

A note on representing uncertainty in Inquisitive Semantics

Dylan Bumford

July 7, 2014

A classical probability space $\langle \Omega, \chi, p \rangle$ consists of a set Ω (referred to as the sample space), a σ -algebra $\chi \subseteq \wp(\Omega)$ (i.e. a family of sets closed under union and complementation), and a measure $p: \chi \rightarrow [0, 1]$ mapping subsets of the sample space Ω to real numbers in such a way that for any $E \in \chi$ and any set of disjoint $E_i \in \chi$:

- $p(E) \geq 0$
- $p(\Omega) = 1$
- $p(\bigcup E_i) = \sum p(E_i)$

This classical conception fits cleanly around a possible worlds semantics for classical logic. To see this, let $\llbracket \cdot \rrbracket: \mathcal{L} \rightarrow \wp(W)$ be the standard Tarskian interpretation function mapping classical formulas to the sets of worlds in which they're true. Then from any probability function $p_W: \wp(W) \rightarrow [0, 1]$ defined on W , the set of possible worlds, we can read off a probability function $p: \mathcal{L} \rightarrow [0, 1]$ defined on classical formulas by composing p_W with $\llbracket \cdot \rrbracket$. That is, we set $p(\phi) = p_W(\llbracket \phi \rrbracket)$.

Under conditions of certainty, this transformation will respect classical entailment and validity: if $\Gamma \vdash \phi$, then every probability function that assigns 1 to each $\gamma \in \Gamma$ will also assign 1 to ϕ . The converse is also true: if every probability function that assigns 1 to each $\gamma \in \Gamma$ also assigns 1 to ϕ , then $\Gamma \vdash \phi$. As a result, this notion of probabilistic validity is sound and complete with respect to classical logic (Leblanc 1983). More generally, the compositional procedure that routes probabilistic interpretation through classical possible worlds semantics will verify the following equations, for any formula ϕ and any set of mutually inconsistent formulas ϕ_i :

- $p(\phi) \geq 0$
- $p(\top) = 1$
- $p(\bigvee \phi_i) = \sum p(\phi_i)$

For the class of real-valued interpretation functions satisfying these constraints, Adams (1975) established that probability preservation is sound and complete with respect to classical logic. That is, for all sentences A and B , $A \vdash B$ iff every probability function guarantees that $p(A) \leq p(B)$. For proofs, generalizations, and overviews, see Adams (1998); Williamson (2002); Hájek (2001), and references therein.

For instance, the probability of a conjunction $p(\phi \wedge \psi)$ is always less than or equal to the minimum of $p(\phi)$ and $p(\psi)$ because $\llbracket \phi \wedge \psi \rrbracket$ is a subset of both $\llbracket \phi \rrbracket$ and $\llbracket \psi \rrbracket$, and the probability axioms above guarantee that subsets are never more likely than their supersets.

This connection between classical probability and classical logic is generally presupposed in informal presentations of the probability calculus (Hájek 2003). For example, the last thirty years of research on what has come to be known as the conjunction fallacy is predicated on the observation about $\phi \wedge \psi$ above. Tversky & Kahneman (1983) asked subjects to rate the comparative likelihoods of two statements, a conjunction vs. one of its conjuncts. They found that in some circumstances, people gave more credence to the conjunct than the conjunction, which they declared irrational on the basis of the probability axioms. But properly speaking, the axioms place constraints on the

probabilities of sets, not sentences. So Tversky and Kahneman, and much of the literature that they inspired, implicitly assume that (i) their question targets judgments of the relative probabilities of the *denotations* of the expressions they use, and (ii) those denotations are given by a classical interpretation function.¹

However, a variety of semantic programs for natural language depart from the simple Tarskian picture. In the dynamic semantic tradition, for instance, the meanings of expressions are identified not with the set of worlds in which they are true, but with the effect that they have on the information represented in the conversational context. This enriched notion of meaning opens the door to theories of anaphora, presupposition, and the coordination of conversants on various open informational parameters, including quantificational and modal domains, vagueness thresholds, lexical extensions, variable co-indexations, etc.

In a similar but independent vein, language researchers have recently pointed toward a handful of sentence types that seem to express speaker opinions, beliefs, preferences, or obligations, in ways that cannot be captured if sentences denote simple sets of worlds. For example, Yalcin (2007) argues that sentences with epistemic modals like ‘might’ do not express properties that any particular world can have, but rather properties of information states. That is, when I utter the sentence ‘It might be raining’, all I mean to say is that one of the worlds I imagine to represent a live candidate for being actual is such that it’s raining. But I don’t thereby announce a constraint that I believe the actual world satisfies, as I would with the sentence ‘It is raining’. In fact, the actual world needn’t be any particular way to sincerely utter the modal claim. Rather, what justifies my expression is the configuration of my beliefs, and in particular my opinion about the possibility of rain.

Yet another strain of semantic research that rejects the simplicity of propositional denotations is motivated by connections between questions, existentials, and disjunctions. Like those concerned with talk of subjective information states, these theorists have also moved toward a denotational picture that includes alternative possibilities as first-class citizens. The basic idea, played out in several different keys (e.g. Groenendijk & Stokhof 1991, Kratzer & Shimoyama 2002, Ciardelli et al. 2013), is that some sentences fundamentally denote choices, and it is these sets of choices that connectives and logical operators target. For instance, Alonso-Ovalle (2006) identifies the meaning of a disjunction with the set containing the propositions denoted by its disjuncts: $[I \text{ saw John or Mary}] = \{[I \text{ saw John}], [I \text{ saw Mary}]\}$.

This last line of research brings the challenge to classical probabilistic interpretations of natural language expressions into stark relief. Here the discussion is not about arguably pragmatic phenomena — e.g. anaphora and expressions of opinion — that appear to shave semantic interpretation a little too close for comfort, but about simple, basic boolean assertions. Yet even these sentences, it is thought, don’t denote sets of fully specified, mutually exclusive worlds, but rather sets of partial or indeterminate events, or perhaps sets of *sets* of determinate worlds. But then the assignment of probabilities to sentences cannot be computed directly in terms of the probabilities of their denotations, since denotations are not subsets of the sample space.

The obvious thing to try, then, is to lay a probability measure over the *alternatives themselves*, rather than the worlds they circumscribe. Then we could try and identify the probability of an

¹They might also justify the application of probability laws to natural language in a proof theoretic way, by saying that it is simply a property of English that expressions of the form A and B entail expressions of the form A , and that the probabilities associated with expressions should preserve entailment ordering. But then the reliance of the argument on assumptions of classical logic is even more apparent. In particular, it depends on the validity of conjunction elimination, which as far as natural language semantics is concerned, is not universally accepted (Fine 2012).

expression with the cumulative mass of the alternatives in its denotation. The problem is that alternatives are not necessarily incompatible with one another, the way that worlds are. To make this concrete, let's take alternatives to comprise sets of worlds, as in [Kratzer & Shimoyama \(2002\)](#), [Alonso-Ovalle \(2006\)](#), [Groenendijk & Roelofsen \(2009\)](#), among others. A probability measure defined on these sorts of alternatives would assign unit probability to the powerset of worlds $\wp(W)$ and zero probability to the empty set \emptyset . But we would be free to distribute mass over the singletons (each containing some $A \in \wp(W)$) any which way we wanted, completely independently of their relative overlap or disjointness. For instance, one such probability measure p_ϕ might assign $\phi = \{A\}$ and $\phi' = \{\bar{A}\}$ both probability 0.5, so that $p_\phi(\phi \cup \phi') = p_\phi(\{A, \bar{A}\}) = 1$, and $p_\phi(\phi \cap \phi') = p_\phi(\emptyset) = 0$, as expected. However, p_ϕ would have no more mass left to assign to any of other set in $\wp(W)$. So for any $\psi = \{B\}$, where $B \notin \{A, \bar{A}\}$, $p_\phi(\psi) = p_\phi(\psi') = 0$, and thus $p_\phi(\psi \vee \psi') = p_\phi(\{B, \bar{B}\}) = 0$. In general, any such probability function over sets of worlds will assign certainty to at most one tautology.

This is presumably not an epistemic constraint that our semantics should commit us to. Fortunately, there is another way to assign probabilistic values directly to the propositions in $\wp(W)$. The idea comes from work in formal epistemology in the tradition of Arthur Dempster (1967) and Glenn Shafer (1976). Dempster Shafer Theory (DST) formalizes a notion of the extent to which an event provides *evidence* for various conclusions, and a mechanism for pooling evidence from different sources. The theory has many superficial similarities to probability theory, and in some circumstances actually generalizes basic Bayesian conditionalization ([Dempster 1968](#)). But at the core of DST (at least from our perspective) is a mass function that distributes a unit amount of credence directly over alternative propositions (subsets of the sample space). Various probability-like functions can then be constructed from this credence distribution that jointly represent the range of true probabilities one could reasonably assign to worlds in light of the evidence at hand.

Formally, an event or piece of information, is modeled by an assignment of “basic belief” to *subsets* of a sample space Ω . Each such basic belief assignment (BBA) m is normalized so as to map the sets in $\wp(\Omega)$ onto positive real numbers in such a way that $1 = \sum_{A \subseteq \Omega} m(A)$. The crucial difference between a BBA m and a classical probability measure p is that p is completely determined by its behavior on the singleton subsets of $\wp(\Omega)$. This is guaranteed by the additivity axiom, $p(\bigcup E_i) = \sum p(E_i)$; since every subset $A \subseteq \Omega$ is the union $\bigcup \{\{\omega\} \mid \omega \in A\}$, the probability $p(A) = \sum_{\omega \in A} p(\{\omega\})$.

But m is not required to be additive. It distributes a unit's worth of mass over the sets of worlds in $\wp(\Omega)$ without reference to the mass of the individual worlds in those sets. So where p gives weight to a proposition A in proportion to the cumulative evidence that supports each of the worlds compatible with A (i.e. by implicitly marginalizing over all of the other possible propositions that might distinguish worlds), m gives weight to a proposition A in accordance with the evidence that can be said to *exactly support* A , independent of any other variable. That is, in light of some body of evidence, m gives A credence to the extent that the evidence supports A but nothing more specific than A .

So technically, a BBA is any function $m: \wp(\Omega) \rightarrow [0, 1]$ such that for any $E \in \wp(\Omega)$:

- $m(E) \geq 0$
- $m(\Omega) = 1$
- $\sum_{A \subseteq \Omega} m(A) = 1$

The subsets of Ω that m assigns nonzero mass to are called the *focal elements* of m .

Every BBA additionally determines two other functions bel and pl , glossed respectively as the *belief* and the *plausibility* that m assigns to each $A \subseteq \Omega$. The belief function determined by m maps every set A to the cumulative mass assigned to subsets of A . The plausibility function determined by m maps every set A to the cumulative mass assigned to sets that are consistent with A .

$$bel(A) = \sum_{B \subseteq A} m(B) \quad pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

Dempster (1967) showed that all such belief and plausibility functions definable in this way can be independently axiomatized without reference to an underlying mass function. Belief functions are exactly those functions $bel: \wp(\Omega) \rightarrow [0, 1]$ such that for any set of disjoint $E_i \in \wp(\Omega)$:

- $bel(E) \geq 0$
- $bel(\Omega) = 1$
- $bel(\bigcup E_i) \geq \sum bel(E_i)$

And to every belief function, there is a dual plausibility function such that $pl(A) = 1 - bel(\bar{A})$.

The only difference between belief functions and probability functions is the third axiom. Where probabilities are exactly additive, in the sense that the probability of two disjoint sets is the sum of their probabilities, belief functions are *superadditive*; the probability of two disjoint sets is *at least* the sum of their probabilities. Thus, it is possible in DST to speak of a belief function that assigns full credence in a tautology $\phi \vee \neg\phi$ without assigning *any* credence whatsoever to either ϕ or $\neg\phi$. In fact, the vacuous BBA $m = \Omega \mapsto 1$ represents the situation in which every such tautology is known with certainty — i.e. $bel(\Omega) = 1$ — without any commitment to any particular proposition: $bel(A) = 0, \forall A \subset \Omega$.

Every probability space $\langle \Omega, \chi, p \rangle$ determines a corresponding belief function

$$bel_p(A) = \sum_{B \subseteq A} m(B), \quad \text{where } m(B) = \begin{cases} p(B) & \text{if } B \text{ is a basis element for } \chi \\ 0 & \text{otherwise} \end{cases}$$

This belief function corresponds to the lower envelope of the probability measures extending p to Ω , (i.e. $bel_p(A) = \inf\{p'(A) \mid p': \wp(\Omega) \rightarrow [0, 1], \text{ and } p'(X) = p(X), \forall X \in \chi\}$). Conversely, every belief function is equivalent to the lower envelope of probability measures that are consistent with it. That is, for an arbitrary particular belief function bel , $bel(A) = \inf\{p(A) \mid p \in \mathcal{P}\}$, where $\mathcal{P} = \{p' \mid p'(A) \geq bel(A), \forall A \in \Omega\}$

Dually, plausibility functions correspond to upper envelopes on probability spaces. These results imply that belief functions and plausibility functions provide lower and upper bounds, respectively, for probability measures over a sample space. Thus we may think of a BBA m defined over propositions in $\wp(\Omega)$ as determining a set of probability functions \mathcal{P} whose members respect the boundaries of bel_m and pl_m , the belief and plausibility functions determined by m . That is, every BBA m determines a canonical set of measures $\mathcal{P} = \{p \mid bel_m(A) \leq p(A) \leq pl_m(A), \forall A \subseteq \Omega\}$.

When $\chi = \wp(\Omega)$, $bel_p = p$. More generally, when m assigns mass only among basis elements of χ , then $bel_m = pl_m$ are both classical probability functions for χ .

Taken together, these facts underscore the sense in which DST generalizes classical probability theory. Dempster himself appears to have thought of mass functions as essentially compact representations of families of probability distributions (Dempster 1967). Shafer & others (1976) argues

against this understanding of the mathematical theory, promoting instead a view of belief functions as defining a theory of *evidence*, independent of classical probability theory. The difference between these interpretations is not relevant here, and I will float between the language of imprecise probability and evidential support freely (see Halpern & Fagin (1992) for discussion of the relationship between these approaches).

For instance, when it comes to the issue of combining belief functions, we may sometimes wish to think of the process as a form of probabilistic update, and at other times as a kind of evidence pooling. In the former light, belief combination should resemble a generalized form of Bayesian conditionalization, tailor-made for the sort of meta-indeterminacy (indeterminacy *about* probabilities) that BBAs encode. In the latter light, combination should aggregate credence on outcomes supported by both information states, and ignore inconsistencies between them.

The classic DST combination rule, going back to Dempster (1968), is given below. Given two BBAs m_1 and m_2 , we construct a joint mass function by intersecting the focal elements of m_1 and m_2 pointwise, and multiplying their probabilities. For those pairs of elements that happen to be disjoint (representing a state the two bodies of evidence provide support for respectively inconsistent alternatives), the associated credences are redistributed evenly over the other possibilities.

$$[m_1 \oplus m_2](A) = \frac{f(A)}{1 - f(\emptyset)}, \quad \text{where } f(Z) = \sum_{X \cap Y = Z} m_1(X) m_2(Y)$$

Against this backdrop, we can return to the question of how to meaningfully distribute credences over sentences, in an alternative-oriented semantics. One such semantics, championed recently by Roelofsen (2013), building on work by Ciardelli et al. (2009), is couched in the framework of Inquisitive Semantics, which takes sentences to denote sets of nonempty propositions. Roelofsen proposes the following recursive semantics for propositional formulas:

$$\begin{aligned} [p] &= \{\llbracket p \rrbracket\}, \text{ for atomic formulas } p, \text{ where } \llbracket p \rrbracket = \{w \mid w(p)\} \\ [\neg\phi] &= \{\overline{U[\phi]}\} \\ [\phi \vee \psi] &= [\phi] \cup [\psi] \\ [\phi \wedge \psi] &= \{\alpha_{[\psi]} \mid \alpha \in [\phi]\} \cup \{\beta_{[\phi]} \mid \beta \in [\psi]\}, \text{ where } A_\Gamma = A \cap \bigcup \Gamma \\ [\diamond\phi] &= [\phi] \cup \{W\} \end{aligned}$$

Atomic formulas denote the singleton set of the worlds that verify them. Negations denote the singleton set of the worlds not included in any of the alternatives of their prejacent. Disjunctions accumulate the alternatives of their disjuncts. Conjunctions intersect each alternative in each conjunct with the union of the alternatives in the other conjunct. And epistemic possibility modals “highlight” the alternatives of their hosts. For example, Δp denotes a set containing two alternatives, one nested within the other. The outer alternative is the trivial set W , indicating that Δp does not actually provide any information, in the sense that it does not rule any state of affairs out. The inner alternative is $\llbracket p \rrbracket$, which draws attention to ϕ .

It is mostly straightforward to translate Roelofsen’s semantics into a DST-friendly interpretation system. Only conjunction — to which we return shortly — requires a bit of care. The basic strategy for interpreting the language into a Dempster Shafer Theoretic universe, will be to provide a recursive scheme that assigns BBAs to formulas in such a way that the *alternatives* of Inquisitive Semantics are preserved as *focal elements* of the corresponding BBAs. Such a scheme is presented below.

$$\begin{aligned}
\langle\langle p \rangle\rangle &= [\llbracket p \rrbracket \mapsto 1] \\
\langle\langle \neg\phi \rangle\rangle &= [\overline{\llbracket \phi \rrbracket} \mapsto 1] \\
\langle\langle \phi \vee \psi \rangle\rangle &= \left[A \mapsto \frac{1}{2} (\langle\langle \phi \rangle\rangle(A) + \langle\langle \psi \rangle\rangle(A)) \right] \\
\langle\langle \diamond\phi \rangle\rangle &= \left[\begin{array}{l} A \neq W \\ W \end{array} \mapsto \begin{array}{l} \theta \langle\langle \phi \rangle\rangle(A) \\ 1 - \theta \end{array} \right] \text{ where } \theta \ll 1 \\
\langle\langle \triangle\phi \rangle\rangle &= \left[\begin{array}{l} A \neq W \\ W \end{array} \mapsto \begin{array}{l} \rho \langle\langle \phi \rangle\rangle(A) \\ 1 - \theta \end{array} \right] \text{ where } \rho \gg 0
\end{aligned}$$

Uninquisitive sentences like ‘ p ’, for some atom p , denote categorical BBAs, all of whose mass is concentrated on $\llbracket p \rrbracket$. Note that this represents a body of evidence exactly trained on the content of p , remaining silent about everything else. But the belief function that this BBA induces will, in addition to accepting p with certainty, also accept every ψ that p entails. $\neg\phi$ likewise denote categorical BBAs, which concentrate mass entirely on the complement of the alternatives (considered jointly) in ϕ . The focal elements of a disjunction are weighted averages of the focal elements of the disjuncts.

Epistemic possibility modals, which following [Yalcin \(2010\)](#) may as well include graded modals like ‘probably’, allocate a context-sensitive amount of credence to their hosts, and reserve the rest for the trivial state W . For modals low on the possibility scale, like ‘might’, the credence apportioned to the host’s alternatives is low. For higher-likelihood modals like ‘probably’, this credence is substantially higher (presumably above 0.5 in the case of ‘probably’).

In the simplest case, we could identify logical conjunction with the Dempster Rule of Combination, defined above by $m_1 \oplus m_2$. In fact, this seems to be the best way to make sense of an earlier Inquisitive treatment for the same fragment of language which cached out conjunction in terms of pointwise intersection of alternatives ([Ciardelli et al. 2009](#)). But because of the special symmetric way that Roelofsen defines conjunction (his intention being to preserve a particular algebraic characterization of the logical operation), we have to be somewhat more careful. Rather than intersecting every focal element x in m_1 with a focal element in m_2 (and vice versa), we intersect x with the union of the focal elements in m_2 (and vice versa). Then we compute joint mass as a product of the mass assigned to x together with the masses of all y in m_2 that are consistent with x . This strategy suggests the following definition:

$$\begin{aligned}
[m_1 \otimes m_2](A) &= k_1 \sum_{\substack{A=X_{m_2} \\ Y \cap A \neq \emptyset}} m_1(X) m_2(Y) + k_2 \sum_{\substack{A=Y_{m_1} \\ X \cap A \neq \emptyset}} m_1(X) m_2(Y) \\
&= k_1 \left(\sum_{Y \cap A \neq \emptyset} m_2(Y) \right) \sum_{A=X_{m_2}} m_1(X) + k_2 \left(\sum_{X \cap A \neq \emptyset} m_1(X) \right) \sum_{A=Y_{m_1}} m_2(Y) \\
&= k_1 pl_2(A) \sum_{A=X_{m_2}} m_1(X) + k_2 pl_1(A) \sum_{A=Y_{m_1}} m_2(Y)
\end{aligned}$$

where k_1 and k_2 are normalizing constants, and $A_\Gamma = A \cap \bigcup \Gamma$, as before. As it happens, this procedure can also be expressed in terms of DST plausibility functions. From this perspective, to combine BBAs m_1 and m_2 , we first intersect each alternative of m_1 with the union of alternatives

in m_2 , and then scale the result by the plausibility that m_2 attributes to that intersection. Then we repeat the process with the alternatives in m_2 and the plausibilities determined by m_1 . Finally, we renormalize so that everything sums to one.

To complete the fragment, we set the denotation of conjunction equal to this combinatory operation on the denotations of the conjuncts.

$$\llbracket \phi \wedge \psi \rrbracket = \left[A \mapsto [m_1 \otimes m_2](A) \right]$$

These denotations guarantee that whenever a formula $\llbracket \phi \rrbracket$ has alternatives $\{\alpha_1, \dots, \alpha_n\}$, the corresponding mass function $\llbracket \phi \rrbracket$ will have $\{\alpha_1, \dots, \alpha_n\}$ exactly as its focal elements. This ensures that we preserve, among other things, the nice empirical predictions of Roelofsen’s semantics with respect to epistemic modals and disjunction. For instance, in English, all of the sentences in (1)-(3) are understood to have the same semantic effect; they all intuitively present the listener with a pair of epistemic options, without committing to either.

- (1) John might be in Paris or in London
- (2) John might be in Paris or he might be in London
- (3) John might be in Paris and he might be in London

Accordingly, given the Inquisitive interpretation function above, the three sentences will all denote the same three-element set of alternatives. One alternative will correspond to the worlds in which John is in Paris, another to the worlds in which he is in London, and a third to the trivial proposition containing all worlds (representing the fact that the sentence contains no definitive information).

Likewise, the DST interpretation function will produce BBAs for the three sentences that each contain the same three focal elements. The sentence in (1) will split a small amount of mass between the two propositions indicating John’s potential location, and reserve the rest for the trivial information state. Similarly, for (2), the proposition that John is in London will accumulate the little weight that the first conjunct affords it, and the proposition that he’s in Paris the weight that the second conjunct affords it. The trivial state will collect mass from both noncommittal disjuncts, and then everything will renormalize. Finally, in (3), since none of the alternatives in either of the conjuncts are inconsistent, all of the plausibilities will be 1. And because both conjuncts contain the trivial proposition as an alternative, every alternative will survive into the final mass function (since $A \cap W = A$, for all $A \subseteq W$). As in (2), the possibilities raised in each conjunct will keep their masses, and the trivial state will accumulate weight from both conjuncts, before renormalizing.

This is all good and well, but we still have not yet reached the point where we can assign probabilities to sentences. So far, all we have done is translate formulas into richly structured continuously-valued representations. The situation is familiar to many flavors of natural language semantics, where a variety of “closure” and “lowering” operations are needed to reduce articulated structures semantic structures to truth conditions. For instance, in the Inquisitive framework we have been considering, to determine whether a sentence ϕ is true at a particular world, we typically ask whether that world is a member of of *the union* of the alternatives denoted by ϕ . We cannot inspect the denotation directly. Similarly, most decision- and game-theoretic models that depend on formal theories of partial belief expect genuine probabilities over worlds (or hypotheses).

Fortunately, there are a variety of transformations available to DST models that convert belief functions to probability functions (for a overview, see [Daniel 2006](#)). One of the simplest such

transformations, called the plausibility transformation, leverages the structure of the DST plausibility function to compute probability values for individual worlds, based on a BBA. So, for any DST mass function m , let

$$p_m(w) = \frac{pl_m(\{w\})}{\sum_{w' \in W} pl_m(\{w'\})}$$

be the probability function that assigns to each world w the plausibility of w according to m , divided by the cumulative plausibilities over W . Halpern & Fagin (1992) and Pearl (1990) both argue that this plausibility-transformed probability function has much in common with the likelihood function of normal Bayesian update. The resemblance is in fact so clear that both authors suggest factoring probabilistic update into a two-step algorithm that routes through DST mass functions.

Halpern & Fagin (1992) advocate for a reconceptualization of “evidence” in dynamic terms. Rather than associate a piece of information or body of data with a DST belief function simpliciter, they suggest that we identify it with a particular function from prior probability distributions to posterior probability distributions. That is, for some evidence encoded as an assignment of credence m to a set of focal alternatives, Halpern & Fagin (1992) define the update function $\lambda p' . p' \oplus p_m$, where \oplus is Dempster’s Rule of Combination, and p_m is the plausibility transformation of the mass function p . This definition takes advantage of the fact that all probability functions are special cases of belief functions, and so can be combined using the tools laid out above.

To some extent, Halpern and Fagin’s dynamic turn anticipates recent expressivist approaches to the meanings of epistemically modalized sentences. Yalcin (2012), for instance, interprets expressions as instructions for updating (sets of) probability functions. For example, the atomic sentence ‘It rained’ encodes the function that conditionalizes any prior probability distribution on the set of worlds in which it rained. The more complex sentence ‘It will probably rain’, however, encodes the function that *tests* prior probability distributions to make sure that they assign sufficiently high credence to the potential for rain.

In this respect, the simple DST semantics I’ve provided for the tiny fragment above provides a sort of bridge between the linguistically-motivated representational models of natural language semantics, and the nuanced, decision-theoretic models of partial belief. Puzzled by the way people *talk about* their own epistemic states, Pearl (1990) even suggests that “there might be some phase of reasoning where people use belief functions, possibly in interpreting the impact of an observation, or during the mental exercise of estimating probabilities from a qualitative knowledge”, and furthermore that DST “could serve as a useful interface between a human probability assessor and a mechanical reasoner—a translator from colloquial assertions about probability assignments into likelihood functions.” So if the Inquisitive take on the vocabulary of subjective uncertainty, including modals like “might” and possibility-generators like “or” is a good one, then its resemblance to the basic structure of DST goes some way towards explaining how natural languages manage to express uncertainty in a way that is useful for the reasoning agent. And this sort of DST semantics provides a small step towards a robust language-reasoning interface.

References

Adams, Ernest. 1998. *A primer of probability logic*. CSLI Publications. <http://philpapers.org/rec/ADAAPO>.

- Adams, Ernest Wilcox. 1975. *The logic of conditionals: An application of probability to deductive logic* 86. Springer. <http://books.google.com/books?hl=en&lr=&id=ZnlhnGQzzU0C&oi=fnd&pg=PR9&dq=adams+probability+logic&ots=EzXIpAJcMx&sig=VINgUAXwE6wuaWRCq7FUzI2IHYI>.
- Alonso-Ovalle, Luis. 2006. *Disjunction in alternative semantics*: University of Massachusetts Amherst dissertation. <http://semanticsarchive.net/Archive/TVkY2ZIM/alonso-ovalle2006.pdf>.
- Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2009. Attention! 'might' in inquisitive semantics. In *Proceedings of SALT*, vol. 19, 91–108. <http://elanguage.net/journals/index.php/salt/article/view/19.6>.
- Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass* 7(9). 459–476. doi:10.1111/lnc3.12037.
- Daniel, Milan. 2006. On transformations of belief functions to probabilities. *International Journal of Intelligent Systems* 21(3). 261–282. <http://onlinelibrary.wiley.com/doi/10.1002/int.20134/abstract>.
- Dempster, Arthur P. 1967. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics* 325–339. <http://www.jstor.org/stable/2239146>.
- Dempster, Arthur P. 1968. A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)* 205–247. <http://www.jstor.org/stable/2984504>.
- Fine, Kit. 2012. Counterfactuals without possible worlds. *The Journal of Philosophy* 109(3). 221–246.
- Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14. 39–100.
- Groenendijk, Jeroen AG & Floris Roelofsen. 2009. Inquisitive semantics and pragmatics <http://dare.uva.nl/record/301473>.
- Halpern, Joseph Y. & Ronald Fagin. 1992. Two views of belief: belief as generalized probability and belief as evidence. *Artificial intelligence* 54(3). 275–317. <http://www.sciencedirect.com/science/article/pii/0004370292900483>.
- Hájek, Alan. 2001. Probability, logic, and probability logic. *The Blackwell guide to philosophical logic* 362–384.
- Hájek, Alan. 2003. Interpretations of probability. In *In the stanford encyclopedia of philosophy*, .
- Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from japanese. In *3rd tokyo conference on psycholinguistics*, <https://drupalnode1.ncs.mcgill.ca/linguistics/sites/mcgill.ca/linguistics/files/KratzerShimoyama2002.pdf>.
- Leblanc, Hugues. 1983. Alternatives to standard first-order semantics. In *Handbook of philosophical logic*, 189–274. Springer. http://link.springer.com/chapter/10.1007/978-94-009-7066-3_3.
- Pearl, Judea. 1990. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning* 4(5-6). 363–389. doi:10.1016/0888-613X(90)90013-R. <http://ezproxy.library.nyu.edu:2060/science/article/pii/0888613X9090013R>.
- Roelofsen, Floris. 2013. A bare bone attentive semantics for might. *The dynamic, inquisitive, and visionary life of ϕ* http://www.illc.uva.nl/Festschrift-JMF/papers/25_Roelofsen.pdf.
- Shafer, Glenn & others. 1976. *A mathematical theory of evidence*, vol. 1. Princeton university press Princeton. <http://pup.princeton.edu/titles/2439.html>.
- Tversky, Amos & Daniel Kahneman. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90(4). 293–315. doi:10.1037/0033-295X.90.4.293.

- Williamson, Jon. 2002. Probability logic. *Handbook of the Logic of Inference and Argument: The Turn Toward the Practical* 1. 397–424. http://books.google.com/books?hl=en&lr=&id=c-bthgk3DQAC&oi=fnd&pg=PA397&dq=adams+primer+probability+logic&ots=yqdiXnJa4P&sig=NZZeipGukv3JxP-q0Chaxh0g_yw.
- Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. <http://mind.oxfordjournals.org/content/116/464/983.short>.
- Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937. <http://onlinelibrary.wiley.com/doi/10.1111/j.1747-9991.2010.00360.x/full>.
- Yalcin, Seth. 2012. Context probabilism. In *Logic, language and meaning* 7218, 12–21. http://link.springer.com/chapter/10.1007/978-3-642-31482-7_2.