Evidence with Uncertain Likelihoods*

Joseph Y. Halpern
Cornell University
Ithaca, NY 14853 USA
halpern@cs.cornell.edu

Riccardo Pucella
Northeastern University
Boston, MA 02115 USA
riccardo@ccs.neu.edu

Abstract

An agent often has a number of hypotheses, and must choose among them based on observations, or outcomes of experiments. Each of these observations can be viewed as providing *evidence* for or against various hypotheses. All the attempts to formalize this intuition up to now have assumed that associated with each hypothesis h there is a *likelihood function* μ_h, which is a probability measure that intuitively describes how likely each observation is, conditional on h being the correct hypothesis. We consider an extension of this framework where there is uncertainty as to which of a number of likelihood functions is appropriate, and discuss how one formal approach to defining evidence, which views evidence as a function from priors to posteriors, can be generalized to accommodate this uncertainty.

1 Introduction

An agent often has a number of hypotheses, and must choose among them based on observations, or outcomes of experiments. Each of these observations can be viewed as providing *evidence* for or against various hypotheses. The following simple example illustrates the situation.

Example 1.1 Suppose that Alice and Bob each have a coin. Alice's coin is double-headed, Bob's coin is fair. Charlie knows all of this. Alice and Bob give their coin to some third party, Zoe, who chooses one of the coins, and tosses it. Charlie is not privy to Zoe's choice, but gets to see the outcome of the toss. Charlie is interested in two events (which are called hypotheses in this context):

> A: the coin is Alice's coin
>
> B: the coin is Bob's coin.

Now Charlie observes the coin land heads. What can he say about the probability of the events A and B? If Charlie has no prior probability on A and B, then he can draw no conclusions about their posterior probability; the probability of A could be any number in $[0, 1]$. The same remains true if the coin lands heads 100 times in a row. ∎

*A preliminary version of this paper appeared in the *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 243–250, 2005. Most of this work was done while the second author was at Cornell University.

Clearly Charlie learns something from seeing 100 (or even one) coin toss land heads. This has traditionally been modeled in terms of evidence: the more times Charlie sees heads, the more evidence he has for the coin being heads. There have been a number of ways of modeling evidence in the literature; see [Kyburg 1983] for an overview. All of them make use of the *likelihood function*. More precisely, they assume that for each hypothesis h of interest, there is a probability μ_h (called a likelihood function) on the space of possible observations. In the example above, if the coin is tossed once, the two possible observations are *heads* and *tails*. Clearly $\mu_A(heads) = 1/2$ and $\mu_B(heads) = 1$. If the coin is tossed 100 times, then there are 2^{100} possible observations (sequences of coin tosses). Again, μ_A and μ_B put obvious probabilities on this space. In particular, if *100heads* is the observation of seeing 100 heads in a row, then $\mu_A(\mathit{100heads}) = 1/2^{100}$ and $\mu_B(\mathit{100heads}) = 1$. Most of the approaches compute the relative weight of evidence of a particular observation ob for two hypotheses A and B by comparing $\mu_A(ob)$ and $\mu_B(ob)$.

Our goal in this paper is to understand what can be done when an hypothesis h does not determine a unique probability μ_h. To understand the issues that arise, consider the following somewhat contrived variant of Example 1.1.

Example 1.2 Suppose that Alice has two coins, one that is double-headed and one that is biased 3/4 towards heads, and chooses which one to give Zoe. Again, Zoe chooses either Alice's coin or Bob's coin and tosses it. Charlie, who knows the whole setup, sees the coin land heads. What does this tell him about the likelihood that the coin tossed was Alice's? ∎

The problem is that now we do not have a probability μ_A on observations corresponding to the coin being Alice's coin, since Charlie does not know if Alice's coin is double-headed or biased 3/4 towards heads. It seems that there is an obvious solution to this problem. We simply split the hypothesis "the coin is Alice's coin" into two hypotheses:

A_1: the coin is Alice's coin and it is double-headed

A_2: the coin is Alice's coin and it is the biased coin.

Now we can certainly apply standard techniques for computing evidence to the three hypotheses A_1, A_2, and B. The question now is what do the answers tell us about the evidence in favor of the coin being Alice's coin?

Situations like that in Example 1.2 arise frequently. For example, Epstein and Schneider [Epstein and Schneider 2005] show how multiple likelihoods can arise in investment decisions in the stock market, and the impact they can have on hedging strategies.[1] For example, consider a robot equipped with an unreliable sensor for navigation. This sensor returns the distance to the wall in front of the robot, with some known error. For simplicity, suppose that distances are measured in integral units $0, 1, 2, \dots$, and that if the wall is at distance m, then the sensor will return a reading of $m-1$ with probability 1/4, a reading of m with probability 1/2, and a reading of $m+1$ with probability 1/4. Suppose the robot wants to

[1] Epstein and Schneider present a general model of decision making in the presence of multiple likelihoods, although they do not attempt to quantify the evidence provided by observations in the presence of multiple likelihoods.

stop if it is exactly close to the wall, where "close" is interpreted as being within 3 units of the wall, and go forward if it is farther than 3 units. So again, we have two hypotheses of interest. However, while for each specific distance m we have a probability μ_m on sensor readings, we do not have a probability on sensor readings corresponding to the hypothesis *far*: "the robot is farther than 3 from the wall". While standard techniques will certainly give us the weight of evidence of a particular sensor reading for the hypothesis "the robot is distance m from the wall", it is not clear what the weight of evidence should be for the hypothesis *far*.

To examine the problem carefully, we consider one particular approach for determining the weight of evidence, due to Shafer [1982], which is a generalization of a method advocated by Good [1950]. The idea is to assign to every observation and hypothesis a number between 0 and 1—the weight of evidence for the hypothesis provided by the observation—that represents how much the observation supports the hypothesis. The closer a weight is to 1, the more the observation supports the hypothesis. This weight of evidence is computed using the likelihood functions described earlier. This way of computing the weight of evidence has several good properties, and is related to Shafer's theory of belief functions [Shafer 1976]; for instance, the theory gives a way to combine the weight of evidence from independent observations. We give full details in Section 2. For now, we illustrate how the problems described above manifest themselves in Shafer's setting.

Let an evidence space $\mathcal{E}$ consist of a set $\mathcal{H}$ of possible hypotheses, a set $\mathcal{O}$ of observations, and a probability μ_h on observations for each $h \in \mathcal{H}$. We take the weight of evidence for hypothesis h provided by observation ob in evidence space $\mathcal{E}$, denoted $w_{\mathcal{E}}(ob, h)$, to be

$$w_{\mathcal{E}}(ob, h) = \frac{\mu_h(ob)}{\sum_{h' \in \mathcal{H}} \mu_{h'}(ob)}.$$

It is easy to see that $w_{\mathcal{E}}(ob, \cdot)$ acts like a probability on $\mathcal{H}$, in that $\sum_{h \in \mathcal{H}} w_{\mathcal{E}}(ob, h) = 1$. With this definition, it is easy to compute the weight of evidence for Alice's coin when Charlie sees heads in Example 1.1 is $2/3$, and the weight of evidence when Charlie sees 100 heads is $2^{100}/(2^{100}+1)$. As expected, the more often Charlie sees heads, the more evidence he has in favor of the coin being double-headed (provided that he does not see tails).

In Example 1.2, if we consider the three hypotheses A_1, A_2, and B, then the weight of evidence for A_1 when Charlie sees heads is $1/(1 + 3/4 + 1/2) = 4/9$; similarly, the weight of evidence for A_2 is $1/3$ and the weight of evidence for B is $2/9$. Since weight of evidence acts like a probability, it might then seem reasonable to take the weight of evidence for A (the coin used was Alice's coin) to be $4/9 + 1/3 = 7/9$. (Indeed, this approach was implicitly suggested in our earlier paper [Halpern and Pucella 2003a].) But is this reasonable? A first hint that it might not be is the observation that the weight of evidence for A is higher in this case than it is in the case where Alice certainly had a double-headed coin.

To analyze this issue, we need an independent way of understanding what evidence is telling us. As observed by Halpern and Fagin [1992], weight of evidence can be viewed as a function from priors to posteriors. That is, given a prior on hypotheses, we can combine the prior with the weight of evidence to get the posterior. In particular, if there are two hypotheses, say H_1 and H_2, the weight of evidence for H_1 is α, and the prior probability of H_1 is β, then the posterior probability of H_1 (that is, the probability of H_1 in light of the evidence) is

$$\frac{\alpha\beta}{\alpha\beta+(1-\alpha)(1-\beta)}.$$

Thus, for example, by deciding to perform an action when the weight of evidence for A is 2/3 (i.e., after Charlie has seen the coin land heads once), Charlie is assured that, if the prior probability of A is at least .01, then the posterior probability of A is at least 2/11; similarly, after Charlie has seen 100 heads, if the prior probability of A is at least .01, then the posterior probability of A is at least $2^{100}/(2^{100}+99)$.

But now consider the situation in Example 1.2. Again, suppose that the prior probability of A is at least .01. Can we conclude that the posterior probability of A is at least $.01(7/9)/(.01(7/9)+.99(2/9)) = 7/205$? As we show, we cannot. The calculation $(\alpha\beta)/(\alpha\beta+(1-\alpha)(1-\beta))$ is appropriate only when there are two hypotheses. If the hypotheses A_1 and A_2 have priors α_1 and α_2 and weights of evidence β_1 and β_2, then the posterior probability of A is

$$\frac{\alpha_1\beta_1+\alpha_2\beta_2}{\alpha_1\beta_1+\alpha_2\beta_2+(1-\alpha_1-\alpha_2)(1-\beta_1-\beta_2)},$$

which is in general quite different from

$$\frac{(\alpha_1+\alpha_2)(\beta_1+\beta_2)}{(\alpha_1+\alpha_2)(\beta_1+\beta_2)+(1-\alpha_1-\alpha_2)(1-\beta_1-\beta_2)}.$$

Moreover, it is easy to show that if $\beta_1 > \beta_2$ (as is the case here), then the posterior of A is somewhere in the interval

$$\left[\frac{\alpha_2\beta_2}{\alpha_2\beta_2+(1-\alpha_2)(1-\beta_2)}, \frac{\alpha_1\beta_1}{\alpha_1\beta_1+(1-\alpha_1)(1-\beta_1)}\right].$$

That is, we get a lower bound on the posterior by acting as if the only possible hypotheses are A_2 and B, and we get an upper bound by acting as if the only possible hypotheses are A_1 and B.

In this paper, we generalize this observation by providing a general approach to dealing with weight of evidence when the likelihood function is unknown. In the special case when the likelihood function is known, our approach reduces to Shafer's approach. Roughly speaking, the idea is to consider all possible evidence spaces consistent with the information. The intuition is that one of them is the right one, but the agent trying to ascribe a weight of evidence does not know which. For example, in Example 1.2, the evidence space either involves hypotheses $\{A_1, B\}$ or hypotheses $\{A_2, B\}$: either Alice's first coin is used or Alice's second coin is used. We can then compute the weight of evidence for Alice's coin being used with respect to each evidence space. This gives us a range of possible weights of evidence, which can be used for decision making in a way that seems most appropriate for the problem at hand (by considering the max, the min, or some other function of the range).

The advantage of this approach is that it allows us to consider cases where there are correlations between the likelihood functions. For example, suppose that, in the robot example, the robot's sensor was manufactured at one of two factories. The sensors at factory 1 are more reliable than those of factory 2. Since the same sensor is used for all readings, the appropriate evidence space either uses all likelihood functions corresponding to factory 1 sensors, or all likelihood functions corresponding to factory 2 sensors.

The rest of this paper is organized as follows. In Section 2, we review Shafer's approach to dealing with evidence. In Section 3, we show how to extend it so as to deal with situation where the likelihood function is uncertain, and argue that our approach is reasonable. In Section 4, we consider how to combine evidence in this setting. We conclude in Section 5. The proofs of our technical results are deferred to the appendix.

2 Evidence: A Review

We briefly review the notion of evidence and its formalization by Shafer [1982], using some terminology from [Halpern and Pucella 2003b].

We start with a finite set $\mathcal{H}$ of hypotheses, which we take to be mutually exclusive and exhaustive; thus, exactly one hypothesis holds at any given time. We also have a set $\mathcal{O}$ of *observations*, which can be understood as outcomes of experiments that can be made. Finally, we assume that for each hypothesis $h \in \mathcal{H}$, there is a probability μ_h (often called a *likelihood function*) on the observations in $\mathcal{O}$. This is formalized as an *evidence space* $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$, where $\mathcal{H}$ and $\mathcal{O}$ are as above, and $\boldsymbol{\mu}$ is a *likelihood mapping*, which assigns to every hypothesis $h \in \mathcal{H}$ a probability measure $\boldsymbol{\mu}(h) = \mu_h$. (For simplicity, we often write μ_h for $\boldsymbol{\mu}(h)$, when the former is clear from context.)

For an evidence space $\mathcal{E}$, the weight of evidence for hypothesis $h \in \mathcal{H}$ provided by observation ob, written $w_{\mathcal{E}}(ob, h)$, is

$$w_{\mathcal{E}}(ob, h) = \frac{\mu_h(ob)}{\sum_{h' \in \mathcal{H}} \mu_{h'}(ob)}. \tag{1}$$

The weight of evidence $w_{\mathcal{E}}$ provided by an observation ob with $\sum_{h \in \mathcal{H}} \mu_h(ob) = 0$ is left undefined by (1). Intuitively, this means that the observation ob is impossible. In the literature on evidence it is typically assumed that this case never arises. More precisely, it is assumed that all observations are possible, so that for every observation ob, there is an hypothesis h such that $\mu_h(ob) > 0$. For simplicity, we make the same assumption here. (We remark that in some application domains this assumption holds because of the structure of the domain, without needing to be assumed explicitly; see [Halpern and Pucella 2003b] for an example.)

The measure $w_{\mathcal{E}}$ always lies between 0 and 1, with 1 indicating that the observation provides full evidence for the hypothesis. Moreover, for each fixed observation ob for which $\sum_{h \in \mathcal{H}} \mu_h(ob) > 0$, $\sum_{h \in \mathcal{H}} w_{\mathcal{E}}(ob, h) = 1$, and thus the weight of evidence $w_{\mathcal{E}}$ looks like a probability measure for each ob. While this has some useful technical consequences, one should not interpret $w_{\mathcal{E}}$ as a probability measure. It is simply a way to assign a weight to hypotheses given observations, and, as we shall soon see, can be seen as a way to update a prior probability on the hypotheses into a posterior probability on those hypotheses, based on the observations made.

Example 2.1 In Example 1.1, the set $\mathcal{H}$ of hypotheses is $\{A, B\}$; the set $\mathcal{O}$ of observations is simply $\{heads, tails\}$, the possible outcomes of a coin toss. From the discussion following the description of the example, it follows that $\boldsymbol{\mu}$ assigns the following likelihood functions to the hypotheses: since $\mu_A(heads)$ is the probability that the coin landed heads if it is Alice's coin (i.e., if it is double-headed), then $\mu_A(heads) = 1$ and $\mu_A(tails) = 0$. Similarly, $\mu_B(heads)$ is the probability that the coin lands heads if it is fair, so $\mu_B(heads) = 1/2$ and $\mu_B(tails) = 1/2$. This can be summarized by the following table:

$\boldsymbol{\mu}$	A	B
heads	1	1/2
tails	0	1/2

Let

$$\mathcal{E} = (\{A, B\}, \{heads, tails\}, \boldsymbol{\mu}).$$

A straightforward computation shows that $w_{\mathcal{E}}(heads, A) = 2/3$ and $w_{\mathcal{E}}(heads, B) = 1/3$. Intuitively, the coin landing heads provides more evidence for the hypothesis A than the hypothesis B. Similarly, $w(tails, A) = 0$ and $w(tails, A) = 1$. Thus, the coin landing tail indicates that the coin must be fair. This information can be represented by the following table:

$w_{\mathcal{E}}$	A	B
heads	2/3	1/3
tails	0	1

∎

It is possible to interpret the weight function w as a prescription for how to update a prior probability on the hypotheses into a posterior probability on those hypotheses, after having considered the observations made [Halpern and Fagin 1992]. There is a precise sense in which $w_{\mathcal{E}}$ can be viewed as a function that maps a prior probability μ_0 on the hypotheses $\mathcal{H}$ to a posterior probability μ_{ob} based on observing ob, by applying Dempster's Rule of Combination [Shafer 1976]. That is,

$$\mu_{ob} = \mu_0 \oplus w_{\mathcal{E}}(ob, \cdot), \tag{2}$$

where $\oplus$ combines two probability distributions on $\mathcal{H}$ to get a new probability distribution on $\mathcal{H}$ as follows:

$$(\mu_1 \oplus \mu_2)(H) = \frac{\sum_{h \in H} \mu_1(h)\mu_2(h)}{\sum_{h \in \mathcal{H}} \mu_1(h)\mu_2(h)}. \tag{3}$$

(Strictly speaking, $\oplus$ is defined for set functions, that is, functions with domain $2^{\mathcal{H}}$. We have defined $w_{\mathcal{E}}(ob, \cdot)$ as a function with domain $\mathcal{H}$, but is is clear from (3) that this is all that is really necessary to compute $\mu_0 \oplus w_{\mathcal{E}}(ob, \cdot)$ in our case.) Note that (3) is not defined if $\sum_{h \in \mathcal{H}} \mu_1(h)\mu_2(h) = 0$—this means that the update (2) is not defined when the weight of evidence provided by observation ob all goes for an hypothesis h with prior probability $\mu_0(h) = 0$

Bayes' Rule is the standard way of updating a prior probability based on an observation, but it is only applicable when we have a joint probability distribution on both the hypotheses and the observations, something which we did not assume we had. Dempster's Rule of Combination essentially "simulates" the effects of Bayes's rule. The relationship between Dempster's Rule and Bayes' Rule is made precise by the following well-known theorem.

Proposition 2.2 [Halpern and Fagin 1992] *Let $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$ be an evidence space. Suppose that P is a probability on $\mathcal{H} \times \mathcal{O}$ such that $P(\mathcal{H} \times \{ob\} | \{h\} \times \mathcal{O}) = \mu_h(ob)$ for all $h \in \mathcal{H}$ and all $ob \in \mathcal{O}$. Let μ_0 be the probability on $\mathcal{H}$ induced by marginalizing P; that is, $\mu_0(h) = P(\{h\} \times \mathcal{O})$. For $ob \in \mathcal{O}$, let $\mu_{ob} = \mu_0 \oplus w_{\mathcal{E}}(ob, \cdot)$. Then $\mu_{ob}(h) = P(\{h\} \times \mathcal{O} | \mathcal{H} \times \{ob\})$.*

In other words, when we do have a joint probability on the hypotheses and observations, then Dempster's Rule of Combination gives us the same result as a straightforward application of Bayes' Rule.

3 Evidence with Uncertain Likelihoods

In Example 1.1, each of the two hypotheses A and B determines a likelihood function. However, in Example 1.2, the hypothesis A does not determine a likelihood function. By viewing it as the compound hypothesis $\{A_1, A_2\}$, as we did in the introduction, we can construct an evidence space with a set $\{A_1, A_2, B\}$ of hypotheses. We then get the following likelihood mapping$\boldsymbol{\mu}$:

$\boldsymbol{\mu}$	A_1	A_2	B
heads	1	3/4	1/2
tails	0	1/4	1/2

Taking

$$\mathcal{E} = (\{A_1, A_2, B\}, \{heads, tails\}, \boldsymbol{\mu}),$$

we can compute the following weights of evidence:

$w_{\mathcal{E}}$	A_1	A_2	B
heads	4/9	1/3	2/9
tails	0	1/3	2/3

If we are now given prior probabilities for A_1, A_2, and B, we can easily use Proposition 2.2 to compute posterior probabilities for each of these events, and then add the posterior probabilities of A_1 and A_2 to get a posterior probability for A.

But what if we are given only a prior probability μ_0 for A and B, and are not given probabilities for A_1 and A_2? As observed in the introduction, if we define $w_{\mathcal{E}}(heads, A) = w_{\mathcal{E}}(heads, A_1) + w_{\mathcal{E}}(heads, A_2) = 7/9$, and then try to compute the posterior probability of A given that heads is observed by naively applying the equation in Proposition 2.2, that is, taking by $\mu_{heads}(A) = (\mu_0 \oplus w_{\mathcal{E}}(heads, \cdot))(A)$, we get an inappropriate answer. In particular, the answer is not the posterior probability in general.

To make this concrete, suppose that $\mu_0(A) = .01$. Then, as observed in the introduction, a naive application of this equation suggests that the posterior probability of A is 7/205. But suppose that in fact $\mu_0(A_1) = \alpha$ for some $\alpha \in [0, .01]$. Then applying Proposition 2.2, we see that $\mu_{heads}(A_1) = \alpha(4/9)/(\alpha(4/9) + (.01 - \alpha)(1/3) + .99(2/9) = 4\alpha/(\alpha + 2.01)$. It is easy to check that $4\alpha/(\alpha+2.01) = 7/205$ iff $\alpha = 1407/81300$. That is, the naive application of the equation in Proposition 2.2 is correct only if we assume a particular (not terribly reasonable) value for the prior probability of A_1.

We now present one approach to dealing with the problem, and argue that it is reasonable.

Define a *generalized evidence space* to be a tuple $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$, where Δ is a finite set of likelihood mappings. As we did in Section 2, we assume that every $\boldsymbol{\mu} \in \Delta$ makes every observation possible: for all $\boldsymbol{\mu} \in \Delta$ and all observations ob, there is an hypothesis h such that $\boldsymbol{\mu}(h)(ob) > 0$. Note for future reference that we can associate with the generalized evidence space $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ the set $\mathcal{S}(\mathcal{G}) = \{(\mathcal{H}, \mathcal{O}, \boldsymbol{\mu}) \mid \boldsymbol{\mu} \in \Delta\}$ of evidence spaces. Thus, given a generalized evidence space $\mathcal{G}$, we can define the *generalized weight of evidence* $w_{\mathcal{G}}$ to be the set $\{w_{\mathcal{E}} : \mathcal{E} \in \mathcal{S}(\mathcal{G})\}$ of weights of evidence. We often treat $w_{\mathcal{G}}$ as a set-valued function, writing $w_{\mathcal{G}}(ob, h)$ for $\{w(ob, h) \mid w \in w_{\mathcal{G}}\}$.

Just as we can combine a prior with the weight of evidence to get a posterior in a standard evidence spaces, given a generalized evidence space, we can combine a prior with a generalized weight of evidence to get a set of posteriors. Given a prior probability μ_0 on a set $\mathcal{H}$ of hypotheses and a generalized weight of evidence $w_{\mathcal{G}}$, let $\mathcal{P}_{\mu_0,ob}$ be the set of posterior probabilities on $\mathcal{H}$ corresponding to an observation ob and prior μ_0, computed according to Proposition 2.2:

$$\mathcal{P}_{\mu_0,ob} = \{\mu_0 \oplus w(ob, \cdot) \mid w \in w_{\mathcal{G}},\ \mu \oplus w(ob, \cdot) \text{ defined}\}. \tag{4}$$

Since $\mu_0 \oplus w(ob, \cdot)$ need not always exist for a given $w \in w_{\mathcal{G}}$, the set $\mathcal{P}_{\mu_0,ob}$ is made up only of those $\mu_0 \oplus w(ob, \cdot)$ that do exist.

Example 3.1 The generalized evidence space for Example 1.2, where Alice's coin is unknown, is

$$\mathcal{G} = (\{A, B\}, \{\mathit{heads}, \mathit{tails}\}, \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2\}),$$

where $\boldsymbol{\mu}_1(A) = \mu_{A_1}$, $\boldsymbol{\mu}_2(A) = \mu_{A_2}$, and $\boldsymbol{\mu}_1(B) = \boldsymbol{\mu}_2(B) = \mu_B$. Thus, the first likelihood mapping corresponds to Alice's coin being double-headed, and the second corresponds to Alice's coin being biased 3/4 towards heads. Then $w_{\mathcal{G}} = \{w_1, w_2\}$, where $w_1(\mathit{heads}, A) = 2/3$ and $w_2(\mathit{heads}, A) = 3/5$. Thus, if $\mu_0(A) = \alpha$, then $\mathcal{P}_{\mu_0,\mathit{heads}}(A) = \{\frac{3\alpha}{\alpha+2}, \frac{2\alpha}{\alpha+1}\}$. ∎

We have now given two approaches for capturing the situation in Example 1.2. The first involves refining the set of hypotheses —that is, replacing the hypothesis A by A_1 and A_2—and using a standard evidence space. The second involves using a generalized evidence space. How do they compare?

To make this precise, we need to first define what a refinement is. We say that the evidence space $(\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ *refines*, or *is a refinement of*, the generalized evidence space $(\mathcal{H}, \mathcal{O}, \Delta)$ via g if $g : \mathcal{H}' \to \mathcal{H}$ is a surjection such that $\boldsymbol{\mu} \in \Delta$ if and only if, for all $h \in \mathcal{H}$, there exists some $h' \in g^{-1}(h)$ such that $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(h')$. That is, taking $\mathcal{P}_h = \{\boldsymbol{\mu}'(h') \mid h' \in g^{-1}(h)\}$, we must have $\Delta = \prod_{h \in \mathcal{H}} \mathcal{P}_h$. Intuitively, the hypothesis $h \in \mathcal{H}$ is refined to the set of hypothesis $g^{-1}(h) \subseteq \mathcal{H}'$; moreover, each likelihood function $\boldsymbol{\mu}(h)$ in a likelihood mapping $\boldsymbol{\mu} \in \Delta$ is the likelihood function $\boldsymbol{\mu}'(h')$ for some hypothesis h' refining h. We say that $\mathcal{G}$ refines, or is a refinement of, $\mathcal{E}$ if $\mathcal{G}$ refines $\mathcal{E}$ via some surjection g. For example, the evidence space $\mathcal{E}$ at the beginning of this section (corresponding to Example 1.2) is a refinement of the generalized evidence space $\mathcal{G}$ in Example 3.1 via the surjection $g : \{A_1, A_2, B\} \to \{A, B\}$ that maps A_1 and A_2 to A and B to B. A prior μ_0' on $\mathcal{H}'$ *extends* a prior μ_0 on $\mathcal{H}$ if for all h,

$$\mu_0'(g^{-1}(h)) = \mu_0(h).$$

Let $Ext(\mu_0)$ consist of all priors on $\mathcal{H}'$ that extend μ_0. Recall that given a set $\mathcal{P}$ of probability measures, the *lower probability* $\mathcal{P}_*(U)$ of a set U is $\inf\{\mu(U) \mid \mu \in \mathcal{P}\}$ and its *upper probability* $\mathcal{P}^*(U)$ is $\sup\{\mu(U) \mid \mu \in \mathcal{P}\}$ [Halpern 2003].

Proposition 3.2 *Let $\mathcal{E} = (\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ be a refinement of the generalized evidence space $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ via g. For all $ob \in \mathcal{O}$ and all $h \in \mathcal{H}$, we have*

$$(\mathcal{P}_{\mu_0,ob})^*(h) = \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}^*(g^{-1}(h))$$

and

$$(\mathcal{P}_{\mu_0,ob})_*(h) = \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}_*(g^{-1}(h)).$$

In other words, if we consider the sets of posteriors obtained by either (1) updating a prior probability μ_0 by the generalized weight of evidence of an observation in $\mathcal{G}$ or (2) updating the set of priors extending μ_0 by the weight of evidence of the same observation in $\mathcal{E}$, the bounds on those two sets are the same. Therefore, this proposition shows that, given a generalized evidence space $\mathcal{G}$, if there an evidence space $\mathcal{E}$ that refines it, then the weight of evidence $w_{\mathcal{G}}$ gives us essentially the same information as $w_{\mathcal{E}}$. But is there always an evidence space $\mathcal{E}$ that refines a generalized evidence space? That is, can we always understand a generalized weight of evidence in terms of a refinement? As we now show, we cannot always do this.

Let $\mathcal{G}$ be a generalized evidence space $(\mathcal{H}, \mathcal{O}, \Delta)$. Note that if $\mathcal{E}$ refines $\mathcal{G}$ then, roughly speaking, the likelihood mappings in Δ consist of all possible ways of combining the likelihood functions corresponding to the hypotheses in $\mathcal{H}$. We now formalize this property. A set Δ of likelihood mappings is *uncorrelated* if there exist sets of probability measures $\mathcal{P}_h$ for each $h \in \mathcal{H}$ such that

$$\Delta = \prod_{h \in \mathcal{H}} \mathcal{P}_h = \{\boldsymbol{\mu} \mid \boldsymbol{\mu}(h) \in \mathcal{P}_h \text{ for all } h \in \mathcal{H}\}.$$

(We say Δ is *correlated* if it is not uncorrelated.) A generalized evidence space $(\mathcal{H}, \mathcal{O}, \Delta)$ is uncorrelated if Δ is uncorrelated.

Observe that if $(\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ refines $(\mathcal{H}, \mathcal{O}, \Delta)$ via g, then $(\mathcal{H}, \mathcal{O}, \Delta)$ is uncorrelated since, as observed above, $\Delta = \prod_{h \in \mathcal{H}} \mathcal{P}_h$, where $\mathcal{P}_h = \{\boldsymbol{\mu}'(h') \mid h' \in g^{-1}(h)\}$. Not only is every refinement uncorrelated, but every uncorrelated evidence space can be viewed as a refinement.

Proposition 3.3 *Let $\mathcal{G}$ be a generalized evidence space. There exists an evidence space $\mathcal{E}$ that refines $\mathcal{G}$ if and only if $\mathcal{G}$ is uncorrelated.*

Thus, if a situation can be modeled using an uncorrelated generalized evidence space, then it can also be modeled by refining the set of hypotheses and using a simple evidence space. The uncorrelated case has a further advantage. It leads to simple formula for calculating the posterior in the special case that there are only two hypotheses (which is the case that has been considered most often in the literature, often to the exclusion of other cases).

Given a generalized evidence space $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ and the corresponding generalized weight of evidence $w_{\mathcal{G}}$, we can define *upper* and *lower* weights of evidence, determined by the maximum and minimum values in the range, somewhat analogous to the notions of upper and lower probability. Define the *upper weight of evidence function* $\overline{w}_{\mathcal{G}}$ by taking

$$\overline{w}_{\mathcal{G}}(ob, h) = \sup\{w(ob, h) \mid w \in w_{\mathcal{G}}\}.$$

Similarly, define the *lower weight of evidence function* $\underline{w}_{\mathcal{G}}$ by taking

$$\underline{w}_{\mathcal{G}}(ob, h) = \inf\{w(ob, h) \mid w \in w_{\mathcal{G}}\}.$$

These upper and lower weights of evidence can be used to compute the bounds on the posteriors obtained by updating a prior probability via the generalized weight of evidence of an observation, in the case where $\mathcal{G}$ is uncorrelated, and when there are two hypotheses.

Proposition 3.4 *Let $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ be an uncorrelated generalized evidence space.*

(a) The following inequalities hold when the denominators are nonzero:

$$(\mathcal{P}_{\mu_0,ob})^*(h) \leq \frac{\overline{w}_{\mathcal{G}}(ob, h)\mu_0(h)}{\overline{w}_{\mathcal{G}}(ob, h)\mu_0(h) + \sum_{h' \neq h} \underline{w}_{\mathcal{G}}(ob, h')\mu_0(h')}; \tag{5}$$

$$(\mathcal{P}_{\mu_0,ob})_*(h) \geq \frac{\underline{w}_{\mathcal{G}}(ob, h)\mu_0(h)}{\underline{w}_{\mathcal{G}}(ob, h)\mu_0(h) + \sum_{h' \neq h} \overline{w}_{\mathcal{G}}(ob, h')\mu_0(h')}. \tag{6}$$

If $|\mathcal{H}| = 2$, these inequalities can be taken to be equalities.

(b) The following equalities hold:

$$\overline{w}_{\mathcal{G}}(ob, h) = \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h} (\mathcal{P}_{h'})_*(ob)};$$

$$\underline{w}_{\mathcal{G}}(ob, h) = \frac{(\mathcal{P}_h)_*(ob)}{(\mathcal{P}_h)_*(ob) + \sum_{h' \neq h} (\mathcal{P}_{h'})^*(ob)},$$

where $\mathcal{P}_h = \{\boldsymbol{\mu}(h) \mid \boldsymbol{\mu} \in \Delta\}$, for all $h \in \mathcal{H}$.

Thus, if have an uncorrelated generalized evidence space with two hypotheses, we can compute the bounds on the posteriors $\mathcal{P}_{\mu_0,ob}$ in terms of upper and lower weights of evidence using Proposition 3.4(a), which consists of equalities in that case. Moreover, we can compute the upper and lower weights of evidence using Proposition 3.4(b). As we now show, the inequalities in Proposition 3.4(a) can be strict if there are more than two hypotheses.

Example 3.5 Let $\mathcal{H} = \{D, E, F\}$ and $\mathcal{O} = \{X, Y\}$, and consider the two probability measures μ_1 and μ_2, where $\mu_1(X) = 1/3$ and $\mu_2(X) = 2/3$. Let $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$, where $\Delta = \{\boldsymbol{\mu} \mid \boldsymbol{\mu}(h) \in \{\mu_1, \mu_2\}\}$. Clearly, Δ is uncorrelated. Let μ_0 be the uniform prior on $\mathcal{H}$, so that $\mu_0(D) = \mu_0(E) = \mu_0(F) = 1/3$. Using Proposition 3.4(b), we can compute that the upper and lower weights of evidence are as given in the following tables:

$\overline{w}_{\mathcal{G}}$	D	E	F
X	1/2	1/2	1/2
Y	1/2	1/2	1/2

$\underline{w}_{\mathcal{G}}$	D	E	F
X	1/5	1/5	1/5
Y	1/5	1/5	1/5

The uniform measure is the identity for $\oplus$, and therefore $\mu_0 \oplus w(ob, \cdot) = w(ob, \cdot)$. It follows that $\mathcal{P}_{\mu_0,X} = \{w(X, \cdot) \mid w \in w_{\mathcal{G}}\}$. Hence, $(\mathcal{P}_{\mu_0,X})^*(D) = 1/2$ and $(\mathcal{P}_{\mu_0,X})_*(D) = 1/5$. But the right-hand sides of (5) and (6) are 5/9 and 1/6, respectively, and similarly for hypotheses E and F. Thus, in this case, the inequalities in Proposition 3.4(a) are strict. ∎

While uncorrelated generalized evidence spaces are certainly of interest, correlated spaces arise in natural settings. To see this, first consider the following somewhat contrived example.

Example 3.6 Consider the following variant of Example 1.2. Alice has two coins, one that is double-headed and one that is biased 3/4 towards heads, and chooses which one to give Zoe. Bob also has two coins, one that is fair and one that is biased 2/3 towards tails, and chooses which one to give Zoe. Zoe chooses one of the two coins she was given and tosses it. The hypotheses are $\{A, B\}$ and the observations are $\{\mathit{heads}, \mathit{tails}\}$, as in Example 1.2. The likelihood function μ_1 for Alice's double-headed coin is given by $\mu_1(\mathit{heads}) = 1$, while the likelihood function μ_2 for Alice's biased coin is given by $\mu_2(\mathit{heads}) = 3/4$. Similarly, the likelihood function μ_3 for Bob's fair coin is given by $\mu_3(\mathit{heads}) = 1/2$, and the likelihood function μ_4 for Bob's biased coin is given by $\mu_4(\mathit{heads}) = 1/3$.

If Alice and Bob each make their choice of which coin to give Zoe independently, we can use the following generalized evidence space to model the situation:

$$\mathcal{G}_1 = (\{A, B\}, \{\mathit{heads}, \mathit{tails}\}, \Delta_1),$$

where

$$\Delta_1 = \{(\mu_1, \mu_3), (\mu_1, \mu_4), (\mu_2, \mu_3), (\mu_2, \mu_4)\}.$$

Clearly, Δ_1 is uncorrelated, since it is equal to $\{\mu_1, \mu_2\} \times \{\mu_3, \mu_4\}$.

On the other hand, suppose that Alice and Bob agree beforehand that either Alice gives Zoe her double-headed coin and Bob gives Zoe his fair coin, or Alice gives Zoe her biased coin and Bob gives Zoe his biased coin. This situation can be modeled using the following generalized evidence space:

$$\mathcal{G}_2 = (\{A, B\}, \{\mathit{heads}, \mathit{tails}\}, \Delta_2),$$

where

$$\Delta_2 = \{(\mu_1, \mu_3), (\mu_2, \mu_4)\}.$$

Here, note that Δ_2 is a correlated set of likelihood mappings. ∎

While this example is artificial, the example in the introduction, where the robot's sensors could have come from either factory 1 or factory 2, is a perhaps more realistic case where correlated evidence spaces arise. The key point here is that these examples show that we need to go beyond just refining hypotheses to capture a situation.

4 Combining Evidence

An important property of Shafer's [1982] representation of evidence is that it is possible to combine the weight of evidence of independent observations to obtain the weight of evidence of a sequence of observations. The purpose of this section is to show that our framework enjoys a similar property, but, rather unsurprisingly, new subtleties arise due to the presence of uncertainty. For simplicity, in this section we concentrate exclusively on combining the evidence of a sequence of two observations; the general case follows in a straightforward way.

Recall how combining evidence is handled in Shafer's approach. Let $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$ be an evidence space. We define the likelihood functions μ_h on pairs of observations, by taking $\mu_h(\langle ob_1, ob_2 \rangle) = \mu_h(ob_1)\mu_h(ob_2)$. In other words, the probability of observing a particular sequence of observations given h is the product of the probability of making each observation in the sequence. Thus, we are implicitly assuming that the observations are independent. It is well known (see, for example, [Halpern and Fagin 1992, Theorem 4.3]) that Dempster's Rule of Combination can be used to combine evidence; that is,

$$w_{\mathcal{E}}(\langle ob_1, ob_2 \rangle, \cdot) = w_{\mathcal{E}}(ob_1, \cdot) \oplus w_{\mathcal{E}}(ob_2, \cdot).$$

If we let μ_0 be a prior probability on the hypotheses, and $\mu_{\langle ob_1, ob_2 \rangle}$ be the probability on the hypotheses after observing ob_1 and ob_2, we can verify that

$$\mu_{\langle ob_1, ob_2 \rangle} = \mu_0 \oplus w_{\mathcal{E}}(\langle ob_1, ob_2 \rangle, \cdot).$$

Here we are assuming that exactly one hypothesis holds, and it holds each time we make an observation. That is, if Zoe picks the double-headed coin, she uses it for both coin tosses.

Example 4.1 Recall Example 2.1, where Alice just has a double-headed coin and Bob just has a fair coin. Suppose that Zoe, after being given the coins and choosing one of them, tosses it twice, and it lands heads both times. It is straightforward to compute that

$w_{\mathcal{E}}$	A	B
$\langle heads, heads \rangle$	$4/5$	$1/5$
$\langle heads, tails \rangle$	0	1
$\langle tails, heads \rangle$	0	1
$\langle tails, tails \rangle$	0	1

Not surprisingly, if either of the observations is *tails*, the coin cannot be Alice's. In the case where the observations are $\langle heads, heads \rangle$, the evidence for the coin being Alice's (that is, double-headed) is greater than if a single heads is observed, since from Example 2.1, $w_{\mathcal{E}}(heads, A) = 2/3$. This agrees with our intuition that seeing two heads in a row provides more evidence for a coin to be double-headed than if a single heads is observed. ∎

How should we combine evidence for a sequence of observations when we have a generalized evidence space? That depends on how we interpret the assumption that the "same" hypothesis holds for each observation. In a generalized evidence space, we have possibly many likelihood functions for each hypothesis. The real issue is whether we use the same likelihood function each time we evaluate an observation, or whether we can use a different likelihood function associated with that hypothesis. The following examples show that this distinction can be critical.

Example 4.2 Consider Example 1.2 again, where Alice has two coins (one double-headed, one biased toward heads), and Bob has a fair coin. Alice chooses a coin and gives it to Zoe; Bob gives his coin to Zoe. As we observed, there are two likelihood mappings in this case, giving rise to the weights of evidence we called w_1 and w_2; w_1 corresponds to Alice's coin being double-headed, and w_2 corresponds to the coin being biased 3/4 towards heads. Suppose that Zoe tosses the coin twice. Since she is tossing the same coin, it seems most appropriate to consider the generalized weight of evidence

$$\{w' \mid w'(\langle ob_1, ob_2\rangle, \cdot) = w_i(ob_1, \cdot) \oplus w_i(ob_2, \cdot), i \in \{1, 2\}\}.$$

On the other hand, suppose Zoe first chooses whether she will always use Alice's or Bob's coin. If she chooses Bob, then she obviously uses his coin for both tosses. If she chooses Alice, before each toss, she asks Alice for a coin and tosses it; however, she does not have to use the same coin of Alice's for each toss. Now the likelihood function associated with each observation can change. Thus, the appropriate generalized weight of evidence is

$$\{w' \mid w'(\langle ob_1, ob_2\rangle, \cdot) = w_i(ob_1, \cdot) \oplus w_j(ob_2, \cdot), i, j \in \{1, 2\}\}.$$

∎

Fundamentally, combining evidence in generalized evidence spaces relies on Dempster's rule of combination, just like in Shafer's approach. However, as Example 4.2 shows, the exact details depends on our understanding of the experiment. While the first approach used in Example 4.2 seems more appropriate in most cases that we could think of, we suspect that there will be cases where something like the second approach may be appropriate.

5 Conclusion

In the literature on evidence, it is generally assumed that there is a single likelihood function associated with each hypothesis. There are natural examples, however, which violate this assumption. While it may appear that a simple step of refining the set of hypotheses allows us to use standard techniques, we have shown that this approach can lead to counterintuitive results when evidence is used as a basis for making decisions. To solve this problem, we proposed a generalization of a popular approach to representing evidence. This generalization behaves correctly under updating, and gives the same bounds on the posterior probability as that obtained by refining the set of hypotheses when there is no correlation between the various likelihood functions for the hypotheses. As we show, this is the one situation where we can identify a generalized evidence space with the space obtained by refining the hypotheses. One advantage of our approach is that we can also reason about situations where the likelihood functions are correlated, something that cannot be done by refining the set of hypotheses.

We have also looked at how to combine evidence in a generalized evidence space. While the basic ideas from standard evidence spaces carry over, that is, the combination is essentially obtained using Dempster's rule of combination, the exact details of how this combination should be performed depend on the specifics of how the likelihood functions change for each observation. A more detailed dynamic model would be helpful in understanding the combination of evidence in a generalized evidence space setting; we leave this exploration for future work.

Acknowledgments

Work supported in part by NSF under grants CTC-0208535 and ITR-0325453, by ONR under grant N00014-02-1-0455, by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grants N00014-01-1-0795 and N00014-04-1-0725, and by AFOSR under grant F49620-02-1-0101. The second author was also supported in part by AFOSR grants F49620-00-1-0198 and F49620-03-1-0156, National Science Foundation Grants 9703470 and 0430161, and ONR Grant N00014-01-1-0968.

A Proofs

We first establish some results that are useful for proving Proposition 3.2. The following lemma gives an alternate way of updating a prior probability by a weight of evidence.

Lemma A.1 *Let $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$ be an evidence space. For all ob and $H \subseteq \mathcal{H}$,*

$$(\mu_0 \oplus w_{\mathcal{E}}(ob, \cdot))(H) = \frac{\sum_{h\in H} \mu_0(h)\boldsymbol{\mu}(h)(ob)}{\sum_{h\in\mathcal{H}} \mu_0(h)\boldsymbol{\mu}(h)(ob)}.$$

Proof. By the definition of $\oplus$ and $w_{\mathcal{E}}$,

$$\begin{aligned}(\mu_0 \oplus w_{\mathcal{E}}(ob, \cdot))(H) &= \frac{\sum_{h\in H} \mu_0(h) w_{\mathcal{E}}(ob, h)}{\sum_{h\in\mathcal{H}} \mu_0(h) w_{\mathcal{E}}(ob, h)} \\ &= \frac{\sum_{h\in H} \mu_0(h)\frac{\boldsymbol{\mu}(h)(ob)}{\sum_{\overline{h}\in\mathcal{H}}\boldsymbol{\mu}(\overline{h})(ob)}}{\sum_{h\in\mathcal{H}} \mu_0(h)\frac{\boldsymbol{\mu}(h)(ob)}{\sum_{\overline{h}\in\mathcal{H}}\boldsymbol{\mu}(\overline{h})(ob)}} \\ &= \frac{\sum_{h\in H} \mu_0(h)\boldsymbol{\mu}(h)(ob)}{\sum_{h\in\mathcal{H}} \mu_0(h)\boldsymbol{\mu}(h)(ob)}.\end{aligned}$$

∎

Some notation will simplify the presentation of the other results. Suppose that $\mathcal{E} = (\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ is a refinement of $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ via g. Given a prior probability μ_0 on $\mathcal{H}$, recall that $Ext(\mu_0)$ consists of all priors on $\mathcal{H}'$ that extend μ_0. Let $Ext^A(\mu_0)$ be the subset of $Ext(\mu_0)$ consisting of all priors μ' on $\mathcal{H}$ such that, for all $h \in \mathcal{H}$, there exists some $h' \in g^{-1}(h)$ such that $\mu_0'(h') = \mu_0(h)$. In other words, the probability measures in $Ext^A(\mu_0)$ place all the probability $\mu_0(g^{-1}(h))$ onto a single hypothesis in $g^{-1}(h)$. If $\boldsymbol{\mu} \in \Delta$, let $\mathcal{E}_{\boldsymbol{\mu}}$ be the evidence space $(\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$.

Lemma A.2 *Let μ_0 be a prior probability on $\mathcal{H}$.*

(a) For every $\boldsymbol{\mu} \in \Delta$, there is a $\mu_0' \in Ext^A(\mu_0)$ such that, for all $h \in \mathcal{H}$ and $ob \in \mathcal{O}$, $(\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob, \cdot))(h) = (\mu_0' \oplus w_{\mathcal{E}}(ob, \cdot))(g^{-1}(h))$.

(b) For every $\mu_0' \in Ext^A(\mu_0)$, there is a $\boldsymbol{\mu} \in \Delta$ such that, for all $h \in \mathcal{H}$ and $ob \in \mathcal{O}$, $(\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob, \cdot))(h) = (\mu_0' \oplus w_{\mathcal{E}}(ob, \cdot))(g^{-1}(h))$.

Proof. Let μ_0 be a prior probability on $\mathcal{H}$. To prove (a), let $\boldsymbol{\mu}$ be a likelihood mapping in Δ. By the definition of refinement, there is a function $f_{\boldsymbol{\mu}} : \mathcal{H} \to \mathcal{H}'$ such that $f_{\boldsymbol{\mu}}(h) \in g^{-1}(h)$ and $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(f_{\boldsymbol{\mu}}(h))$. (Of course, there can be more than one such function.) Define μ_0' by taking $\mu_0'(h') = \mu_0(h)$ if $h' = f_{\boldsymbol{\mu}}(h)$ for some h, and $\mu_0'(h') = 0$ otherwise. Clearly, $\mu_0' \in Ext^A(\mu_0)$, and

$$\begin{aligned}
(\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob,\cdot))(h) &= \frac{\mu_0(h)\boldsymbol{\mu}(h)(ob)}{\sum_{\overline{h}\in\mathcal{H}} \mu_0(\overline{h})\boldsymbol{\mu}(\overline{h})(ob)} && \text{[by Lemma A.1]}\\
&= \frac{\mu_0'(f_{bmu}(h))\boldsymbol{\mu}'(f_{\boldsymbol{\mu}}(h))(ob)}{\sum_{\overline{h}\in\mathcal{H}} \mu_0'(f_{\boldsymbol{\mu}}(\overline{h}))\boldsymbol{\mu}'(f_{\boldsymbol{\mu}}(\overline{h}))(ob)} && \text{[by definition of } \mu_0' \text{ and } f_{\boldsymbol{\mu}}]\\
&= \frac{\sum_{h'\in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{\overline{h}'\in\mathcal{H}'} \mu_0'(\overline{h}')\boldsymbol{\mu}'(\overline{h}')(ob)} && \text{[adding zero terms]}\\
&= (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)).
\end{aligned}$$

The proof of (b) is analogous. Let μ_0' be a prior probability in $Ext^A(\mu_0)$. Define $f_{\mu_0'} : \mathcal{H} \to \mathcal{H}'$ so that for all $h \in \mathcal{H}$, $f_{\mu_0'}(h)$ is the unique $h' \in g^{-1}(h)$ such that $\mu_0(h) = \mu_0'(h')$. Again, by the definition of refinement, this means there is a $\boldsymbol{\mu} \in \Delta$ such that $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(f_{\mu_0'}(h))$. A straightforward computation shows that

$$\begin{aligned}
(\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob,\cdot))(h) &= \frac{\mu_0(h)\boldsymbol{\mu}(h)(ob)}{\sum_{\overline{h}\in\mathcal{H}} \mu_0(\overline{h})\boldsymbol{\mu}(\overline{h})(ob)} && \text{[by Lemma A.1]}\\
&= \frac{\mu_0'(f_{\mu_0'}(h))\boldsymbol{\mu}'(f_{\mu_0'}(h))(ob)}{\sum_{\overline{h}\in\mathcal{H}} \mu_0'(f_{\mu_0'}(\overline{h}))\boldsymbol{\mu}'(f_{\mu_0'}(\overline{h}))(ob)} && \text{[by definition of } \mu_0' \text{ and } f_{\mu_0'}]\\
&= \frac{\sum_{h'\in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{\overline{h}'\in\mathcal{H}'} \mu_0'(\overline{h}')\boldsymbol{\mu}'(\overline{h}')(ob)} && \text{[adding zero terms]}\\
&= (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)).
\end{aligned}$$

∎

Proposition 3.2 *Let $\mathcal{E} = (\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ be a refinement of the generalized evidence space $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ via g. For all $ob \in \mathcal{O}$ and all $h \in \mathcal{H}$, we have*

$$(\mathcal{P}_{\mu_0,ob})^*(h) = \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}^*(g^{-1}(h))$$

and

$$(\mathcal{P}_{\mu_0,ob})_*(h) = \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}_*(g^{-1}(h)).$$

Proof. We prove the first equality; the second follows by a similar argument. First, we prove that $(\mathcal{P}_{\mu_0,ob})^*(h) \leq \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}^*(g^{-1}(h))$. This follows almost immediately from Lemma A.2, which says that for all $w \in w_{\mathcal{G}}$, there is a measure $\mu_0' \in Ext^A(\mu_0) \subseteq Ext(\mu_0)$ such that

$$\begin{aligned}
(\mu_0 \oplus w(ob,\cdot))(h) &= (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h))\\
&\leq \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}^*(g^{-1}(h)).
\end{aligned}$$

Since $w \in w_{\mathcal{G}}$ was chosen arbitrarily, by the properties of sup, we have

$$\begin{aligned}
(\mathcal{P}_{\mu_0,ob})^*(h) &= \sup\{(\mu_0 \oplus w(ob,\cdot))(h) \mid w \in w_{\mathcal{G}}\}\\
&\leq \{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in Ext(\mu_0)\}^*(g^{-1}(h)),
\end{aligned}$$

as required.

To prove the reverse inequality, it suffices to show that for every ob and h, and for every $\mu_0' \in Ext(\mu_0)$, there is a measure $\mu_0'' \in Ext^A(\mu_0)$ such that

$$(\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) \le (\mu_0'' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)). \tag{7}$$

To prove (7), we first define a function $f_{\mu_0'} : \mathcal{H} \to \mathcal{H}'$ such that $f_{\mu_0'}(h)$ is an hypothesis in $g^{-1}(h)$ that maximizes $\boldsymbol{\mu}'(h'')(ob)$ over all $h'' \in g^{-1}(h)$, and $f_{\mu_0'}(\overline{h})$ for $\overline{h} \neq h$ is an hypothesis in $g^{-1}(\overline{h})$ that minimizes $\boldsymbol{\mu}'(h'')(ob)$ over all $h'' \in g^{-1}(\overline{h})$.

Define $\mu_0''(h')$ as follows:

$$\mu_0''(h') = \begin{cases} \mu_0'(g^{-1}(\overline{h})) & \text{if } h' = f_{\mu_0'}(\overline{h}) \text{ for some } \overline{h} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, μ_0'' is in $Ext^A(\mu_0)$. We now show that (7) holds. There are two cases. If $\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob) = 0$, then

$$\begin{array}{ll} (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) & = \frac{\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in \mathcal{H}'} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)} \quad \text{[by Lemma A.1]} \\ & = 0 \\ & \le (\mu_0'' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) \quad \text{[since } \mu_0'' \oplus w_{\mathcal{E}} \text{ is a probability].} \end{array}$$

Thus, (7) holds in this case. If $\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob) > 0$, then

$$\begin{array}{lll} (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) & = \frac{\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in \mathcal{H}'} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)} & \text{[by Lemma A.1]} \\ & = \frac{1}{1+\frac{\sum_{h' \in \mathcal{H}'-g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}} & \\ & = \frac{1}{1+\frac{\sum_{\overline{h} \in \mathcal{H}-\{h\}} \sum_{h' \in g^{-1}(\overline{h})} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(h')(ob)}} & \\ & \le \frac{1}{1+\frac{\sum_{\overline{h} \in \mathcal{H}-\{h\}} \sum_{h' \in g^{-1}(\overline{h})} \mu_0'(h')\boldsymbol{\mu}'(f_{\mu_0'}(\overline{h}))(ob)}{\sum_{h' \in g^{-1}(h)} \mu_0'(h')\boldsymbol{\mu}'(f_{\mu_0'}(h))(ob)}} & \text{[by definition of } f_{\mu_0'}] \\ & = \frac{1}{1+\frac{\sum_{\overline{h} \in \mathcal{H}-\{h\}} \boldsymbol{\mu}'(f_{\mu_0'}(\overline{h}))(ob) \sum_{h' \in g^{-1}(\overline{h})} \mu_0'(h')}{\boldsymbol{\mu}'(f_{\mu_0'}(h))(ob) \sum_{h' \in g^{-1}(h)} \mu_0'(h')}} & \\ & = \frac{1}{1+\frac{\sum_{\overline{h} \in \mathcal{H}-\{h\}} \boldsymbol{\mu}'(f_{\mu_0'}(\overline{h}))(ob) \sum_{h' \in g^{-1}(\overline{h})} \mu_0''(h')}{\boldsymbol{\mu}'(f_{\mu_0'}(h))(ob) \sum_{h' \in g^{-1}(h)} \mu_0''(h')}} & \text{[by definition of } \mu_0''] \\ & = \frac{1}{1+\frac{\sum_{\overline{h} \in \mathcal{H}-\{h\}} \sum_{h' \in g^{-1}(\overline{h})} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in g^{-1}(h)} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)}} & \text{[since } \mu_0''(h') = 0 \text{ if } h' \neq f_{\mu_0'}(\overline{h})] \\ & = \frac{1}{1+\frac{\sum_{h' \in \mathcal{H}'-g^{-1}(h)} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in g^{-1}(h)} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)}} & \\ & = \frac{\sum_{h' \in g^{-1}(h)} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)}{\sum_{h' \in \mathcal{H}'} \mu_0''(h')\boldsymbol{\mu}'(h')(ob)} & \\ & = (\mu_0'' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)). & \end{array}$$

Therefore, (7) holds in this case as well.

Now by Lemma A.2, corresponding to this μ_0'', there exists some $\boldsymbol{\mu} \in \Delta$ such that $(\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob,\cdot))(h) = (\mu_0'' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h))$. Thus,

$$\begin{aligned} (\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) &\le (\mu_0'' \oplus w_{\mathcal{E}}(ob,\cdot))(g^{-1}(h)) \\ &= (\mu_0 \oplus w_{\mathcal{E}_{\boldsymbol{\mu}}}(ob,\cdot))(h) \\ &\le (\mathcal{P}_{\mu_0,ob})^*(h). \end{aligned}$$

Since μ_0' was chosen arbitrarily, by the properties of sup, we have

$$\{\mu_0' \oplus w_{\mathcal{E}}(ob,\cdot) \mid \mu_0' \in \mathit{Ext}(\mu_0)\}^*(g^{-1}(h)) \le (\mathcal{P}_{\mu_0,ob})^*(h),$$

as required. ∎

Proposition 3.3 *Let $\mathcal{G}$ be a generalized evidence space. There exists an evidence space $\mathcal{E}$ that refines $\mathcal{G}$ if and only if $\mathcal{G}$ is uncorrelated.*

Proof. Let $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$. For the forward direction, suppose that $\mathcal{E} = (\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$ refines $\mathcal{G}$ via g. For $h \in \mathcal{H}$, let $\mathcal{P}_h = \{\boldsymbol{\mu}'(h') \mid h' \in g^{-1}(h)\}$. We show that $\Delta = \prod_{h\in\mathcal{H}} \mathcal{P}_h$. By the definition of refinement, $\boldsymbol{\mu} \in \Delta$ if and only if, for all $h \in \mathcal{H}$, there there exists some $h' \in g^{-1}(h)$ such that $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(h')$, which is the case if and only if, for all $h \in \mathcal{H}$, $\boldsymbol{\mu}(h) \in \mathcal{P}_h$, that is, $\boldsymbol{\mu} \in \prod_{h\in\mathcal{H}} \mathcal{P}_h$. Thus, $\Delta = \prod_{h\in\mathcal{H}} \mathcal{P}_h$.

For the converse, suppose that $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ is uncorrelated, so that $\Delta = \prod_{h\in\mathcal{H}} \mathcal{P}_h$, for some sets $\mathcal{P}_h$. Let $\mathcal{H}' = \{(h,\mu) \mid h \in \mathcal{H}, \mu \in \mathcal{P}_h\}$, define $\boldsymbol{\mu}'$ by taking $\boldsymbol{\mu}'((h,\mu)) = \mu$, and set $\mathcal{E} = (\mathcal{H}', \mathcal{O}, \boldsymbol{\mu}')$. We show that $\mathcal{E}$ refines $\mathcal{G}$ via g, where $g((h,\mu)) = h$. Since $\Delta = \prod_{h\in\mathcal{H}} \mathcal{P}_h$, if $\boldsymbol{\mu} \in \Delta$, then for every $\boldsymbol{\mu}(h) \in \mathcal{P}_h$, there is an $h' \in g^{-1}(h)$ (namely, $h' = (h, \boldsymbol{\mu}(h))$) such that $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(h')$, by definition of $\boldsymbol{\mu}'$. Conversely, if $\boldsymbol{\mu}$ is such that for all $h \in \mathcal{H}$, there exists some $h' \in g^{-1}(h)$ such that $\boldsymbol{\mu}(h) = \boldsymbol{\mu}'(h')$. Then $h' = (h,\mu)$ for some $\mu \in \mathcal{P}_h$, and thus $\boldsymbol{\mu}(h) \in \mathcal{P}_h$, and $\boldsymbol{\mu} \in \prod_{h\in\mathcal{H}} \mathcal{P}_h = \Delta$. This proves that $\mathcal{E}$ refines $\mathcal{G}$. ∎

Proposition 3.4 *Let $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ be an uncorrelated generalized evidence space.*

(a) The following inequalities hold when the denominators are nonzero:

$$(\mathcal{P}_{\mu_0,ob})^*(h) \le \frac{\overline{w}_{\mathcal{G}}(ob,h)\mu_0(h)}{\overline{w}_{\mathcal{G}}(ob,h)\mu_0(h) + \sum_{h'\neq h} \underline{w}_{\mathcal{G}}(ob,h')\mu_0(h')};$$

$$(\mathcal{P}_{\mu_0,ob})_*(h) \ge \frac{\underline{w}_{\mathcal{G}}(ob,h)\mu_0(h)}{\underline{w}_{\mathcal{G}}(ob,h)\mu_0(h) + \sum_{h'\neq h} \overline{w}_{\mathcal{G}}(ob,h')\mu_0(h')}.$$

If $|\mathcal{H}| = 2$, these inequalities can be taken to be equalities.

(b) The following equalities hold:

$$\overline{w}_{\mathcal{G}}(ob,h) = \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h'\neq h} (\mathcal{P}_{h'})_*(ob)};$$

$$\underline{w}_{\mathcal{G}}(ob,h) = \frac{(\mathcal{P}_h)_*(ob)}{(\mathcal{P}_h)_*(ob) + \sum_{h'\neq h} (\mathcal{P}_{h'})^*(ob)},$$

where $\mathcal{P}_h = \{\boldsymbol{\mu}(h) \mid \boldsymbol{\mu} \in \Delta\}$, for all $h \in \mathcal{H}$.

Proof. For part (a), we just prove the first inequality; the second follows by a symmetric argument. Assume that $\overline{w}_{\mathcal{G}}(ob,h)\mu_0(h) + \sum_{h'\neq h} \underline{w}_{\mathcal{G}}(ob,h')\mu_0(h') > 0$. It is clearly sufficient to show that for all $w \in w_{\mathcal{G}}$, $(\mu_0 \oplus w(ob,\cdot))(h) \leq (\overline{w}_{\mathcal{G}}(ob,h)\mu_0)/(\overline{w}_{\mathcal{G}}(ob,h)\mu_0 + \sum_{h'\neq h} \underline{w}_{\mathcal{G}}(ob,h'))$. The desired inequality then follows by properties of sup.

Given $w \in w_{\mathcal{G}}$, by definition of $\overline{w}_{\mathcal{G}}$ and $\underline{w}_{\mathcal{G}}$, $w(ob,h) \leq \overline{w}_{\mathcal{G}}(ob,h)$ and $\underline{w}_{\mathcal{G}}(ob,h) \leq w(ob,h)$, for all $h \in \mathcal{H}$ and $ob \in \mathcal{O}$. Thus,

$$\sum_{h'\neq h} \mu_0(h)\mu_0(h')w(ob,h)\underline{w}_{\mathcal{G}}(ob,h') \leq \sum_{h'\neq h} \mu_0(h)\mu_0(h')\overline{w}_{\mathcal{G}}(ob,h)w(ob,h'),$$

so

$$\begin{aligned}\mu_0(h)\mu_0(h)w(ob,h)\overline{w}_{\mathcal{G}}(ob,h) + \sum_{h'\neq h} \mu_0(h)\mu_0(h')w(ob,h)\underline{w}_{\mathcal{G}}(ob,h') \leq \\ \mu_0(h)\mu_0(h)w(ob,h)\overline{w}_{\mathcal{G}}(ob,h) + \sum_{h'\neq h} \mu_0(h)\mu_0(h')\overline{w}_{\mathcal{G}}(ob,h)w(ob,h').\end{aligned}$$

It easily follows that

$$\begin{aligned}(\mu_0 \oplus w(ob,\cdot))(h) &= \frac{\mu_0(h)w(ob,h)}{\mu_0(h)w(ob,h) + \sum_{h'\neq h}\mu_0(h')w(ob,h')} \\ &\leq \frac{\mu_0(h)\overline{w}_{\mathcal{G}}(ob,h)}{\mu_0(h)\overline{w}_{\mathcal{G}}(ob,h) + \sum_{h'\neq h}\mu_0(h')\underline{w}_{\mathcal{G}}(ob,h')},\end{aligned}$$

as required.

If $|\mathcal{H}| = 2$, we show that the inequality can be strengthened into an equality. Assume $\mathcal{H} = \{h_1, h_2\}$. Without loss of generality, it suffices to show that

$$(\mathcal{P}_{\mu_0,ob})^*(h_1) \geq \frac{\mu_0(h_1)\overline{w}_{\mathcal{G}}(ob,h_1)}{\mu_0(h_1)\overline{w}_{\mathcal{G}}(ob,h_1) + \mu_0(h_2)\underline{w}_{\mathcal{G}}(ob,h_2)}. \tag{8}$$

The key step in the argument is establishing that for $w \in w_{\mathcal{G}}$, if $\overline{w}_{\mathcal{G}}(ob,h_1) = w(ob,h_1)$ then $\underline{w}_{\mathcal{G}}(ob,h_2) = w(ob,h_2)$. We know that for every fixed ob and every w, $w(ob,h_1) = 1 - w(ob,h_2)$. If $w(ob,h_1) = \overline{w}_{\mathcal{G}}(ob,h_1)$ and $w(ob,h_2) > \underline{w}_{\mathcal{G}}(ob,h_2)$, then there must exist $w' \in w_{\mathcal{G}}$ with $w(ob,h_2) > w'(ob,h_2)$; but then we must have $w(ob,h_1) = 1 - w(ob,h_2) < 1 - w'(ob,h_2) = w'(ob,h_1)$, contradicting the fact that $w(ob,h_1) = \overline{w}_{\mathcal{G}}(ob,h_1)$. Thus, $w(ob,h_2) \leq \underline{w}_{\mathcal{G}}(ob,h_2)$, so that $w(ob,h_2) = \underline{w}_{\mathcal{G}}(ob,h_2)$. To prove (8), we now proceed as follows. Let $w \in w_{\mathcal{G}}$ be such that $w(ob,h_1) = \overline{w}_{\mathcal{G}}(ob,h_1)$. (We know such a w exists since $w_{\mathcal{G}}$ is finite.) We now get that

$$\begin{aligned}(\mathcal{P}_{\mu_0,ob})^*(h_1) &\geq (\mu_0 \oplus w(ob,\cdot))(h_1) \\ &= \frac{\mu_0(h_1)w(ob,h_1)}{\mu_0(h_1)w(ob,h_1) + \mu_0(h_2)w(ob,h_2)} \\ &= \frac{\mu_0(h_1)\overline{w}_{\mathcal{G}}(ob,h_1)}{\mu_0(h_1)\overline{w}_{\mathcal{G}}(ob,h_1) + \mu_0(h_2)\underline{w}_{\mathcal{G}}(ob,h_2)},\end{aligned}$$

as required.

For part (b), we again prove only the first equality; the second again follows by a symmetric argument. We first show that $(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob) > 0$, to establish that the right-hand side is well defined. By way of contradiction, assume that $(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob) = 0$. Since $(\mathcal{P}_h)^*(ob) = 0$, we have $\mu_h(ob) = 0$ for all $\mu_h \in \mathcal{P}_h$; similarly, for every $h' \neq h$, since $(\mathcal{P}_{h'})_*(ob) = 0$ and $\mathcal{P}'_h$ is finite, there exists $\mu_{h'} \in \mathcal{P}_{h'}$ such that $\mu_{h'}(ob) = 0$. Because Δ is uncorrelated, we can find a $\boldsymbol{\mu} \in \Delta$ such that $\boldsymbol{\mu}(h')(ob) = 0$ for every $h' \in \mathcal{H}$, contradicting the assumption we made that Δ contains only likelihood mappings that make every observation possible.

Since $\mathcal{G} = (\mathcal{H}, \mathcal{O}, \Delta)$ is uncorrelated, $\Delta = \prod_{h \in \mathcal{H}} \mathcal{P}_h$ for some sets $\mathcal{P}_h$. Thus, there exists a $\boldsymbol{\mu} \in \Delta$ such that $\boldsymbol{\mu}(h)(ob)$ is $(\mathcal{P}_h)^*(ob)$ and $\boldsymbol{\mu}(h')(ob)$ is $(\mathcal{P}_{h'})_*(ob)$ when $h' \neq h$. (The bounds are attained because each $\mathcal{P}_h$ is finite.) Since $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu}) \in S(\mathcal{G})$, we have

$$\begin{aligned} w_{\mathcal{E}}(ob, h) &= \frac{\boldsymbol{\mu}(h)(ob)}{\sum_{h' \in \mathcal{H}} \boldsymbol{\mu}(h')(ob)} \\ &= \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)} \end{aligned}$$

and thus

$$\overline{w}_{\mathcal{G}}(ob, h) \geq \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)}. \tag{9}$$

To prove equality, it suffices to show that $(\mathcal{P}_h)^*(ob)/((\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)) > w(ob, h)$ for all $w \in w_{\mathcal{G}}$. So choose $w \in w_{\mathcal{G}}$ and let the corresponding evidence space be $\mathcal{E} = (\mathcal{H}, \mathcal{O}, \boldsymbol{\mu})$ in $S(\mathcal{G})$. Given $h \in \mathcal{H}$ and $ob \in \mathcal{O}$, there are two cases. If $\boldsymbol{\mu}(h)(ob) = 0$, then

$$w(ob, h) = 0 \leq \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)}.$$

If $\boldsymbol{\mu}(h)(ob) > 0$, then $(\mathcal{P}_h)^*(ob) > \boldsymbol{\mu}(h)(ob) > 0$, so

$$\begin{aligned} w(ob, h) &= \frac{\boldsymbol{\mu}(h)(ob)}{\sum_{h' \in \mathcal{H}} \boldsymbol{\mu}(h')(ob)} \\ &= \frac{1}{1 + \frac{\sum_{h' \neq h} \boldsymbol{\mu}(h')(ob)}{\boldsymbol{\mu}(h)(ob)}} \\ &\leq \frac{1}{1 + \frac{\sum_{h' \neq h}(\mathcal{P}_{h'})_*(ob)}{(\mathcal{P}_h)^*(ob)}} \\ &= \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)}. \end{aligned}$$

Since w was arbitrary, by the properties of sup,

$$\overline{w}_{\mathcal{G}}(ob, h) \leq \frac{(\mathcal{P}_h)^*(ob)}{(\mathcal{P}_h)^*(ob) + \sum_{h' \neq h}(\mathcal{P}_h)_*(ob)}. \tag{10}$$

Equations (9) and (10) together give the result. ∎

References

Epstein, L. and M. Schneider (2005). Learning under ambiguity. Unpublished manuscript, available at http://www.econ.rochester.edu/Faculty/Epstein.html.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin & Co. Ltd.

Halpern, J. Y. (2003). *Reasoning About Uncertainty*. MIT Press.

Halpern, J. Y. and R. Fagin (1992). Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence 54*, 275–317.

Halpern, J. Y. and R. Pucella (2003a). A logic for reasoning about evidence. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI'03)*, pp. 297–304.

Halpern, J. Y. and R. Pucella (2003b). Probabilistic algorithmic knowledge. In *Proc. 9th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'03)*, pp. 118–130.

Kyburg, Jr., H. E. (1983). Recent work in inductive logic. In T. Machan and K. Lucey (Eds.), *Recent Work in Philosophy*, pp. 87–150. Rowman & Allanheld.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Shafer, G. (1982). Belief functions and parametric models (with commentary). *Journal of the Royal Statistical Society, Series B 44*, 322–352.