Communicating with Epistemic Modals in Stochastic λ-Calculus

Daniel Lassiter and Noah D. Goodman
Stanford Psychology
{danlassiter, ngoodman}@stanford.edu

Draft: comments welcome.

Abstract Dynamic semantic theories generally treat epistemic modals as operators which test that their propositional argument bears some relation to the conversational common ground (e.g., entaiment or partial overlap), but do not change it. This means that epistemically modalized sentences do not carry any information about the world, except perhaps as a meta-comment on the state of the conversation and the information that its participants have. We argue that this prediction is incorrect, and discuss corresponding technical challenges for modeling the informational effect of epistemic modals in a recent probabilistic model of dynamics. We show how the challenges can be overcome using a model in which agents' knowledge is structured using hierarchical generative models and a semantics and pragmatics making crucial use of the stochastic lambda calculus implemented in the Church probabilistic programming language. This account makes precise quantitative predictions about the information that epistemic sentences convey about contingencies in the world.

Keywords: Epistemic modality, dynamic semantics, Bayesian pragmatics, probabilistic models, reasoning

What information do sentences containing epistemic modals convey? Consider the exchange:

(1) a. A: What will happen in the game tonight between team X and team Y?
    b. B: Team X is likely to win.

After hearing this utterance, A will typically have more information about the outcome of the game than he did before. A may also learn something about other factors which have an influence on the outcome: however B arrived at the conclusion summarized by his utterance, if his utterance is well-supported he must have evidence that relevant factors have conspired to favor team X.

The observation that (1b) carries information about the game's outcome seems fairly banal, but it is surprisingly difficult to devise a formal theory of the flow of information in discourse that can capture the fact that utterances containing epistemic modals convey non-trivial information about the world. In particular, mainstream dynamic semantic theories treat (1b) as conveying no information, or information only about the conversation between A and B (rather than the game).

A probabilistic theory of discourse which treats update as conditionalization offers a promising extension to dynamic semantics, and interfaces cleanly with cognitively motivated probabilistic models of language understanding and knowledge representation and reasoning. It would also appear to fit well with recently proposed probabilistic semantics for epistemic modals (Yalcin 2010; Lassiter 2011a). §2 sketches such a model and notes a serious problem: it is not possible to conditionalize a probability measure directly on a probability condition such as "$pr(p) > .5$", and so update with an epistemically modalized sentence is not even defined on these assumptions. In §3 we describe a probabilistic model of dynamics drawing inspiration from work on hierarchical generative models from psychology and AI, which is implemented in the probabilistic functional programming language Church. We give a new sampling semantics for probability operators, show how to model informative epistemic statements on this interpretation, and argue that this theory makes intuitively reasonable predictions about the information conveyed by epistemically modalized sentences. The technical appendix gives details of the Church model and simulation results.

1 Epistemic modality in dynamic semantics

In mainstream dynamic semantics information states (or partially shared information states known as *common grounds*) are modeled as sets of possible worlds. Typical non-modal sentences have their informational effect by narrowing down the set of worlds (roughly, by intersecting the current common ground with the set of worlds in the proposition that the sentence denotes in context). Epistemically modalized sentences, however, are modeled as *tests*: partial identity functions on the current common ground which are defined only if the state in question has some particular form (Veltman 1996). In this account, an epistemically modalized sentence, uttered and accepted, does not change the information state/common ground, and so does not carry any information about the world. Indirectly, epistemic modals may carry some meta-information by functioning as a public acknowledgement of relevant features of the common ground — which are, however, assumed to be known already to all involved (cf. von Fintel & Gillies 2007).

A model of this sort has difficulty in accounting for the interaction in (1). B's utterance will typically lead to a net gain in information for A, rather than simply checking that A and B already (publicly) share this information. As Beaver (2001) points out, this apparent problem can be avoided by upgrading the representation of common ground: instead of treating the common ground as a single set of worlds, we may treat it as a set of such states or as a more complicated object such as a preference order or probability distribution over information states (cf. Roussarie 2009). However, the revised dynamic model continues to predict that epistemically modalized sentences can only be used in situations in which the common ground *might* have the appropriate shape. This does not quite capture (1) either: this interaction could be felicitious even if it is clearly **not** common ground that *Team X will win* is likely (for example, if the game has not been discussed and A has no idea what information B possesses on this topic). This suggests that a model relying on uncertainty about common ground to model informative epistemic statements is not adequate.

2 Probabilistic dynamics and modality

Probabilistic theories of information flow in discourse combine mainstream dynamic models of discourse with the claim that information states are not *sets* of worlds but *probability distributions* over sets of worlds. This richer structure supports more fine-grained inferences in various ways and is backed by much evidence from recent cognitive science (cf. Griffiths, Kemp & Tenenbaum 2008 and references therein). The probabilistic theory of Lassiter (2012), for instance, models conversational participants' epistemic states and their interaction without making assumptions of shared information. An agent who accepts the information conveyed by a sentence will add this information to her mental model of the world in the normative Bayesian manner by conditionalizing on it. Specifically, if an agent's information state prior to utterance u in context c is given by the probability measure $pr(\cdot)$, and the denotation of u in c is proposition p, then her posterior information state is given by the conditional probability measure $pr(\cdot|p)$, defined as $\lambda q.pr(p \wedge q)/pr(p)$.

Recent work on the semantics of epistemic modals such as *likely* (our example from (1b)) suggests that they are also best treated in probabilistic terms (Yalcin 2007, 2010; Lassiter 2010, 2011a; Klecha 2012). For example, *p is likely* requires that the probability of *p* be greater than a threshold θ_{likely}, which (for present purposes) we may set at .5. This would seem to be the natural theory of modality to pair with a probabilistic theory of dynamics, but there is a problem: the operation of conditionalization requires a proposition as input, and this semantics for epistemic modals does not give us one. That is, if $[\![p \textit{ is likely}]\!]^c$ denotes the condition that $pr(p) > .5$, it does not make sense to conditionalize on this statement.

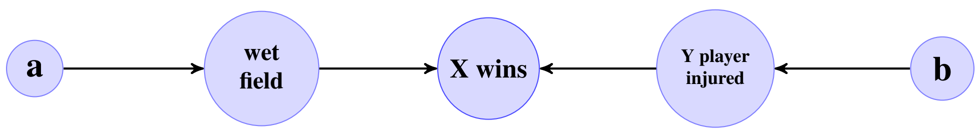
One possible response is to treat the meaning of *p is likely* not as directly expressing a condition on probabilities, but as expressing the proposition that *p* has high probability in the epistemic state of some relevant individual(s). We find this approach unsatisfactory: what (1b) expresses is not about the speaker's epistemic state or anyone else's, as an overt *I believe that Team X will win* would be. A listener who accepts the information contained in an utterance of (1b) will not simply add to her beliefs the information that *speaker B believes* that team X will likely win, but the information that team X *will* likely win. We thus conclude that epistemic statements are not covert belief statements, but rather convey information about (potentially) non-mental events.

A second response is to make use of sets of probabilities, probability ranges, higher-order probabilities, or other enriched representations (Halpern 2003). Models built around these tools are able to model informative epistemic statements (cf. Yalcin 2012, who makes this point using sets of probability measures). However, with the exception of higher-order probabilities (on which see below), these approaches face an empirical difficulty. A realistic system should be able to model a scenario of the following sort: at $t' < t$, *A* utters "It is likely to rain at t", and *B* accepts this statement. Then, upon receiving information at a later $t'' < t$ favoring the contrary conclusion, *A* utters "It is unlikely to rain at t", and *B* responds by adjusting her information state accordingly. With ranges or sets of probability assignments, once the probability assignments that fail to make *It is likely to rain* true have been discarded they cannot be recovered, and the second utterance should lead to contradiction. That is, an update model based on these tools would seem to predict that information gain should monotonically reduce the admissible probability assignments. The model that we will present captures the core insight that uncertainty about probability assignments is needed, but does so without problematic monotonicity predictions or special-purpose machinery: instead we use generative models, which are compact, computationally efficient, and independently motivated.

3 Conditioning on uncertain information

We assume, along with much work in recent psychology and artificial intelligence, that agents' beliefs about the world can be represented as probabilistic generative models (a.k.a. causal models; Pearl 2000; Griffiths et al. 2008). Generative models specify probabilistic dependencies between variables and propositions of interest. The dependencies represented in a generative model capture the knowledge summarized by English statements such as *If the field is wet, team X is more likely to win* or *If Team Y's star player is injured, they are less likely to win*. Variables that have a causal influence on the variable of interest may themselves be dependent on further causes. Learning about the value of any of these variables can allow us to make more informed inferences about others: for instance, if an agent discovers that it is raining, this increases the chance that the field will be wet, which (as the example goes) increases the likelihood that team X will win.

Following de Finetti (1977), Pearl (1987, 1988) argues that generative models automatically implement uncertainty about probability assignments, and thus effectively handle phenomena sometimes thought to enriched representations of probabilities. Rather than stipulating that the probability of team X winning falls between, say, .3 and .7, the range of probabilities for the game's outcome is integrated into our uncertainty about the values of causally related variables. In this way generative models implicitly define a restricted kind of higher-order probability. We illustrate with a simple graphical model of relevant portions of *A*'s belief state, where **a** and **b** are parameters representing further causal factors not explicitly represented here (such as rain).



The structure of this model tells us that the probability of **X wins** will vary depending on the values of the causally related variables **wet field** and of **Y player injured**. We specify the model more fully using the probabilistic functional programming language Church, which implements the stochastic λ-calculus underpinning our model. `flip` is a stochastic function which "flips a weighted coin", returning `true` with probability equal to its argument. `X-wins` is defined as a noisy-OR gate whose parameters depend on the values of the higher variables `Y-player-injured` and `wet-field`. Each of these depends in turn on higher parameters.[1]

```
(define (X-wins) (or [and (true-in-w?  wet-field) (flip .6)]
                     [and (true-in-w?  Y-player-injured) (flip .4)]
                     [flip .2]))
```

1 A complete working model with more detailed notes on interpretation is given in the Technical Appendix. The syntax and semantics of Church (Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008), a probabilistic variant of the Scheme dialect of Lisp, is described in detail at http://projects.csail.mit.edu/church/wiki/Church. Functions precede their arguments, and statements wrapped in parentheses are functions, whether or not they have arguments. The predicate `true-in-w?` ensures that stochastic propositions maintain their truth-values in a world, in the style of random-worlds semantics (McAllester, Milch & Goodman 2008). It plays a crucial but subtle role in the inference procedure, and is discussed in detail in the appendix.

```
(define (wet-field) (flip a))
(define (Y-player-injured) (flip b))
```

The model defines a probability distribution and samples a (partial) world w in accordance with this distribution each time it is run, assigning to each variable the value that it has in w. Depending on the truth-values of the higher variables `Y-player-injured` and `wet-field` in the sampled world w, the probability that `X-wins` will be true in w ranges between a maximum of .808 and a minimum of .2. These are, respectively, the probabilities if both enabling variables are true, and if neither is.

`likely` is defined as a function which checks that the probability of its argument `p` is greater than 0.5, approximated by independently sampling `n` worlds and finding the proportion which satisfy `p`. The choice of `n` controls the speed-accuracy tradeoff in the estimate of $pr(p)$.

```
(define (likely p) (> (prob p) .5))
(define (prob p) (/ (num-true (repeat n p)) n))
```

Inference (conditionalization) in Church involves finding the value of a *query* expression in a world sampled from a *model* given a Boolean *condition*, as shown schematically on the left below.

```
(church-query                          (church-query
   (define model ...)                     (define model ...)  ;; as above
   query-expression                       (prob wet-field)
   conditioning-expression)               (likely X-wins))
```

The crucial new semantic feature of our model is that probability operators such as `likely`, when embedded inside `church-query`, take into account the randomness that adheres to a particular node *given the actual values of all nodes that causally influence that node*. This means that the probability of a proposition `p` will vary from world to world: for instance, the estimated probability of `X-wins` in world w_1 will differ from the estimated probability of `X-wins` in world w_2 if these worlds differ in the truth-value of a causally connected node such as `wet-field`.

If the conditioning expression happens to be `(likely X-wins)`, this has the effect of restricting attention to w in which the values of variables causally related to `X-wins` have their truth-values set so as to render `X-wins` likely. More generally, learning that p is likely means that the probabilities of other variables q which are causally connected to p must be arranged in a way that supports this fact. Inferences involving q connected to p will be sensitive to this.

In this way, generative models and the semantics for probability operators that we have described here makes it possible to conditionalize on a probability statement such as "$pr(p) > .5$", with non-trivial informational consequences. Testing various queries in the place of `(prob wet-field)` gives results consonant with our intuitive understanding of what follows from "Team X is likely to win" in this context. For instance, in this model A has no idea prior to the utterance whether the field is wet or Y's star player is injured. Learning that X is likely to win will lead A, quite reasonably, to consider it more likely that the enabling factors in her model are fulfilled. Correspondingly, the marginal probabilities of these variables increase after conditionalizing on "X is likely to win" (see the simulation results in the Technical Appendix, §5.2). Note that the model makes a further empirical prediction: the informational effects of epistemically modalized sentences should be sensitive to the *structure* of the generative model. Here, `(likely X-wins)` is easier to satisfy if both causally relevant nodes — `wet-field` and `Y-player-injured` — are true. Adding further causally connected nodes or modifying their interaction would affect the total informational effect of an utterance of "Team X is likely to win" on *A*'s information state. The prediction of structure-dependence strikes us as very reasonable, but further work will be needed to verify it.

4 The information conveyed by epistemic utterances

The probabilistic semantics sketched here differs from mainstream dynamic semantics in treating epistemically modalized sentences as first-class bearers of information. According to our model of the dialogue in (1), if the relevant portion of A's mental model of the world is captured by the model given here, learning from B that team X is likely to win will cause A to conditionalize on `(likely X-wins)`. This leads to revised probabilities for both the proposition `X-wins` and for other variables in her causal model which are causally related to it. Crucially, however, the extent of revision to other variables will depend on how strict the condition imposed by the epistemic modal is, i.e. how high the threshold is. Although we have only discussed *likely*, it is easy to introduce epistemic operators with higher or lower threshold values such as *certain*, *possible*, *might*, or *must*. Changing the modal in the target utterance will give rise to different effects on the estimated probabilities of other variables, capturing the intuition that these utterances convey different information.

5 Technical Appendix

5.1 Church Implementation and Notes

Church is a variant of the Scheme dialect of Lisp enriched with stochastic primitives and various library functions useful for probabilistic modeling. See the Church Wiki[2] for a detailed presentation of its syntax and semantics along with philosophical, psychological, and computational motivation and many examples of models and applications.

We now specify the model of the dynamics of A's belief state in the course of the interaction in (1). We begin with a simplified model which is semantically clearer, and then present below a slightly more complex working version. First, the probability of `p` is approximated by applying to it a function `prob` which samples repeatedly and independently from the distribution over worlds defined by the generative model and finds the proportion of samples in which `p` is true. The `n` parameter in the definition of `prob` controls the speed/accuracy trade-off in the approximation to the probability of `p`.

```
(define (prob p) (/ (num-true (repeat n p)) n))
```

We interpret epistemic modals as operators which check that the probability of their propositional argument is above or below a threshold value which is lexically specified by the modal in question (Lassiter 2011a).[3]

```
(define (likely p) (> (prob p) .5))
(define (unlikely p) (< (prob p) .5))
(define (might p) (> (prob p) 0))
(define (possible p) (> (prob p) 0))
(define (certain p) (= (prob p) 1))
```

`church-query`, as described in the main text, is a placeholder for a variety of conditional inference techniques. The simplest implementation uses rejection sampling, approximating the conditional probability of `p` given `q` by repeatedly sampling from the distribution on possible worlds but discarding samples which are not `q`-worlds, and then finding the proportion of the remaining worlds which are `p`-worlds. Rejection sampling is frequently very slow, however, and we generally prefer to use more efficient sampling methods. Here we use `mh-query`, based on the Metropolis-Hastings algorithm which implements a form of Markov Chain Monte Carlo inference. In most cases `mh-query` will return samples from a conditional distribution much more quickly than `rejection-query`.

2 http://projects.csail.mit.edu/church/wiki/Probabilistic_Models_of_Cognition

3 The assumption of a fixed threshold for each modal and for the assert operator is a simplification which could be relaxed by defining a probability distribution over threshold values for each item. This feature would add the flexibility needed to make inferences over threshold values, which is crucial for a probabilistic model of the context-sensitivity of many epistemic modals (Yalcin 2010; Lassiter 2011a: ch.4) and of the assertibility threshold (e.g., Lewis 1979). See also Frazee & Beaver 2010; Lassiter 2011b for arguments that probability distributions over threshold values are a semantic feature of vague gradable adjectives generally, a category which includes the epistemic adjectives that we are considering here.

A key feature of the model is the function `true-in-w?`, which we attach to the definitions of all propositions which depend for their probabilities on the truth-values of other propositions.

```
(define true-in-w?  (mem (lambda (p) (sample p))))
```

The purpose of calling this function in the definitions of stochastic propositions which depend on other propositions is to make sure that propositions maintain their truth-value within a world. The truth-value maintenance is done via a built-in Church function `mem` which *memoizes* the value of `(sample p)` and returns the same value every time `(true-in-w? p)` is called thereafter within a single query. In effect, `(true-in-w? p)` is the Church equivalent of $[\![p]\!]^w$: in each sampled world w, it maps all stochastic propositions to their truth-value in w. This feature is semantically crucial because propositions must maintain stable truth-values in individual worlds, but probability operators must also have access to their local stochastic behavior, i.e., they must be able to "see" the randomness that remains in the model when the values of all variables which influence `p` are determined. We may think of this as a "worlds-within-worlds" semantics for probability operators, analogous to the fact that, in modal logic, $\Box p$ and $\Diamond p$ when evaluated at w look to other worlds accessible from w to find the value of p there. The essential difference is that what is epistemically accessible from w is not a set of worlds but a distribution over worlds, and that we do not make the assumption that all epistemically possible worlds are associated with the same distribution.

Some further notes on interpreting the code on the following page (see also footnote 1):

- `(flip b)` is a function which "flips a coin" with bias `b`, returning `true` with probability `b` and `false` otherwise. For example, `(flip .7))` will return `true` 70% of the time.
- The line `(define (wet-field) (flip b))` defines `wet-field` as a function with no arguments (a "thunk"). Outside of its definition, surrounding a thunk with parentheses, as in `(wet-field)`, causes it to be evaluated (sampled from). Each evaluation returns an independent sample from the given distribution. To see the importance of this difference, suppose we define `foo` as in `(define (foo) (flip .5))` and define `bar` as in `(define bar (foo))`. This definition will lead `bar` to be equal to some sample from `foo`, i.e., to either `true` or `false` with equal probability. Thus the formula `(equal?  bar bar)` will always be true. However, the superficially similar `(equal?  (foo) (foo))` will be true only 50% of the time, because each evaluation of `(foo)` returns an independent sample.
- The two numbers after `mh-query` tell the program, respectively, how many samples to take in performing conditional inference and how many random steps to take between samples.

```
(mh-query 100 300
   (define true-in-w? (mem (lambda (p) (sample p))))
   (define a (uniform 0 1))
   (define b (uniform 0 1))
   (define (Y-player-injured) (flip a))
   (define (wet-field) (flip b))
   (define (X-wins) (or [and (true-in-w? wet-field) (flip .6)]
                        [and (true-in-w? Y-player-injured) (flip .4)]
                        [flip .2]))

   (prob wet-field)            ;; the query: something A is interested in knowing

   (likely X-wins))                        ;; the conditioner: what A has learned
```

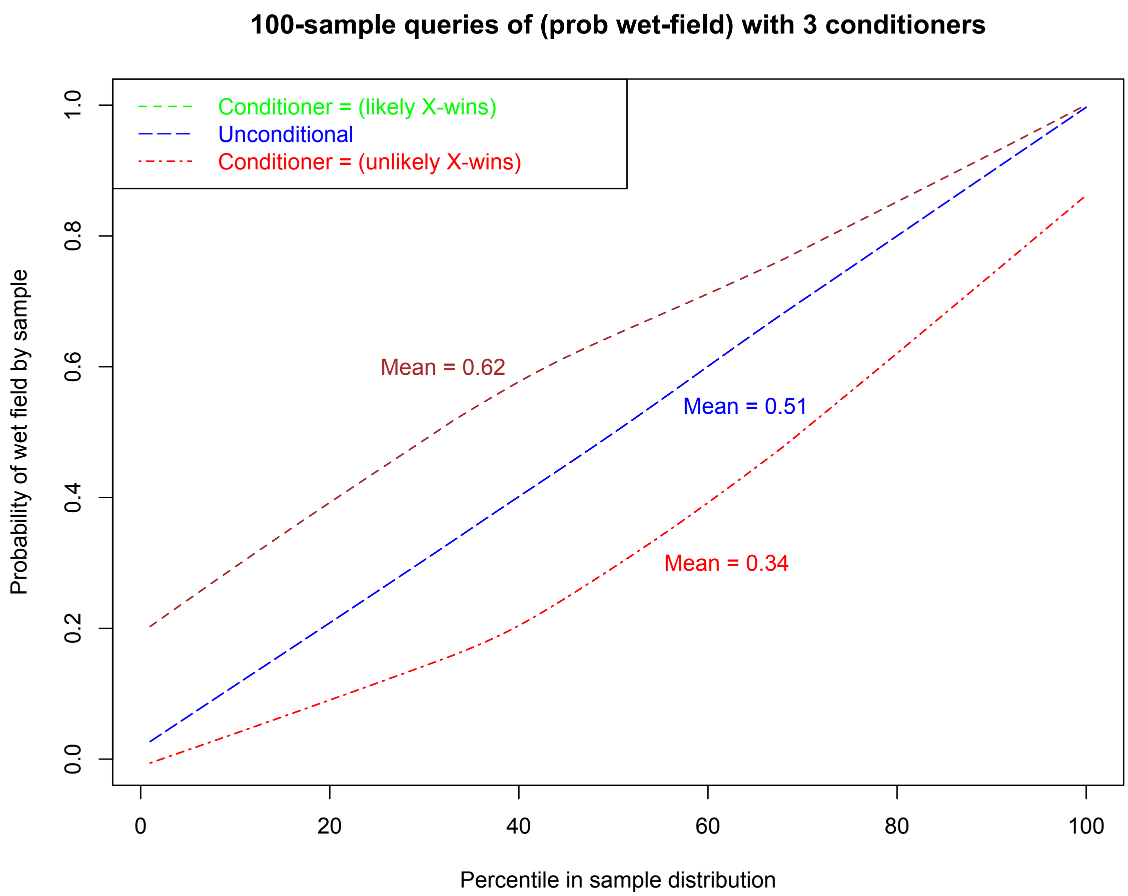
The final components of the model are a handful of parameters and helper functions. The only component of note is the `n` parameter, which controls the speed-accuracy tradeoff in calculating probabilities of stochastic propositions in the definition of the `prob` function above.

```
(define n 100)
(define (num-true lst)
   (if (equal? lst '()) 0
      (if (equal? (first lst) true) (+ 1 (num-true (rest lst)))
         (num-true (rest lst)))))
```

The next subsection gives illustrative simulation results.

5.2 Simulation Results

We illustrate the informational effect of epistemically modalized utterances with results from three model runs querying the value of `(prob wet-field)` with different conditioning expressions. The results are given in rank order on the *x*-axis, and the *y*-axis shows the estimated probability in that sample. Note that the estimated probability that the field is wet tends to be substantially higher when we conditionalize on `(likely X-wins)` than it is unconditionally (i.e. with conditioner `true`), and it is substantially lower when we conditionalize on `(unlikely X-wins)`. This illustrates the informational impact of epistemically modalized sentences in our treatment.



5.3 Full working version

This section gives a slight modification of the above code which runs as is using the current version of Bher, an implementation of Church which can be accessed through a web-based interface at the Church Wiki or can be installed locally.[4] It is semantically equivalent to the version presented in §5.1 but modifies the `true-in-w?` function in order to avoid difficulties in memoizing certain functions that take stochastic functions as arguments in the current version of Bher. Syntactically, the differences are that arguments of `true-in-w?` must be quoted (preceded by a `'`), and each stochastic proposition defined must be matched with a quoted form using `get-procedure`. This is merely a workaround, and can be ignored except by readers who are interested in running or modifying the program themselves.

```
(define (prob p) (/ (num-true (repeat n p)) n))
(define (likely p) (> (prob p) .5))
(define (unlikely p) (< (prob p) .5))
(define (might p) (> (prob p) 0))
(define (possible p) (> (prob p) 0))
(define (certain p) (= (prob p) 1))

(mh-query 100 300
   (define true-in-w? (mem (lambda (p) ((get-procedure p)))))
   (define a (uniform 0 1))
   (define b (uniform 0 1))
   (define (Y-player-injured) (flip a))
   (define (wet-field) (flip b))
   (define (X-wins) (or [and (true-in-w? 'wet-field) (flip .6)]
                        [and (true-in-w? 'Y-player-injured) (flip .4)]
                        [flip .2]))
   (define (get-procedure name)
      (cond [(equal? name 'Y-player-injured) Y-player-injured]
            [(equal? name 'wet-field) wet-field]
            [(equal? name 'X-wins) X-wins]
            [else (error name "unknown procedure")]))

   (prob wet-field)            ;; the query: something A is interested in knowing

   (likely X-wins))                        ;; the conditioner: what A has learned

(define n 100)
(define (num-true lst)
   (if (equal? lst '()) 0
      (if (equal? (first lst) true) (+ 1 (num-true (rest lst)))
         (num-true (rest lst)))))
```

4 Installation instructions are available at http://projects.csail.mit.edu/church/wiki/Installing_Bher.

References

Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. CSLI.

de Finetti, Bruno. 1977. Probabilities of probabilities: a real problem or a misunderstanding? In A. Aykac & C. Brumet (eds.), *New developments in the applications of Bayesian methods*, North-Holland.

von Fintel, Kai & Anthony S. Gillies. 2007. An opinionated guide to epistemic modality. *Oxford studies in epistemology* 2. 32–62.

Frazee, Joey & David Beaver. 2010. Vagueness is rational under uncertainty. In Maria Aloni, Harald Bastiaanse, Tikitu de Jager & Katrin Schulz (eds.), *Logic, language and meaning: 17th amsterdam colloquium, amsterdam, the netherlands, december 16-18, 2009, revised selected papers*, 153–162. Springer.

Goodman, N.D., V.K. Mansinghka, D. Roy, K. Bonawitz & J.B. Tenenbaum. 2008. Church: a language for generative models. In *Uncertainty in artificial intelligence*, vol. 22, 23.

Griffiths, Thomas L., Charles Kemp & Joshua B. Tenenbaum. 2008. Bayesian models of cognition. In Ron Sun (ed.), *Cambridge handbook of computational psychology*, 59–100. Cambridge University Press.

Halpern, Joseph Y. 2003. *Reasoning about Uncertainty*. MIT Press.

Klecha, Peter. 2012. Positive and conditional semantics for gradable modals. In Anna Chernilovskaya Ana Aguilar Guevara & Rick Nouwen (eds.), *Sinn und bedeutung 16*, vol. 2, 363–376. MIT Press. http://mitwpl.mit.edu/open/sub16/.

Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. In Nan Li & David Lutz (eds.), *Semantics & Linguistic Theory (SALT) 20*, 197–215. CLC Publications.

Lassiter, Daniel. 2011a. Measurement and Modality: The Scalar Basis of Modal Semantics. Ph.D. thesis, New York University.

Lassiter, Daniel. 2011b. Vagueness as probabilistic linguistic knowledge. In R. Nouwen, R. van Rooij, U. Sauerland & H.-C. Schmitz (eds.), *Vagueness in communication*, 127–150. Springer.

Lassiter, Daniel. 2012. Presuppositions, provisos, and probability. *Semantics & Pragmatics* 5. 1–37.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359. doi:10.1007/BF00258436.

McAllester, D., B. Milch & N.D. Goodman. 2008. Random-world semantics and syntactic independence for expressive languages. Technical Report MIT-CSAIL-TR-2008-025, Massachusetts Institute of Technology.

Pearl, Judea. 1987. Do we need higher-order probabilities, and, if so, what do they mean? In *Proceedings of the AAAI workshop on uncertainty in AI*, 47–60.

Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pearl, Judea. 2000. *Causality: Models, reasoning and inference*. Cambridge University Press.

Roussarie, Laurent. 2009. What might be known: Epistemic modality and uncertain contexts. In Ed Cormany, Satoshi Ito & David Lutz (eds.), *Semantics and Linguistic Theory (SALT) 19*, 379–394.

Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25(3). 221–261.

doi:10.1007/BF00248150.
Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026.
Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937.
Yalcin, Seth. 2012. Context probabilism. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz & M. Westera (eds.), *Logic, language and meaning*, vol. 7218 Lecture Notes in Computer Science, 12–21. Springer.