Sven Ove Hansson

Measuring Uncertainty

Abstract. Two types of measures of probabilistic uncertainty are introduced and investigated. Dispersion measures report how diffused the agent's second-order probability distribution is over the range of first-order probabilities. Robustness measures reflect the extent to which the agent's assessment of the prior (objective) probability of an event is perturbed by information about whether or not the event actually took place. The properties of both types of measures are investigated. The most obvious type of robustness measure is shown to coincide with one of the major candidates for a dispersion measure, the mean square deviation measure.

Keywords: uncertainty measure, second-order probability, readjustment, second-order probability, subjective probability, objective probability, dispersion measure, robustness measure.

1. Introduction

Since Knight's [13] classic account of the distinction between risk and uncertainty, the latter term has usually been reserved for such lack of knowledge that is not fully expressible in terms of a single probability statement. The following example can be used to illustrate the nature of (epistemic) uncertainty: A dime has been found among the property of a deceased cardsharp. We suspect that the coin may be unfair, but we have no clue to whether it is in that case biased towards heads or tails. Now suppose that someone decides to toss the coin. If I have to assign a probability that it will yield heads, then I will say 0.5. This is the same answer that I would give before someone tossed a coin that I knew to be fair. However, although the probabilities are the same, I am much more uncertain about the behaviour of the cardsharp's coin than about that of the ordinary coin. Although the probability is the same in the two cases, they differ in the degrees of uncertainty. Probabilities can be measured in numerical terms. Can degrees of uncertainty also be numerically measured?

It is the purpose of the present contribution to propose and compare two types of numerical measures of degrees of uncertainty. In Section 2, a framework for these developments will be informally introduced. It is based on the

Presented by **Heinrich Wansing**; *Received* September 4, 2007

Studia Logica (2009) 93: 21–40
DOI: 10.1007/s11225-009-9207-0

coexistence of objective and subjective probabilities. In Section 3, a formal representation of that framework is provided. In Section 4 two measures of uncertainty are introduced that are both based on the assumption that a high degree of uncertainty is associated with a large dispersion of second-order probability. In Section 5, an uncertainty measure is proposed that is based on another principle, namely that uncertainty concerning one's estimate of the (objective) chance of a possible event is associated with willingness to revise that estimate. A theorem that connects the two approaches concludes the section. All proofs of formal results are deferred to an Appendix.

2. A framework for uncertainty measurement

Although uncertainty is by definition not representable with a single probability function, it may be representable in other, more elaborate ways. One of the most common representations of uncertainty is in terms of second-order probabilities. [26, 24, 11] Second-order probabilities have mostly been discussed in relation to decision-rules that make use of them. A variety of such decision rules have been proposed. [3, 14, 15, 18, 5, 6]. In the 1980s a lively debate took place on the relative merits of some of these proposals. [16, 17, 18, 19, 6, 25, 22]. However, in this paper decisions or decision rules will not be discussed. Our focus will be on how to measure uncertainty, not on how do make decisions in its presence.

Examples with fair and biased coins are instructive, and will therefore be used throughout this paper. To begin with, suppose that I am certain that a particular coin is fair. Consider the second-order probability function that represents my subjective beliefs about the objective probabilities that rule the behaviour of this coin. This function will assign the probability 1 to the statement that the coin has the probability 0.5 of yielding heads. Next, suppose instead that I believe that the coin may be biased. Then the corresponding second-order probability assigns non-zero probability to various statements according to which the probability of heads differs from 0.5. The most obvious interpretation of such a two-levelled structure is that the first-order probability is objective (i.e. chance in Lewis's [20] terminology) whereas second-order probability is subjective (i.e. credence in Lewis's terminology; cf. also [2]). This is the interpretation to which I will refer in what follows. Since objective and subjective probabilities represent quite different entities, this interpretation has the advantage of being immune against the standard objection to second-order probabilities, namely that they have no barrier against repeating the reasoning that took us from first- to second-order probabilities, and thus no means to avoid an infinite hierarchy of higher and higher orders of probabilities. (Other interpretations of two-levelled probabilities are also possible. Hence first-order probabilities could represent the subjective probability assignments we would have made if we had access to a certain body of information. See [1].)

In what follows measures of uncertainty will be developed in a framework containing this two-levelled structure. The constructions will be based on three assumptions about subjective and objective probabilities:

> *First assumption*: Subjective probabilities and objective probabilities are separate entities each of which satisfies the standard probability axioms.

> *Second assumption*: Subjective probabilities can refer to what the objective probabilities were, are, or will be, at any point in time.

For our present purposes it is particularly important to note that subjective probabilities can refer to chances in the past. Suppose that you threw a fair dice yesterday and obtained a six. If I know this, then I may trivially say (i) that given what we know today, the (subjective or objective) probability that the die landed on six yesterday is equal to 1. However, I may also (ii) provide the best estimate of what the objective probability of landing on six was yesterday before the die was tossed. Given what we know today this estimate will expectedly be 1/6. The second type of statement is common in everyday parlance, but not easily expressible in models of probabilistic reasoning that merge all probabilistic statements into one single probability function.

> *Third assumption*: Subjective probabilities referring to what the objective probability of an event E was at a certain point in time can be revised in retrospect in response to various types of information, including information on whether or not E actually took place.

In other words, information about whether an event took place is among the information that can influence a post-event estimate of its pre-event probability. For a simple example, consider again the cardsharp's coin. Before we have tossed the coin, I have a strong suspicion that it is biased, but I have no clue as to the direction of the bias. Therefore, the best estimate that I can give of its propensity to yield heads in that toss is 0.5. Next, the coin is tossed, and yields heads. This outcome gives some support to the supposition that the coin is biased towards heads rather than towards tails. Therefore, my best estimate of the coin's propensity towards heads should now be higher than 0.5. If the coin is flipped repeatedly, and yields heads each time, then the estimated value should increase after each toss.

Posterior adjustments of pre-event probability are most easily accounted for in relation to a repeatable event type such as coin-flipping. However, it is important to observe that the probabilities of unique or non-repeatable events can be readjusted in the same way. Suppose that an accident occurs that we believed to be highly unlikely. A possible reaction is of course: "This was extremely improbable, so it must have been a case of unusually bad luck." However, a more common response, not least among accident investigators, is: "Since it happened it seems to have been more probable than we thought." This reaction involves an adjustment of the previous probability estimate (not necessarily expressed or even expressible in terms of second-order probabilities).

I will use the term *readjustment* for such post-event revisions of estimates of a pre-event probability that are triggered by information about whether the event actually took place. Readjustments seem to be a common and arguably indispensable component of informal probabilistic reasoning. However, they are not part of the standard repertoire of decision theory or probability theory. In order to account for them we need to distinguish carefully between the different types of probabilities that are involved. Otherwise, a readjustment will seem to be a conditionalization of an event on itself. Such conditional statements have the value 1 for all events. (The probability of E given that E is of course 1.) Authors who have overlooked these distinctions have often denounced readjustments as resulting from an irrational "hindsight bias". [4, 12])

In what follows, two approaches to the measurement of uncertainty will be investigated. They rely on different principles. *Dispersion measures* are based on the assumption that the more uncertain one is about the value of an objective probability, the more diffused is one's second-order probability measure over the range of first-order (objective) probability values. To exemplify the principle we can use the following version of the coins example: We know that the cardsharp's coin is one of three types of coin but we do not know which. It can be a "tails dime" with the (objective) probability 0 of yielding heads, a fair dime for which the corresponding probability is 0.5, or a "heads dime" for which it is 1. We may further assume for instance that the subjective probability is 0.25 that this is a tails dime, 0.5 that it is a fair dime, and 0.25 that it is a heads dime. Based on this principle, we can measure a person's uncertainty in a certain matter by measuring the dispersion of the estimated probability over the range between 0 and 1. For this we need a plausible dispersion measure. This approach will be developed in Section 4.

The second approach is based on the equally plausible assumption that a higher degree of uncertainty concerning a possible event is associated with greater willingness to revise one's subjective estimate of the (objective) chance of that event. If I am quite certain about some probability, then it takes more to make me change my estimate of it than if I am uncertain. However, willingness to revise cannot be measured in a fully general way. Typically, there are many potential pieces of information that can make us change our view of the probability of some event E. We cannot include all of them in a manageable measure. Fortunately, one of them can be used as a standard, thereby providing us with a method to measure willingness to revise.

To introduce the standard, suppose first that I am certain that a particular coin is fair. Someone tosses it, and the outcome is heads. My original estimate of the (objective) probability that this would happen was 0.5. The new information (one toss yielding heads) gives me no reason to change this estimate. Next, suppose instead that the cardsharp's coin is tossed and yields heads. As above, my subjective probability is 0.25 that this is a tails dime, 0.5 that it is a fair dime, and 0.25 that it is a heads dime. In this case, my initial estimate that there was a 0.5 chance of obtaining heads would have to be adjusted upwards. The reason for this is that we now know that the coin is not a tails dime, and that the hypothesis that it is a heads dime has been strengthened in comparison to the hypothesis that it is a fair dime. The extent to which my estimate of the (objective) probability of the event (heads in the first toss of the coin) is adjusted when I learn that the event actually took place indicates how uncertain I was concerning that probability. This readjustment can be used as the standard revision for determining willingness to revise that we need to measure uncertainty in a uniform manner. The details of this approach will be developed in Section 5.

3. Formal preliminaries

The usual notation for probabilities will be used. $p(H)$ denotes the probability of H, and p is assumed to satisfy the standard probability axioms. For second-order probabilities, the following notation will be used:

Definition 3.1. A *second-order probability distribution* is a set of pairs

$$\alpha = \{\langle w_1, p_1\rangle, \ldots \langle w_n, p_n\rangle\},$$

where each p_k is a probability function,

$$0 < w_k \leq 1 \text{ for each } w_k, \text{ and } \sum_{k=1}^{n} w_k = 1.$$

A second-order probability distribution is *unary* if it has exactly one element and *dual* if it has exactly two elements.

Two second-order probability distributions α and α' are *equivalent*, $\alpha \equiv \alpha'$, if and only if (i) $\alpha = \alpha'$, or (ii) one of them is obtainable through cleavage of an element of the other, e.g. $\alpha = \{...\langle x+y, p_k\rangle...\}$ and $\alpha' = \{...\langle x, p_k\rangle, \langle y, p_k\rangle...\}$, or (iii) they form the endpoints of a series $\alpha, \beta_1, ...\beta_n, \alpha'$ such that if two distributions are adjacent in this series, then one of them is obtainable through cleavage of an element of the other.

For any two second-order distributions $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_s, p_s\rangle\}$ and $\beta = \{\langle w'_1, p'_1\rangle, ...\langle w'_t, p'_t\rangle\}$ and any number k with $0 < k < 1$, the *mixture* $k\alpha \cup (1-k)\beta$ is equal to $\{\langle kw_1, p_1\rangle, ...\langle kw_s, p_s\rangle, \langle(1-k)w'_1, p'_1\rangle, ...\langle(1-k)w'_t, p'_t\}$.

For simplicity and in order to uphold cognitive realism [9, 10], all second-order probability distributions will be assumed to have a finite number of elements. The overall probability function associated with a second-order probability distribution is obtained in the obvious way:

$$p(H \uparrow \alpha) = \sum_{\langle w_k, p_k\rangle \in \alpha} w_k \times p_k(H)$$

Uncertainty measures are denoted by m. Subscripts to m are used to distinguish between different such measures. Hence, $m(H)$ denotes the uncertainty concerning H, and $m(H \uparrow \alpha)$ the uncertainty concerning H that is inherent in the second-order probability distribution α.

4. Dispersion measures

This section is devoted to the introduction and characterization of uncertainty measures that operate by measuring the dispersion in second-order probability distributions. The use of dispersion measures to measure uncertainty is, of course, not a new idea. Ever since Markowitz's seminal work [23], it is part of the standard repertoire of economic analysis to use the mean variance of the (first-order) probability over a variable as a measure of its uncertainty. Here, dispersion measures will instead be applied to the second-order probability distribution over first-order probability.

In the statistical literature, two major types of dispersion measures have been developed, namely those based on (1) the average of the squared difference between a particular value and the average value (statistical variance),

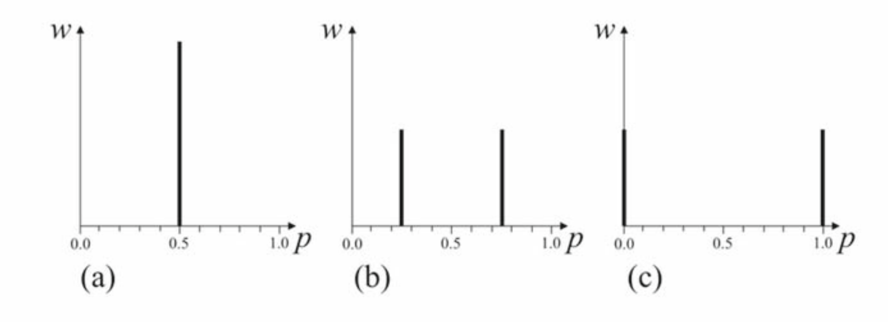

Figure 1.

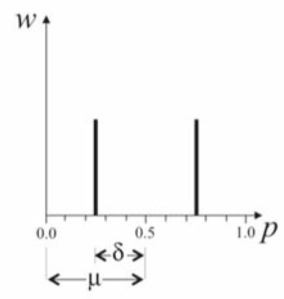

Figure 2.

respectively (2) the average of the absolute value of the difference between a particular value and the average value (average absolute deviation). These types of measures both have a long tradition, and neither can be rejected prior to a careful investigation of its properties. We will therefore investigate both types of measures, beginning with the former.

A measure based on squared deviations can be developed in four steps, going from a very simple class of cases to a fully general one. The simplest cases are those of dual distributions in which both elements have the same weight and the average is 0.5, i.e. $p(H) = 0.5$. This is illustrated in Figure 1. In case (a), the whole mass of the distribution is on the probability value 0.5. This diagram can illustrate a distribution $\{\langle 1, p_1 \rangle\}$ such that $p_1(H) = 0.5$. In case (b), the second-order probability is equally divided between two equally probable options, namely that the coin has either probability 0.25 or 0.75 of yielding heads. This diagram can illustrate a distribution $\{\langle 0.5, p_1 \rangle, \langle 0.5, p_2 \rangle\}$ such that $p_1(H) = 0.25$ and $p_2(H) = 0.75$. Diagram (c), finally, illustrates a distribution of the form $\{\langle 0.5, p_1 \rangle, \langle 0.5, p_2 \rangle\}$ with $p_1(H) = 0$ and $p_2(H) = 1$.

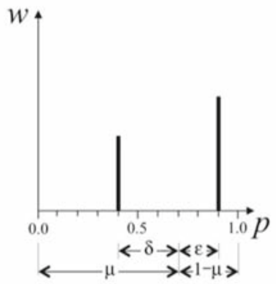

Figure 3.

Clearly, the dispersion of probability values increases as we go from (a) to (b) and from (b) to (c). The most obvious way to measure the distance from the mean is the ratio $\frac{\delta}{\mu}$ as shown in Figure 2. Here δ is the distance from a probability value to the mean $p(H)$. (Note that in the cases referred to in this first step, there are only two probability values, and due to the symmetry of the distribution they both have the same distance to the mean of the distribution.) We divide δ by the largest possible such distance, which (again due to the symmetry) is equal to μ, the mean of the distribution. The square of this ratio, $(\frac{\delta}{\mu})^2$, is the chosen measure. In Figure 1 its value is 0 in case (a), 0.25 in case (b), and 1 in case (c).

In the second step we turn to an arbitrary dual distribution. See Figure 3. Here we have two deviations, one to the left of $p(H)$ and the other to its right. The leftwards deviation is $\frac{\delta}{\mu}$ and the rightwards deviation is $\frac{\varepsilon}{1-\mu}$. A reasonable approach is to use the geometric mean of these two deviations, $\sqrt{\frac{\delta\varepsilon}{\mu(1-\mu)}}$. Its square is the proposed measure, to be applied to a dual distribution:

Definition 4.1. Let $\alpha = \{\langle w_1, p_1\rangle, \langle w_2, p_2\rangle\}$ be a dual second-order probability distribution with $p_1 < p_2$. The *mean square deviation* (MSD) of α is the measure $m_{\text{MSD}}(\,\uparrow \alpha)$ such that for all H, $m_{\text{MSD}}(H \uparrow \alpha) = \frac{\delta\varepsilon}{\mu(1-\mu)}$, where $\mu = w_1 p_1(H) + w_2 p_2(H)$, $\delta = \mu - p_1(H)$, and $\varepsilon = p_2(H) - \mu$.

The third step is to extend this measure to a particular class of non-dual distributions, namely those that are mixtures of two or more dual distributions with the same (overall) probability. The obvious solution is to use the weighted average of the mean square deviations of the component dual distributions. Hence, if $\alpha = k_1\alpha_1 \cup k_2\alpha_2$, where α_1 and α_2 are dual, and $p(H \uparrow \alpha_1) = p(H \uparrow \alpha_2)$, then we have:

$$m_{\text{MSD}}(H \uparrow \alpha) = k_1 \times m_{\text{MSD}}(H \uparrow \alpha_1) + k_2 \times m_{\text{MSD}}(H \uparrow \alpha_2).$$

The fourth and final step in the construction of this measure is to extend it to arbitrary second-order probability distributions. This can be done straightforwardly, due to the following two observations:

DEFINITION 4.2. Let α be a second-order probability distribution and H an event. Then a *paired decomposition* of α for H is a set $\{\langle k_1, \alpha_1\rangle, ...\langle k_n, \alpha_n\rangle\}$ such that $\alpha = k_1\alpha_1 \cup ... \cup k_n\alpha_n$ and $\sum_{m=1}^{n} k_m = 1$, and that for all m with $1 \leq m \leq n$: (i) $0 < k_m < 1$, (ii) α_m is an at most dual second-order probability distribution, and (iii) $p(H \uparrow \alpha_m) = p(H \uparrow \alpha)$.

OBSERVATION 4.3. *Let α be a finite second-order probability distribution and H an event. Then there is a finite paired decomposition of α for H.*

OBSERVATION 4.4. *Let α be a second-order probability distribution and H an event. Let $\pi' = \{\langle k'_1, \alpha'_1\rangle, ...\langle k'_s, \alpha'_s\rangle\}$ and $\pi'' = \{\langle k''_1, \alpha''_1\rangle, ...\langle k''_t, \alpha''_t\rangle\}$ be two paired decompositions of α. Then:*

$$\sum_{r=1}^{s} k'_r \times m_{\text{MSD}}(H \uparrow \alpha'_r) = \sum_{r=1}^{t} k''_t \times m_{\text{MSD}}(H \uparrow \alpha''_r)$$

Hence, each second-order probability distribution can be reconstructed as a mixture of a set of dual distributions with the same probability, and if there are several such reconstructions then it makes no difference which of these we use to calculate the weighted average of the mean square deviation. We can therefore define the mean square deviation of an arbitrary second-order probability distribution through the following extension of Definition 4.1.

DEFINITION 4.5. For any second-order probability distribution α and event H:

$$m_{\text{MSD}}(H \uparrow \alpha) = \sum_{r=1}^{s} k_r \times m_{\text{MSD}}(H \uparrow \alpha_r),$$

where $\{\langle k_1, \alpha_1\rangle, ...\langle k_s, \alpha_s\rangle\}$ is a paired decomposition of α for H.

The other major type of dispersion measure is the average absolute deviation. Consider a second-order probability distribution $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$ with the associated overall probability $p(H \uparrow \alpha)$. For each vector $\langle w_k, p_k\rangle$ in α, the absolute deviation from $p(H \uparrow \alpha)$ is $|p_k(H) - p(H \uparrow \alpha)|$, and the weight that it should be assigned in the calculation of the average is w_k. This could lead us to use

$$\sum_{k=1}^{n} w_k \times |p_k(H) - p(H \uparrow \alpha)|$$

as a measure. However, as can easily be verified, although its minimal value is 0, its maximal value is 0.5. By doubling it we obtain a measure that has 0 as its minimal and 1 as its maximal value, just like the MSD measure:

Definition 4.6. Let $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$. Then the *average absolute deviation* (AAD) of α is the measure $m_{\text{AAD}}(\ \uparrow \alpha)$ such that for all H:

$$m_{\text{AAD}}(H \uparrow \alpha) = 2 \times \sum_{k=1}^{n} w_k \times |p_k(H) - p(H \uparrow \alpha)|$$

The AAD measure can be axiomatically characterized as follows: (I will return shortly to the intuitive motivations of the least transparent of these postulates.)

Theorem 4.7. *An uncertainty measure m coincides with m_{AAD} if and only if it satisfies the postulates:*

Spread: *Let $\alpha = \{\langle w_1, p_1\rangle, \langle w_2, p_2\rangle\}$ and $\alpha' = \{\langle w_1, p'_1\rangle, \langle w_2, p'_2\rangle\}$ with $p(H \uparrow \alpha) = p(H \uparrow \alpha')$ and $p'_1(H) < p_1(H) \leq p_2(H) < p'_2(H)$, and let $0 < k \leq 1$. Let $\beta = k\alpha + (1-k)\gamma$ and $\beta' = k\alpha' + (1-k)\gamma$. Then $m(H \uparrow \beta') > m(H \uparrow \beta)$.*

Calibration: *The minimal value of m is 0 and its maximal value 1.*

Linear interpolation: *If $p(H \uparrow \alpha) = p(H \uparrow \beta)$ then*
$m(H \uparrow k\alpha \cup (1-k)\beta) = k \times m(H \uparrow \alpha) + (1-k) \times m(H \uparrow \beta)$.

Leverage: *If $0 \leq k \leq 1$ then*

$$m(H \uparrow \{\langle x, p(H \uparrow \alpha) + ky\rangle, ...\}) = m(H \uparrow \{\langle kx, p(H \uparrow \alpha) + y\rangle, \langle (1-k)x, p(H \uparrow \alpha)\rangle, ...\}$$

Translation: *If there is some c such that $p'_k(H) = p_k(H) + c$ for all $1 \leq k \leq n$, then*
$p(H \uparrow \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}) = p(H \uparrow \{\langle w_1, p'_1\rangle, ...\langle w_n, p'_n\rangle\})$.

Observation 4.8. *m_{AAD} satisfies:*

Independence of other probabilities: *If* $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$, $\beta = \{\langle w_1, p'_1\rangle, ...\langle w_n, p'_n\rangle\}$, *and* $p_k(H) = p'_k(H)$ *holds for all* k, *then* $m(H \uparrow \alpha) = m(H \uparrow \beta)$.

Certainty: *If* α *is unary, then* $m(H \uparrow \alpha) \leq m(H \uparrow \beta)$ *for all* β.

Negation symmetry: $m(H \uparrow \alpha) = m(\neg H \uparrow \alpha)$.

Spread ensures that uncertainty increases as probability mass is moved away from the centre of the distribution. *Leverage* ensures that the effect of such movements of probability mass away from the centre is proportionate to how far it is removed.

The MSD measure satisfies three of the five axioms that characterize the AAD measure; the two exceptions are both closely related to the linearity of the AAD measure.

OBSERVATION 4.9. (1) m_{MSD} *satisfies* Spread, Calibration, Linear interpolation, Independence of other probabilities, Certainty, *and* Negation symmetry.

(2) m_{MSD} *does not satisfy* Leverage *or* Translation.

The axiomatic characterization of the MSD measure is left as an open issue.

5. The robustness measure

As explained in Section 2, the robustness of an agent's estimate of an (objective) first-order probability can be measured by determining how much it is perturbed in a readjustment, i.e. a revision by information about whether the event actually took place. However, readjustments are not easily representable in a one-levelled model of probability. Clearly, standard conditionalization of probabilities cannot be used to represent them, for the simple reason that $p(H|H) = 1$. For repeatable event types, readjustments can be expressed in terms of probabilities of event sequences. Hence, in the coin example let H denote that the coin lands heads in the first throw and HH that it does so in both the first and the second throw. Then $p(H)$ is the initial and $p(HH|H)$ the readjusted estimate of the coin's propensity to yield heads. However, for non-repeatable events this representation cannot be used. A possible solution is to introduce a more limited form of conditionalization, that does not use all the information derivable from the fact that the event in question has occurred, but only those parts of this information that concern what the world was like before up to the actual occurrence of the event. For any proposition E and point in time t, let $[E]_t$ be the maximal proposition implied by E that contains all the information that E carries about particular facts before t and about (probabilistic and other) laws, but contains no information about events at t or later (other than what can be inferred from the facts before t and from the laws).

To further clarify the meaning of this notation, let A denote the statement that a certain die yielded six when it was thrown at a point in time t, namely eight o'clock yesterday morning. Let $p(A)$ denote our best estimate of the propensity for this to happen. Assuming that we have no specific information about the die in question, it can be assumed that $p(A) = 1/6$. Let B denote the statement that the die was tested with 99 throws yesterday evening and that 59 of them yielded a six. Suppose that A and B are both true and that they are all the information that we have about the die's propensities. Then, based on the information $A\&B$ our best estimate of what the objective chance of A was is 0.6. This can be written $p(A|[A\&B]_t) = 0.6$. Note that $[A\&B]_t$ differs from $A\&B$ in containing no information about events that happened at time t or later.[1] It does however contain information discovered after t about what the world was like before t, and this information has impact on our best estimate of what the chance was for A to happen. Hence, $p(A|[A\&B]_t) \neq p(A|A\&B) = 1$. It is important to observe that p is not an objective probability but a subjective estimate of an objective probability. (Otherwise the notation would not have a reasonable interpretation, since the restriction of informational content represented by $[\ ]_t$ makes sense in a condition on a subjective credence function, but not in a condition for an objective chance function.)

For our present purposes, the time index t of $[E]_t$ can be suppressed without loss of clarity.

In the case of the cardsharp's coin, $p(H|[H])$ is the readjusted estimate of the coin's propensity to land heads, after we learn that it actually did so. Similarly, $p(H|[\neg H])$ is the readjusted estimate of the coin's propensity to land heads, after we learn that it actually landed tails. (Both $[H]$ and $[\neg H]$ are admissible in David Lewis's sense. Cf. [20, 21, 7, 8].)

The difference between the two readjusted probabilities $p(H|[H])$ and $p(H|[\neg H])$ can be used as a measure of uncertainty. This will be called the robustness measure since it tells us how robust a probability estimate is against readjustments:

---

[1] Or more precisely: about events at time t or later that cannot be inferred with certainty from information about what the world was like before t.

DEFINITION 5.1. The *robustness measure* m_{R} associated with a probability function p is the measure such that for all H:

$$m_{\mathrm{R}}(H) = p(H|[H]) - p(H|[\neg H]).$$

It is important to note that the robustness measure can be applied without knowledge (or existence) of second-order probabilities. It is based on our probability estimates and how we revise them when we receive new information. The only addition to the standard apparatus of probability theory that we need in order to introduce this measure is the temporal restriction of information, as denoted by [ ]. However, it is nevertheless interesting to investigate how the robustness measure relates to second-order probability distributions in the cases when such distributions exist and our readjustments are based on them.

To see how this can be done, let us again assume that the cardsharp's coin has to belong to one of three types: it is either a tails dime with the (objective) probability 0 of yielding heads, a fair dime for which the corresponding probability is 0.5, or a heads dime for which it is 1. The pre-event second-order probabilities that the dime had these properties were 0.25, 0.5, and 0.25, respectively. Thus, we have a second-order probability distribution $\alpha = \{\langle 0.25, p_t\rangle, \langle 0.5, p_f\rangle, \langle 0.25, p_h\rangle\}$, where $p_t(H) = 0$, $p_f(H) = 0.5$, and $p_h(H) = 1$. Now suppose the coin is flipped and yields heads. This gives us reason to replace α with a revised second-order probability distribution, to be denoted by $\alpha * H$. If this revision is performed on the w-values (0.25, 0.50, and 0.25) with standard probabilistic conditionalization, then we have $\alpha * H = \{\langle 0.5, p_f\rangle, \langle 0.5, p_h\rangle\}$, and consequently $p(H|[H]) = p(H \uparrow \alpha * H) = 0.75$. Similarly $p(H|[\neg H]) = p(H \uparrow \alpha * \neg H) = 0.25$. This yields an m_{R}-value based on α, namely

$$m_{\mathrm{R}}(H \uparrow \alpha) = p(H|[H]) - p(H|[\neg H]) = p(H \uparrow \alpha * H) - p(H \uparrow \alpha * \neg H) = 0.5.$$

This procedure is generalized in the following definitions:

DEFINITION 5.2. Let $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$. Then

$$\alpha * H = \left\{\left\langle \frac{w_1 \times p_1(H)}{\sum_{k=1}^{n}(w_k \times p_k(H))}, p_1\right\rangle, ... \left\langle \frac{w_n \times p_n(H)}{\sum_{k=1}^{n}(w_k \times p_k(H))}, p_n\right\rangle\right\}$$

DEFINITION 5.3. The robustness measure of uncertainty m_{R} is based on a second-order probability distribution α if and only if it is the case that for all H, $p(H) = p(H \uparrow \alpha)$ and $p(H|[H]) = p(H \uparrow \alpha * H)$.

Observation 5.4. *Let m_{R} be based on a second-order probability distribution $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$. Then*

$$m_{\mathrm{R}}(H \uparrow \alpha) =$$

$$\left(\frac{1}{p(H)}\sum_{k=1}^{n} w_k \times (p_k(H))^2\right) + \left(\frac{1}{1-p(H)}\sum_{k=1}^{n} w_k \times (1-p_k(H))^2\right) - 1$$

Finally, the following theorem connects our two major approaches to uncertainty measurement. If we base our readjustments on conditionalization of second-order probabilities, then m_{R} and m_{MSD} coincide.

Theorem 5.5. *If a robustness measure of uncertainty m_{R} is based on a second-order probability distribution, then it coincides with the MSD measure m_{MSD} that is based on the same distribution.*

Appendix: Proofs

The following simplifying notation will be used in the proofs:

Definition 5.6. If $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$, then
$\alpha/H = \{\langle w_1, p_1(H)\rangle, ...\langle w_n, p_n(H)\rangle\}$.

Proof of Observation 4.3: Let $\alpha/H = \{\langle x_1, y_1\rangle, ...\langle x_n, y_n\rangle\}$ with $n > 2$ and $y_1 < y_2 ... < y_n$. We are going to show that α can be decomposed into two distributions such that one of them is dual and the other has fewer elements than α. Since α is finite the desired paired decomposition can be obtained through repetition of this procedure a finite number of times. There are two cases:

Case i, $p\left(\left\{\left\langle \frac{x_1}{x_1+x_n}, y_1\right\rangle, \left\langle \frac{x_n}{x_1+x_n}, y_n\right\rangle\right\}\right) \geq p(H \uparrow \alpha)$: Then there is some z such that $0 \leq z \leq x_n$ and $p\left(\left\{\left\langle \frac{x_1}{x_1+z}, y_1\right\rangle, \left\langle \frac{z}{x_1+z}, y_n\right\rangle\right\}\right) = p(H \uparrow \alpha)$. We then have:

$\alpha = \{\langle x_1, y_1\rangle, \langle z, y_n\rangle\} \cup \{\langle x_2, y_2\rangle, ...\langle x_{n-1}, y_{n-1}\rangle, \langle x_n - z, y_n\rangle\} =$
$(x_1 + z)\{\langle \frac{x_1}{x_1+z}, y_1\rangle, \langle \frac{z}{x_1+z}, y_n\rangle\} \cup$
$\cup (1 - x_1 - z)\{\langle \frac{x_2}{1-x_1-z}, y_2\rangle, ...\langle \frac{x_{n-1}}{1-x_1-z}, y_{n-1}\rangle, \langle \frac{x_n - z}{1-x_1-z}, y_n\rangle\}$

Here, $p\left(\{\langle \frac{x_1}{x_1+z}, y_1\rangle, \langle \frac{z}{x_1+z}, y_n\rangle\}\right) =$
$= p\left(\{\langle \frac{x_2}{1-x_1-z}, y_2\rangle, ...\langle \frac{x_{n-1}}{1-x_1-z}, y_{n-1}\rangle, \langle \frac{x_n - z}{1-x_1-z}, y_n\rangle\}\right),$

and the former distribution is dual and the second has one element less than α.

Case ii, $p\left(\left\{\left\langle\frac{x_1}{x_1+x_n}, y_1\right\rangle, \left\langle\frac{x_n}{x_1+x_n}, y_n\right\rangle\right\}\right) < p(H \uparrow \alpha)$: The proof is analogous. Here, $\langle x_1, y_1\rangle$ should be decomposed instead of $\langle x_n, y_n\rangle$.

Proof of Observation 4.4: Let $\alpha \equiv \alpha'$ and $\alpha' = \{\langle w_1, p_1\rangle, \langle w_2, p_2\rangle, \ldots \langle w_n, p_n\rangle\}$, where $\langle w_1, p_1\rangle$ and $\langle w_2, p_2\rangle$ are the two elements that give rise to the first part of the decomposition π' of α, etc. Then $k_1' = w_1 + w_2$ and $\alpha_1' = \left\{\left\langle\frac{w_1}{w_1+w_2}, p_1(H)\right\rangle, \left\langle\frac{w_2}{w_1+w_2}, p_2(H)\right\rangle\right\}$. Letting $\mu = \frac{w_1 p_1(H)+w_2 p_2(H)}{w_1+w_2}$, $\delta = \mu - p_1(H)$, and $\varepsilon = p_2(H) - \mu$ we then have:

$m_{\text{MSD}}(H \uparrow \alpha_1') = \frac{\delta\varepsilon}{\mu(1-\mu)} = \frac{\delta\varepsilon^2+\varepsilon\delta^2}{\mu(1-\mu)(\delta+\varepsilon)}$
$= \left(\frac{\varepsilon}{\delta+\varepsilon}\right)\left(\frac{(\mu-\delta)^2}{\mu} + \frac{(1-(\mu-\delta))^2}{1-\mu} - 1\right) + \left(\frac{\delta}{\delta+\varepsilon}\right)\left(\frac{(\mu+\varepsilon)^2}{\mu} + \frac{(1-(\mu+\varepsilon))^2}{1-\mu} - 1\right)$

Using the above definitions of μ, δ, and ε, we can derive:
$\frac{\varepsilon}{\delta+\varepsilon} = \frac{p_2(H)-\mu}{p_2(H)-p_1(H)} = \frac{p_2(H)-\frac{w_1p_1(H)+w_2p_2(H)}{w_1+w_2}}{p_2(H)-p_1(H)} = \frac{w_1}{w_1+w_2}$ and similarly $\frac{\delta}{\delta+\varepsilon} = \frac{w_2}{w_1+w_2}$. Inserting this we obtain $m_{\text{MSD}}(H \uparrow \alpha_1') =$
$\left(\frac{w_1}{w_1+w_2}\right)\left(\frac{p_1(H)^2}{\mu} + \frac{(1-p_1(H))^2}{1-\mu} - 1\right) + \left(\frac{w_2}{w_1+w_2}\right)\left(\frac{p_2(H)^2}{\mu} + \frac{(1-p_2(H))^2}{1-\mu} - 1\right)$.

The contribution that this first element of the paired decomposition makes to $m_{\text{MSD}}(H \uparrow \alpha')$ is $k_1' \times m_{\text{MSD}}(H \uparrow \alpha_1') =$
$w_1\left(\frac{p_1(H)^2}{\mu} + \frac{(1-p_1(H))^2}{1-\mu} - 1\right) + w_2\left(\frac{p_2(H)^2}{\mu} + \frac{(1-p_2(H))^2}{1-\mu} - 1\right)$,
and it follows from the weighted average construction that
$m_{\text{MSD}}(H \uparrow \alpha') = \sum_{\langle w_k, p_k\rangle \in \alpha'} w_k\left(\frac{p_k(H)^2}{\mu} + \frac{(1-p_k(H))^2}{1-\mu} - 1\right)$
$= \sum_{\langle w_k, p_k\rangle \in \alpha} w_k\left(\frac{p_k(H)^2}{\mu} + \frac{(1-p_k(H))^2}{1-\mu} - 1\right)$.

The same result can be obtained from the other paired decompositions of α.

Proof of Theorem 4.7: *From construction to postulates*: This direction of the proof is straight-forward with the possible exception of the maximal value part of *Calibration*. For that part, note that since $m_{\text{AAD}}(H \uparrow \alpha)$ increases as probability mass is removed from the average, the maximum of $m_{\text{AAD}}(H \uparrow \alpha)$ must be for some α such that $\alpha/H = \{\langle x, 1\rangle, \langle 1-x, 0\rangle\}$. Then $p(H \uparrow \alpha) = x$, and $m_{\text{AAD}}(H \uparrow \alpha) = 2x|1-x| + 2(1-x)|x-0| = 4(x-x^2)$, which has its maximal value $= 1$ when $x = 0.5$.

From postulates to construction: (In this part of the proof we will use *Certainty*, that follows from two of the other postulates, see Observation 4.8.)

It follows from Observation 4.3 that α is a linear combination of at most dual distributions all of which assign the same probability to H. Therefore, for this part of the theorem it is sufficient to show that if m satisfies the postulates, then:

(1) $m(H \uparrow \alpha) = m_{\text{AAD}}(H \uparrow \alpha)$ if α is unary,

(2) $m(H \uparrow \alpha) = m_{\text{AAD}}(H \uparrow \alpha)$ if α is dual, and

(3) $m(H \uparrow \alpha) = m_{\text{AAD}}(H \uparrow \alpha)$ if α is a linear combination of at most dual distributions that all assign the same probability to H.

(1) follows directly from *Calibration* and *Certainty*, and (3) from *Linear interpolation* since m_{AAD} also satisfies this postulate. It remains to prove (2). This proof has two steps. In the first step we are going to show that $m(\{\langle 0.5, 0\rangle, \langle 0.5, 1\rangle\}) = 1$.

First step: We use *Independence of other probabilities* and introduce the notation of Definition 5.6. We know from *Spread* that the maximal value of $m(H \uparrow \alpha)$ must be for some α such that $\alpha/H = \{\langle 1-x, 0\rangle, \langle x, 1\rangle\}$ for some x. We are going to show that $m(\{\langle 1-x, 0\rangle, \langle x, 1\rangle)$ is maximal for $x = 0.5$. It is convenient to divide the proof into two cases, $x \geq 0.5$ and $x \leq 0.5$. In the former case we have:

$m(\{\langle 1-x, 0\rangle, \langle x, 1\rangle\})$

$= 2 \times m\left(\left\{\left\langle \frac{1-x}{2}, 0\right\rangle, \left\langle 0.5, x\right\rangle, \left\langle \frac{x}{2}, 1\right\rangle\right\}\right)$

(*Linear interpolation* and $m(\{\langle 1, x\rangle\}) = 0$ that follows from *Certainty* and *Calibration*)

$= 2 \times m\left(\left\{\left\langle \frac{x}{2}, 2x-1\right\rangle, \left\langle 1-x, x\right\rangle, \left\langle \frac{x}{2}, 1\right\rangle\right\}\right)$ (*Leverage*)

$= 2 \times m\left(\left\{\left\langle \frac{x}{2}, x-0.5\right\rangle, \left\langle 1-x, 0.5\right\rangle, \left\langle \frac{x}{2}, 1.5-x\right\rangle\right\}\right)$ (*Translation*)

$= 2 \times m\left(\left\{\left\langle x-x^2, 0\right\rangle, \left\langle 1-2x+2x^2, 0.5\right\rangle, \left\langle x-x^2, 1\right\rangle\right\}\right)$ (*Leverage*)

$= 2 \times m\left(\left\{\left\langle x-x^2, 0\right\rangle, \left\langle 1-2x+2x^2-0.5, 0.5\right\rangle, \left\langle 0.5, 0.5\right\rangle, \left\langle x-x^2, 1\right\rangle\right\}\right)$

(since $1-2x+2x^2 \geq 0.5$ when $x \geq 0.5$)

$= m\left(\left\{\left\langle 2x-2x^2, 0\right\rangle, \left\langle 1-4x+4x^2, 0.5\right\rangle, \left\langle 2x-2x^2, 1\right\rangle\right\}\right)$

(*Linear interpolation* and $m(\{\langle 1, 0.5\rangle\}) = 0$ that follows from *Certainty* and *Calibration*)

Due to *Spread* this expression is maximal when $1-4x+4x^2$ is minimal, i.e. when $x = 0.5$.

The proof for the other case, $x \leq 0.5$, is similar. Hence $m(H \uparrow \alpha)$ has maximal value when $\alpha/H = \{\langle 0.5, 0\rangle, \langle 0.5, 1\rangle\}$. It follows from *Calibration* that $m(H \uparrow \alpha) = 1$.

Second step: Let $\alpha/H = \{\langle 1-x, a\rangle, \langle x, b\rangle\}$. Without loss of generality we assume that $a < b$. It is convenient to divide the proof into two cases, $x \geq 0.5$ and $x \leq 0.5$. We prove here the case when $x \geq 0.5$. We have $p(H \uparrow \alpha) = a - ax + bx$. For a simplifying notation, let $s = a - ax + bx$. We then have:

$m(\{\langle 1-x, a\rangle, \langle x, b\rangle\}) =$

$= 2m\left(\left\{\left\langle \frac{1-x}{2}, a\right\rangle, \langle 0.5, s\rangle, \left\langle \frac{x}{2}, b\right\rangle\right\}\right)$

(*Linear interpolation* and $m(\{\langle 1, s\rangle\}) = 0$ that follows from *Certainty* and *Calibration*)

$= 2m\left(\left\{\left\langle \frac{(1-x)(s-a)}{2(b-s)}, 2s-b\right\rangle, \left\langle 1 - \frac{x}{2} - \frac{(1-x)(s-a)}{2(b-s)}, s\right\rangle, \left\langle \frac{x}{2}, b\right\rangle\right\}\right)$

(*Leverage*)

$= 2m\left(\left\{\left\langle \frac{(1-x)(s-a)}{2(b-s)}, s-b+0.5\right\rangle, \left\langle 1 - \frac{x}{2} - \frac{(1-x)(s-a)}{2(b-s)}, 0.5\right\rangle, \left\langle \frac{x}{2}, b-s+0.5\right\rangle\right\}\right)$

(*Translation*)

$= 2m\left(\left\{\left\langle \frac{(1-x)(s-a)(b-s)}{2(b-s)\times 0.5}, 0\right\rangle, \left\langle 1 - \frac{(1-x)(s-a)(b-s)}{2(b-s)\times 0.5} - \frac{x(b-s)}{2\times 0.5}, 0.5\right\rangle, \left\langle \frac{x(b-s)}{2\times 0.5}, 1\right\rangle\right\}\right)$

(*Leverage*)

$= 2m\{\langle x(b-s), 0\rangle, \langle 1-2x(b-s), 0.5\rangle, \langle x(b-s), 1\rangle\})$

(since $\frac{(1-x)(s-a)(b-s)}{2(b-s)\times 0.5} = x(b-s)$ due to the definition of s)

$= \frac{2x(b-s)}{0.5} \times m(\{\langle 0.5, 0\rangle, \langle 0.5, 1\rangle\})$

(*Linear interpolation* and *Certainty* and *Calibration* as above)

$= 4x(b-s)$ (since $m(\{\langle 0.5, 0\rangle, \langle 0.5, 1\rangle\} = 1$ according to the first step)

$4(b-a)(x-x^2)$ (since $s = a - ax + bx$)

$= 2((1-x)(a-ax+bx-a) + x(b-a+ax-bx))$

$= m_{\text{AAD}}(H \uparrow \alpha)$

The other case, $x \leq 0.5$, is similar.

Proof of Observation 4.8: *Independence of other probabilities* follows directly from *Translation*, let $c = 0$.

Certainty follows from *Spread* and *Translation*: Let α be a unary distribution. If β is also unary then $m(H \uparrow \alpha) = m(H \uparrow \beta)$ follows from *Translation.* If β is non-unary, then it follows from *Spread* (letting $p_1 = p_2$) and Observation 4.3 that $m(H \uparrow \beta) > m(H \uparrow \{\langle 1, p(H \uparrow \beta)\rangle\})$. Due to *Translation*, $m(H \uparrow \{\langle 1, p(H \uparrow \beta)\rangle\} = m(H \uparrow \alpha)$, thus $m(H \uparrow \beta) > m(H \uparrow \alpha)$.

For *Negation symmetry*, let $\alpha/H = \{\langle x_1, y_1\rangle, ...\langle x_n, y_n\rangle\}$. Then $\alpha/\neg H = \{\langle x_1, 1-y_1\rangle, ...\langle x_n, 1-y_n\rangle\}$, and we have:

$m_{\text{AAD}}(H \uparrow \alpha) = 2 \times \sum_{k=1}^{n} x_k \times |p_k(H) - p(H \uparrow \alpha)|$
$= 2 \times \sum_{k=1}^{n} x_k \times |(1 - p_k(H)) - (1 - p(H \uparrow \alpha))|$
$= 2 \times \sum_{k=1}^{n} x_k \times |p_k(\neg H) - p(\neg H \uparrow \alpha)| = m_{\text{AAD}}(\neg H \uparrow \alpha)$

Proof of Observation 4.9: Part 1: *Linear interpolation*, *Independence of other probabilities*, *Certainty*, and *Negation symmetry* follow directly from the construction. For *Spread*, note that since *Linear interpolation* holds, it is sufficient to show that *Spread* holds for dual distributions. For *Calibration*, use the case $\delta = \varepsilon = \mu = 0.5$.

Part 2: To see that *Leverage* does not hold, it is sufficient to note that the identity

$$m_{\text{MSD}}(H \uparrow \{\langle x, p(H \uparrow \alpha) + ky\rangle, ...\})$$
$$= m_{\text{MSD}}(H \uparrow \{\langle kx, p(H \uparrow \alpha) + y\rangle, \langle (1-k)x, p(H \uparrow \alpha)\rangle, ...\})$$

does not hold when $x = y = p(H \uparrow \alpha) = 0.5$ and $k = 0.2$. That *Translation* does not hold follows from

$$m_{\text{MSD}}(\{\langle 0.5, 0.0\rangle, \langle 0.5, 0.5\rangle\}) \neq m_{\text{MSD}}(\{\langle 0.5, 0.25\rangle, \langle 0.5, 0.75\rangle\}).$$

Proof of Observation 5.4: Let $\alpha = \{\langle w_1, p_1\rangle, ...\langle w_n, p_n\rangle\}$. Then according to Definition 5.2 $\alpha * H = \left\{\left\langle \frac{w_1 \times p_1(H)}{\sum_{k=1}^{n}(w_k \times p_k(H))}, p_1\right\rangle, ... \left\langle \frac{w_n \times p_n(H)}{\sum_{k=1}^{n}(w_k \times p_k(H))}, p_n\right\rangle\right\}$, from which it follows that $p(H|[H]) = p(H \uparrow \alpha * H) = \sum_{k=1}^{n} \frac{w_k \times p_k(H) \times p_k(H)}{\sum_{k=1}^{n} w_k \times p_k(H))}$ $= \frac{1}{p(H)} \sum_{k=1}^{n} w_k \times (p_k(H))^2$. Due to Definitions 5.1 and 5.3 we then have $m_{\text{R}}(H \uparrow \alpha) = p(H|[H]) - p(H|[\neg H]) = p(H|[H]) + p(\neg H|[\neg H]) - 1 =$ $\left(\frac{1}{p(H)} \sum_{k=1}^{n} w_k \times (p_k(H))^2\right) + \left(\frac{1}{1-p(H)} \sum_{k=1}^{n} w_k \times (1 - p_k(H))^2\right) - 1.$

Lemma 1. If a robustness measure of uncertainty m_{R} is based on a second-order probability distribution, then it satisfies *Linear interpolation.*

Proof of Lemma 1: Let $\alpha/H = \{\langle w_1, x_1\rangle, ...\langle w_s, x_s\rangle\}$ and furthermore let $\beta/H = \{\langle v_1, y_1\rangle, ...\langle v_t, y_t\rangle\}$, and $p(H \uparrow \alpha) = p(H \uparrow \beta)$. Let $0 \leq k \leq 1$. We are going to show that $m_{\text{R}}(H \uparrow k\alpha \cup (1-k)\beta) = k \times m_{\text{R}}(H \uparrow \alpha) +$ $(1-k)m_{\text{R}}(H \uparrow k\beta)$. Due to Observation 5.4 we have $m_{\text{R}}(H \uparrow \alpha) =$ $\left(\sum_{r=1}^{s} w_r \times \left(\frac{(x_r)^2}{p(H)} + \frac{(1-x_r)^2}{1-p(H)}\right)\right) - 1$ and in the same way $m_{\text{R}}(H \uparrow \beta) =$ $\left(\sum_{r=1}^{t} v_r \times \left(\frac{(y_r)^2}{p(H)} + \frac{(1-y_r)^2}{1-p(H)}\right)\right) - 1$. From this and $(k\alpha \cup (1-k)\beta)/H =$

$\{\langle kw_1, x_1\rangle, ...\langle kw_s, x_s\rangle, \langle(1-k)v_1, y_1\rangle, ...\langle(1-k)v_t, y_t\rangle\}$ it follows that:

$$m_{\mathrm{R}}(H \uparrow k\alpha \cup (1-k)\beta))$$
$$= \left(\sum_{r=1}^{s} k \times w_r \times \left(\frac{(x_r)^2}{p(H)} + \frac{(1-x_r)^2}{1-p(H)}\right)\right)$$
$$+ \left(\sum_{r=1}^{t} (1-k) \times v_r \times \left(\frac{(y_r)^2}{p(H)} + \frac{(1-y_r)^2}{1-p(H)}\right)\right) - 1$$
$$= k \times (m_{\mathrm{R}}(H \uparrow \alpha) + 1) + (1-k) \times (m_{\mathrm{R}}(H \uparrow \beta) + 1) - 1$$
$$= k \times m_{\mathrm{R}}(H \uparrow \alpha) + (1-k) \times m_{\mathrm{R}}(H \uparrow \beta).$$

Proof of Theorem 5.5: Due to Observation 4.3, Lemma 1, and the *Linear interpolation* property of the MSD measure (Observation 4.9), it is sufficient to show that the theorem holds for dual second-order probability distributions. For this purpose, let $\mu = p(H)$ and let $\alpha/H = \{\langle w_1, \mu-\delta\rangle, \langle w_2, \mu+\varepsilon\rangle\}$. Since $w_1 + w_2 = 1$ we then have $\alpha/H = \{\langle \frac{\varepsilon}{\delta+\varepsilon}, \mu-\delta\rangle, \langle \frac{\delta}{\delta+\varepsilon}, \mu+\varepsilon\rangle\}$. It follows from Observation 5.4 that:

$$m_{\mathrm{R}}(H \uparrow \alpha) =$$
$$= \frac{1}{\mu} \times \left(\frac{\varepsilon(\mu-\delta)^2}{\delta+\varepsilon} + \frac{\delta(\mu+\varepsilon)^2}{\delta+\varepsilon}\right) + \frac{1}{1-\mu} \times \left(\frac{\varepsilon(1-(\mu-\delta))^2}{\delta+\varepsilon} + \frac{\delta(1-(\mu+\varepsilon))^2}{\delta+\varepsilon}\right) - 1$$
$$= \frac{(1-\mu)(\varepsilon(\mu-\delta)^2+\delta(\mu+\varepsilon)^2)+\mu(\varepsilon(1-(\mu-\delta))^2+\delta(1-(\mu+\varepsilon))^2)-\mu(1-\mu)(\delta+\varepsilon)}{\mu(1-\mu)(\delta+\varepsilon)}$$
$$= \frac{\delta\varepsilon^2+\varepsilon\delta^2}{\mu(1-\mu)(\delta+\varepsilon)} = \frac{\delta\varepsilon}{\mu(1-\mu)}.$$

Acknowledgement. I would like to thank Professor Alan Hájek for valuable comments on an earlier version of this paper.

References

[1] Baron, J., 'Second-order probabilities and belief functions', *Theory and Decision* 23: 25–36, 1987.

[2] Carnap, R., 'The Two Concepts of Probability', *Philosophy and Phenomenological Research* 5: 513–532, 1945.

[3] Ellsberg, D., 'Risk, ambiguity, and the Savage axioms' [1961], in P. Gärdenfors and N.-E. Sahlin (eds.), *Decision, probability, and utility*, Cambridge: Cambridge University Press, 1988, pp. 245–269.

[4] Fischhoff, B., 'Perceived Informativeness of Facts', *Human Perception and Performance* 3(2): 349–358, 1977.

[5] Gärdenfors, P., 'Forecasts, decisions and uncertain probabilities', *Erkenntnis* 14: 159–181, 1979.

[6] Gärdenfors, P., and N.-E. Sahlin, 'Unreliable probabilities, risk taking, and decision making' [1982], in P. Gärdenfors and N.-E. Sahlin (eds.), *Decision, probability, and utility*, Cambridge: Cambridge University Press, 1988, pp. 313–334.

[7] Hall, N., 'Correcting the Guide to Objective Chance', *Mind* 103: 505–517, 1994.

[8] Hall, N., 'Two Mistakes About Credence and Chance', *Australasian Journal of Philosophy* 82: 93–111, 2004.

[9] Hansson, S. O., 'Ten Philosophical Problems in Belief Revision', *Journal of Logic and Computation*, 13: 37–49, 2003.

[10] Hansson, S. O. 'Levi's ideals', in E. J. Olsson (ed.), *Knowledge and Inquiry. Essays on the Pragmatism of Isaac Levi*, Cambridge: Cambridge University Press, 2006, pp. 241–247.

[11] Hansson, S. O., 'Do we need second-order probabilities?', *Dialectica* 62: 525–533, 2008.

[12] Kelman, M., D. E. Fallas, and H. Folger, 'Decomposing Hindsight Bias', *Journal of Risk and Uncertainty* 16: 251–269, 1998.

[13] Knight, F. H., *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin [1921], 1935. (http://www.econlib.org/library/Knight/knRUP.html.)

[14] Levi, I., *Gambling with truth*, Cambridge, Mass.: M.I.T. Press 1973.

[15] Levi, I., *The enterprise of knowledge*, Cambridge, Mass.: M.I.T. Press 1980.

[16] Levi, I., 'Ignorance, probability and rational choice', *Synthese* 53: 387–417, 1982.

[17] Levi, I., 'Imprecision and indeterminacy in probability judgment', *Philosophy of Science* 52: 390–409, 1985.

[18] Levi, I., *Hard Choices: Decision Making under Unresolved Conflict*, Cambridge: Cambridge University Press, 1986.

[19] Levi, I., 'Reply to Maher', *Economics and Philosophy* 5: 79–90, 1989.

[20] Lewis, D., 'A Subjectivist's Guide to Objective Chance', in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol II, Berkeley: University of California Press, 1980, pp. 263–293.

[21] Lewis, D., 'Humean Supervenience Debugged', *Mind* 103: 473–490, 1994.

[22] Maher, P., 'Levi on the Allais and Ellsberg paradoxes', *Economics and Philosophy* 5: 69–78, 1989.

[23] Markowitz, H. M., *Portfolio Selection*, New York: John Wiley and Sons, 1959.

[24] Sahlin, N. E., 'On second order probability and the notion of epistemic risk', in B. P. Stigum and F. Wenstøp (eds.), *Foundations of Utility and Risk Theory with Applications*, Dordrecht: Reidel, 1983, pp. 95–104.

[25] Sahlin, N. E., 'Three decision rules for generalized probability representation', *The Behavioral and Brain Sciences* 8: 751–753, 1985.

[26] Skyrms, B., 'Higher order degrees of belief' in D. H. Mellor (ed.), *Prospects for Pragmatism*, Cambridge: Cambridge University Press, 1980, pp. 109–137.

Sven Ove Hansson
Department of Philosophy and the History of Technology
Royal Institute of Technology
100 44 Stockholm, Sweden
soh@kth.se