court-cards. After taking the mark into account our probability will be other than before. (Hosiasson 1931, 23)

As in Ayer's setup, the puzzle arises from the fact that in both cases the probability we assign to the event that the card on the table is a court card is that which we "*should* take into account" (emphasis mine). That is, both probabilities are, in some sense, "correct," given our information. But, Hosiasson notes, there seems to be something *better* about the probability based on the more informative description. Hosiasson devotes the rest of the paper to the question: "Why are we the more satisfied with our probability the more particulars about the given case it takes into consideration?" (24).

Hosiasson considers and rejects several candidate solutions to the value

[4] See Section 3 of this paper.

of evidence problem before presenting the answer which “seems to [her] the most satisfactory” (30) in the paper’s fourth section. In fact, Section 4 includes two closely related arguments. Both arguments concern a decision problem involving repeated bets and invite the reader to compare a case in which the chooser has more information about the events on which she will bet to a case in which she has less.

The first argument shows that, other things being equal, a rational decision maker facing this problem would always *in fact* realize (weakly) greater gains in this setting if she entered it with more rather than less evidence, on the assumption that the probabilities guiding her choices are equal to the empirical relative frequencies of the relevant events. The second argument aims to explain the value of additional evidence without making strong assumptions relating the bettor’s decision-making probabilities to the empirical frequencies of the events on which she bets. It shows that the sum of *expected* gains (relative to the bettor’s subjective probabilities) of the individual bets is weakly larger in the case in which the choosing agent has more information relative to the case in which she has less. The remainder of the present paper is dedicated to reconstructing Hosiasson’s arguments and explicating their relationship to the argument of “On the Principle of Total Evidence."

Though “Probabilities Relative to Many Data" has received much less attention than “On Confirmation," it has not been entirely ignored. In a paper on the “The Sessions on Induction and Probability at the 1935 Paris Congress" (at which Hosiasson was present), Galavotti (2018) highlights Hosiasson’s suggestive comments about the interpretation of probabilities in connection with Ramsey’s “Truth and Probability." And the paper receives passing mention in Hilpinen (1970), Peden (2018), Sznajder (2021, 2022), and Horwich (1982). But in none of these works is the connection with Good’s argument discussed explicitly.

3

Hosiasson opens Section 4 by clarifying how probabilities are to be interpreted in the arguments that follow. She explains: “In a considerable number of cases in ordinary life we take account of [probabilities] by considering the amount of something which could be said to be a mathematical expectation” (30). As an illustration, she offers the following example:

> A photographer has to decide whether to go or not, tomorrow,

> Sunday, to a country town to take some photographs. The photos can be taken only if the weather is fine. Whether he will decide to go or not will not only depend on the probability of fine weather, but also on the gain he may get by going if it is fine and the loss if it rains…he will consider the good he will get by going if it is fine and if it is not fine, on the one side, and the good he will get by not going on the other, and the corresponding probabilities-the whole taken together in the form of a difference of mathematical expectations. (32)

In cases like this one, in which the probabilities we "take into account" are probabilities of individual events, "our procedure consists…in adjusting our action so as to have the biggest mathematical expectation" (35). Hosiasson sometimes uses "gain" and other times "good" to refer to the quantity, representing something like overall desirableness, whose expectation is maximized in rational choice.

That acting so as to maximize expected gain is characteristic of *rational* choice is taken as axiomatic, though Hosiasson considers the possibility of justifying the recommendation to maximize expected gain by pointing out that, if probabilities are equal to the relative frequencies of the relevant events, then "by choosing the greatest mathematical expectation we get—by the realisation of the frequency—the greatest amount of good. (Compare with Ramsey (1931), who treats the principle of expected utility maximization as a law of human psychology.)

Probabilities as they figure in Hosiasson's arguments, then, are to be interpreted in terms of their role in guiding rational choice subject to uncertainty. They are the weights used to calculate expected gain for alternative courses of action. The language of "mathematical expectation" may be borrowed from Ramsey, who in "Truth and Probability" (1931) adopts a use of the phrase similar to Hosiasson's. It is in a footnote following her first characterization of probabilities in terms of mathematical expectations that Hosiasson thanks Braithwaite for furnishing an opportunity to read that paper, and acknowledges a debt to Ramsey for "for clearness on this question," despite having "previously thought along similar lines" herself (30).

Both of Section 4's arguments involve a particular decision problem. Consider an event e, and let $G_i = (k_i, l_i)$ stand for a gamble in which we receive k_i if e occurs and l_i if e does not occur, k_i and l_i denoting quantities of "good" (33). Suppose that in each of n many separate events in which e may or may not occur, we are to choose one gamble from among $\mathbf{G} =$

$G_1, G_2, ..., G_k$. e might, for example, be the event that a certain coin comes up heads. In this example, the coin would be flipped n many times, and before each flip, we would asked to choose one from a set of gambles which specify how much "good" we get in the event that the coin comes up heads, and how much "good" we get in the event that the coin comes up tails (or fails in some other way to land heads-up). After each flip, we receive the payoff corresponding to the outcome of the flip according to the gamble we chose.

Hosiasson invites us to compare two versions of this scenario, α and β. In α, "we know in each of the n cases only one general description of the event, say A, to which we refer its probability [i.e., the probability of e occurring], which is say, $\frac{m}{n}$" (33). In that case, we'll calculate the expected value of each gamble G_i by taking the sum $\frac{m}{n}k_i + (1 - \frac{m}{n})l_i$, and, since we're rational, we'll choose the gamble that maximizes that value relative to our probability for e based on the description A: call that gamble G_a (with k_a and l_a as the payoffs we receive if e occurs or not, respectively).

In β, we get more information:"[W]e consider in the n cases another factor, say a character C in each case of A, and have in n_1 of the n cases a closer description of the event, say AC_1, giving the probability $\frac{m_1}{n_1}$; n_2 of the n cases a closer description of the event, say AC_2, giving the probability $\frac{m_2}{n_2}$;" and so on through AC_s, "where $C_1, C_2, ..., C_s$ are determinates of the determinable C" (33).

For concreteness, we might think of e as the event that a coin pulled from a bag containing coins of various colors comes up heads when flipped. Suppose we know that all the coins of a given color have the same (known) bias, but different colors correspond to different biases. Let A be the description according to which the coin was drawn from a bag in which the *average* bias of the coins in that bag is 0.5 (i.e., the "average coin" from the bag is fair), and let AC_i be a description that gives the color of the particular coin being flipped. In this case, we may think of G_a as the gamble that maximizes expected gain relative to the probability $P(e \mid A) = 0.5$ (since, if all we know about a given coin is that it was drawn from the bag in question, we will expect it to land heads-up with probability 0.5.) This is the gamble we will choose in each of the n coin flips in α.

In β, we are informed, prior to each trial, of the color of the coin to be flipped—given our background knowledge, this informs us of the bias of the coin to be flipped. Suppose, for example, that AC_1 says that the coin to be flipped is red, where red coins are known to land heads-up with chance

0.8, while AC_2 says that the coin to be flipped is blue, where blue coins land heads-up with chance 0.3. Suppose further that we set our probabilities for e equal to these chances when they are known. Then, in each of the n_1 cases in which a red coin is flipped, we will choose the gamble G_1 (with an expected value of $\frac{m_1}{n_1}k_1 + (1 - \frac{m_1}{n_1})l_1 = 0.8k_1 + 0.2l_1$) which maximizes expected gain relative to our color-informed beliefs, according to which the probability of the coin landing heads is equal to $P(e \mid A\&AC_1) = 0.8$. Similarly, in the n_2 cases in which a blue coin is flipped, we'll choose G_2 (with an expected value of $\frac{m_2}{n_2}k_2 + (1 - \frac{m_2}{n_2})l_2 = 0.3k_2 + 0.3l_2$). In this case, our probability that a given coin lands heads-up may vary from trial to trial, and so which gamble maximizes expected gain may vary between trials, too. So, a rational agent's pattern of choices may differ between α and β.

Having established how an expectation-maximizer would choose in α and β, Hosiasson gives her first argument: if in both cases the frequencies of the relevant outcomes are equal to their probabilities, we will *in fact* realize more total good in β than in α. The argument proceeds by showing that the sum of expected gains for each of the n trials, with expectation taken with respect to the true empirical frequencies of each outcome, is necessarily no smaller (and is possibly larger) in β than in α.

Hosiasson assumes in this argument that the probabilities with respect to which we take the expected gain of each gamble are equal to the empirical relative frequencies of the relevant events. In the coin-flipping example above, for example, this assumption would entail that $P(e \mid A\&AC_1)$ is precisely equal to the proportion of flips of red coins that result in the coin landing heads-up in the n_1 cases in which a red coin is flipped. Similarly, $P(e \mid A)$ is assumed to be identical to the empirical relative frequency of heads-up outcomes among all n flips.

Since in both α and β we choose from among the same set of gambles in each of the n cases, G_a is an available option in every choice among gambles we make. So, since we know we will choose the expected value-maximizing gamble in each case, we know that in any of the n cases in β, we will choose a gamble other than G_a, call it G_b, *only if* $\frac{m_j}{n_j}k_b + (1 - \frac{m_j}{n_j})l_b \geqslant \frac{m_j}{n_j}k_a + (1 - \frac{m_j}{n_j})l_a$. That is, in any given individual choice in β, we will choose a gamble other than G_a only if the expected value of choosing G_b is at least as great as that of choosing G_a. It follows that, for all j,

$$\frac{m_j}{n_j}k_j + (1 - \frac{m_j}{n_j})l_j \geqslant \frac{m_j}{n_j}k_a + (1 - \frac{m_j}{n_j})l_a,$$

and so

$$\sum_{j=1}^{s} n_j[\frac{m_j}{n_j}k_j + (1-\frac{m_j}{n_j})l_j] \geqslant \sum_{j=1}^{s} n_j[\frac{m_j}{n_j}k_a + (1-\frac{m_j}{n_j})l_a].$$

Since we assumed that the probabilities guiding our choices are equal to the empirical relative frequencies of the relevant events, we can think of the left-hand side of the above inequality as representing the total amount of "good" we will *in fact realize* by betting rationally in β, while the right-hand side represents how much "good" we will realize by betting rationally in α. The direction of the inequality indicates that we will gain more in β than in α. If there is some b such that $\frac{m_j}{n_j}k_b + (1-\frac{m_j}{n_j})l_b > \frac{m_j}{n_j}k_a + (1-\frac{m_j}{n_j})l_a$, then the inequality is strict. Informally: if in *any* case the extra information we have in β makes a difference to our choice (relative to what we would've chosen in α), then, assuming probabilities to be equal to relative frequencies, we gain strictly more in β than in α.

This is the first argument: assuming that an agent's probability for each event she considers is equal to the empirical relative frequency of that event, then given that an agent acts to maximize expected "good," she will always realize at least as much "good," and sometimes more, in the setting in which her probabilities are based on more evidence (β) than in an otherwise identical setting in which her probabilities are based on less evidence (α).

Hosiasson is not satisfied that this argument gives an adequate response to her central question. As the title of her paper indicates, Hosiasson's goal is to rationalize our preference for *probabilities* based on more evidence over less-informed probabilities, and for Hosiasson there is no necessary connection between probabilities and relative frequencies. In her first argument, Hosiasson takes the proposition that probabilities are equal to relative frequencies as a substantive assumption, and she clearly treats "taking account of gains and expectations" as conceptually distinct from "foreseeing frequencies" (34). So the first argument, insofar as its central result depends on the assumption that probabilities of events are equal to their relative frequencies, does not settle the matter, from Hosiasson's perspective.

In developing her second argument, Hosiasson deals as far as possible in "expectations and gains" alone, avoiding assumptions tying probabilities to frequencies of events. This argument is meant to establish that

> If we took for granted that the best way of acting in different cases is to act so as to make the sum of mathematical expecta-

tions as big as possible (without trying to explain this rule by assuming frequencies to be equal to probabilities), then we could show that we act in a better way, if we take account, in particular cases, of probabilities relative to more data, without assuming *all* frequencies to be equal to probabilities in the considered group of cases. (35)

Hosiasson admits that even this argument will not entirely eliminate assumptions about frequencies: "It is...sufficient to assume only that frequencies of cases with different descriptions are equal to their probabilities; i.e. ...it is sufficient that AC_j occurs in n_j of n cases of A (its probability being $\frac{n_j}{n}$) for $j = 1, 2, ..., n$" (35). The assumption that has been dropped is that the probabilities of the possible outcomes of each trial are equal to their relative frequencies. More is said about this assumption—and, more generally, about the role of frequencies in Hosiasson's arguments—below.

Hosiasson's second argument invites us to compare the sum ranging over the *expected values* (in terms of good/gain) of each member of the sequence of n bets we make in α (on the implicit assumption that we expect with certainty that in each bet we will choose so as to maximize expected value) to the same quantity for β. In the case of α, we have

$$\begin{aligned} E_\alpha &= n[\frac{m}{n}k_a + (1 - \frac{m}{n})l_a] \\ &= n[k_a \sum_{j=1}^{s} \frac{n_j}{n}\frac{n_j}{m_j} + l_a \sum_{j=1}^{s} \frac{n_j}{n}(1 - \frac{n_j}{m_j})] \\ &= \sum_{j=1}^{s} n_j[\frac{m_j}{n_j}k_a + (1 - \frac{m_j}{n_j})l_a] \end{aligned}$$

Whereas for β, we have

$$E_\beta = \sum_{j=1}^{s} n_j[\frac{m_j}{n_j}k_j + (1 - \frac{m_j}{n_j})l_j].$$

By the same mathematical reasoning deployed in the first argument, we have it that

$$E_\beta \geqslant E_\alpha$$

with strict inequality if the additional information in β makes any difference to the gambles we expect to pick.

The role played by relative frequencies in both these arguments is puzzling. As we've seen, in setting up the arguments of Section 4, Hosiasson characterizes probability in terms of subjective uncertainty and indicates her sympathy with Ramsey's "Truth and Probability." But for a Ramsey-style subjectivist, the assumption that the bettor's degrees of belief are equal to the empirical relative frequencies of the events she bets on is hard to motivate. If Hosiasson's interpretation of probabilities really is Ramsey's[5], include this argument?

Hosiasson is characteristically terse in "Probabilities Relative to Many Data"; the text itself does not resolve, or even raise, this puzzle. One possibility is that the inclusion of the first argument reflects the influence of Jan Łucasiewicz, an influential student of Kazimierz Twardowski (as was Hosiasson's doctoral advisor, Kotarbinski) and professor of philosophy at the University of Warsaw from 1915 until 1939 (excepting a one-year break in 1919-20 to serve in the Polish government). Łucasiewicz defended a logical conception of probability with a frequentist flavor. In *Die Logische Grundlagen der Wahrscheinlichkeitsrechnung* (1913), Łucasiewicz identified probability with a non-standard notion of truth value. In his system, given a formula p containing a free variable x ranging over a finite set V, the truth value of p is given by the ratio $|W|/|V|$, where W is the set of all elements w of V such p is made into a true sentence when w is substituted for x (see Niiniluoto, 328). As Ilkka Niiniluoto (1998) notes, when "translated into terms more familiar in probability theory," it is clear that "his definition is equivalent to saying that probability is the relative frequency of an attribute in a reference class": "An indefinite proposition like 'x is black' corresponds to an attribute (being black) of objects or events, and the range of variable x is the reference class" (Niiniluoto, 328). Given Łucasiewicz's considerable influence in the intellectual milieu he shared with Hosiasson, we might speculatively interpret the first argument as an effort to explicate the value of more-informed probabilities in a way that would be compelling to those with views closer to Łucasiewicz's than Ramsey's (leaving what Hosiasson herself regarded as the stronger argument for later).

Even if this speculative suggestion is right, it leaves unexplained the assumption relating the probabilities and empirical frequencies of the AC_js in the second argument. It seems clear that Hosiasson does not need that

[5]While the text of "Probabilities Relative to Many Data" does not clearly settle this question, Hosiasson has been read as an early exponent of subjectivism in the philosophy of probability, as highlighted above.

assumption to get the desired inequality. In fact, *probabilities* for the AC_js do not show up anywhere in Hosiasson's equations. The sum of expected gains for β simply takes the total expected gain for choosing, in all of the trials in which AC_j obtains, the expectation-maximizing gamble G_j to be equal to a fixed value n_j multiplied by the expected gain of choosing G_j in a single trial in which AC_j obtains. No expression of the form "$P(AC_j)$" appears anywhere. This suggests an interpretation on which Hosiasson's bettor *already knows* how many trials will satisfy each AC_j when she calculates expected gains in anticipation of the gambling problem. Hosiasson does not make clear whether this is the interpretation she had in mind. But the absence of reference to the probabilities of the AC_js outside the quotation expressing the assumption that those probabilities be equal to their empirical relative frequencies makes it difficult to identify the role she intended that assumption to play in her second argument.

The reliance on assumptions about empirical frequencies marks a difference between Hosiasson's and Good's resolutions to the value of evidence problems—all the probabilities that appear in Good's argument are interpreted as subjective degrees of belief. But of course, the mathematics at the heart of their arguments is not affected by different interpretational choices.

Besides the different approaches to the interpretation of the probabilities that appear in their results, there are some other differences between Good's and Hosiasson's arguments. The central result in Hosiasson's paper concerns a special, highly structured decision problem (though she suggests that "we could try to justify our desire for closer [i.e., more informative] descriptions in other more complicated cases in a similar way" [34]). Good's result is more general. Another difference, of course, is that Good's result is proved as a theorem of Savage's decision theory, which would not be developed until after well after the publication of "Probabilities Relative to Many Data."

What I want to highlight is that, despite these differences, Hosiasson's argument is substantially similar to Good's. Like Good, Hosiasson offers a resolution to the value of evidence problem appealing to a principle of practical rationality. And although Hosiasson did not have the benefit (as Good did) of writing after significant development of decision theory as a mature discipline in its own right, her argument is developed within a proto-decision-theoretic framework. For Hosiasson, as for Good, the reason we should prefer to act on the basis of probabilities based on more rather than less evidence is that, other things equal, the expected utility of making a more informed choice is always greater than or equal to the expected utility of making the same choice with less information. Since for both Hosiasson and Good rationality requires one's preferences to go by expected utility (Hosiasson's "mathematical expectation"), it follows that, other things equal, we violate a norm of rationality if we do not prefer acting on the larger evidentiary basis to acting on the smaller. This is a kind of pragmatic solution: as Hosiasson notes, "The answer...this paper gives, i.e. taking gains or mathematical expectations into account, could be considered an epistemological answer only from a pragmatistic point of view" (36).

It is notable that Good suggests that the primary contribution of "On The Principle of Total Evidence" lies in highlighting the relationship between Ayer's problem and practical rationality: "Perhaps the main value of the present note is that it makes explicit the connection between Carnap's principle of total evidence and the principle of rationality [i.e., the principle that rational choice maximizes expected utility], a connection that was overlooked by seventeen distinguished philosophers of science [i.e., Ayer and the discussants of his paper at the Colston conference]" (321). It is striking that Hosiasson, like Ramsey, saw Good's central point more than thirty years before the publication of "On the Principle of Total Evidence."

References

[1] Ayer, A. J. (1957). The conception of probability as a logical relation. In *Observation and Interpretation*, ed. by S. Körner. London: Butterworths. Pages 12-30, including discussion.

[2] Broad, C. D. (1914). *Perception, physics, and reality: an enquiry into the information that physical science can supply about the real.* Cambridge: Cambridge University Press.

[3] Galavotti, M. C. (2018). The sessions on induction and probability at the 1935 Paris Congress: An overview. *Philosophia Scientiæ. Travaux d'histoire et de philosophie des sciences*, 22(3), 213-232.

[4] Good, I. J. (1967). On the principle of total evidence. *British Journal for the Philosophy of Science*, 17(4), 319-321.

[5] Graves, P. R. (1989). The total evidence theorem for probability kinematics. *Philosophy of Science*, 56(2), 317-324.

[6] Hilpinen, R. (1970). On the information provided by observations. In *Information and inference*, pp. 97-122. D. Reidel Publishing Company.

[7] Horwich, P. (1982). *Probability and evidence*. Cambridge: Cambridge University Press.

[8] Hosiasson, J. (1931). Why do we prefer probabilities relative to many data?. *Mind*, 40(157), 23-36.

[9] Hosiasson-Lindenbaum, J. (1940). On confirmation. *The Journal of Symbolic Logic*, 5(4), 133-148.

[10] Huttegger, S. (2014). 'Learning experiences and the value of knowledge'. *Philosophical Studies*, 171(2):279–288.

[11] Janssen-Lauret, F. (2021). Women in logical empiricism. In *The Routledge handbook of logical empiricism*, pp. 127-135. London: Routledge.

[12] Jedynak, A. Janina Hosiasson-Lindenbaum—The logic of induction. In *Poznan Studies In the Philosophy of the Sciences and the Humanities, vo. 74* (W. Krajewski, ed.), pp. 97-101. Amsterdam: Editions Rodopi.

[13] Keynes, J.M. (1921). *A treatise on probability*. London: MacMillan.

[14] Lindley, D. V. (1965). *Introduction to probability and statistics, part 2*. Cambridge: Cambridge University Press.

[15] Niiniluoto, I. (1998). Induction and probability in the Lvov-Warsaw school. In *The Lvov-Warsaw School and Contemporary Philosophy* (pp. 323-335). Dordrecht: Springer Netherlands.

[16] Peden, W. (2018). Imprecise probability and the measurement of Keynes's "weight of arguments". *Journal of Applied Logics—IFCoLog Journal of Logics and their Applications*, 5(3), 677-707.

[17] Raiffa, H. and Schlaifer (1961). *Applied statistical decision theory*. Boston: Graduate School of Business Administration, Harvard University.

[18] Ramsey, F. P. (1926). "Truth and probability". In D. H. Mellor, ed., *Philosophical Papers*, Cambridge: Cambridge University Press (1990).

[19] Ramsey, F. P. (1990). Weight or the value of knowledge. *The British Journal for the Philosophy of Science*, 41(1), 1-4.

[20] Skyrms, B. (2006). Discovering "Weight, or the value of knowledge". In Galavotti, M.C. (eds) *Cambridge and Vienna*. *Vienna Circle Institute Yearbook* [2004], vol 12. Springer, Dordrecht.

[21] Skyrms, B. (1990). 'The value of knowledge'. *Minnesota Studies in the Philosophy of Science*, 14:245–266

[22] Sznajder, M. (2021). Probability in logical empiricism. In *The Routledge Handbook of Logical Empiricism*, pp. 220-228. London: Routledge.

[23] Sznajder, M. (2022). Janina Hosiasson-Lindenbaum on analogical reasoning: new sources. *Erkenntnis*. https://doi.org/10.1007/s10670-022-00586-y.