Grounding Lexical Meaning in Core Cognition

Noah D. Goodman

December 2012 (Updated June 2013)

Abstract

Author's note: This document is a slightly updated and reformatted extract from a grant proposal to the ONR. As a proposal, it aims describe useful directions while reviewing existing and pilot work; it has no pretensions to being a systematic, rigorous, or entirely coherent scholarly work. On the other hand, I've found that it provides a useful overview of a few ideas on the architecture of natural language that haven't yet appeared elsewhere. I provide it for those interested, but with all due caveats.

Words are potentially one of the clearest windows on human knowledge and conceptual structure. But what do words mean? In this project we aim to **construct and explore a formal model of lexical semantics grounded, via pragmatic inference, in core conceptual structures**. Flexible human cognition is derived in large part from our ability to imagine possible worlds. A rich set of concepts, intuitive theories, and other mental representations support imagining and reasoning about possible worlds—together we call these core cognition. Here we posit that the collection of core concepts also forms the set of primitive elements available for lexical semantics: word meanings are built from pieces of core cognition. We propose to study lexical semantics in the setting of an architecture for language understanding that integrates literal meaning with pragmatic inference. This architecture supports underspecified and uncertain lexical meaning, leading to subtle interactions between meaning, conceptual structure, and context. We will explore several cases of lexical semantics where these interactions are particularly important: indexicals, scalar adjectives, generics, and modals. We formalize both core cognition and the natural language architecture using the Church probabilistic programming language. In this project we aim to contribute to our understanding of the connection between words and mental representations; from this we expect to gain critical insights into many aspects of psychology, to construct vastly more useful thinking machines, and to interface natural and artificial intelligences more efficiently.

1 Introduction

Words are potentially one of the clearest windows on human knowledge and conceptual structure. If we understood the connection between words and mental representation, we could gain critical insights into almost every aspect of psychology, construct vastly more useful thinking machines, and interface the two. But what do words mean? In this project we aim to **construct and explore a formal model of lexical semantics grounded, via pragmatic inference, in core conceptual structures**. We will do so using a set of modeling tools—chiefly the probabilistic programming language Church [32]—that we have previously developed and used to explain aspects of high-level human cognition.

Flexible human cognition is derived in large part from our ability to imagine (or sample, or simulate) possible worlds. A rich set of concepts, intuitive theories, and other mental representations support imagining and reasoning about possible worlds—together we call these **core cognition**[1]. We can formalize key pieces of core cognition (such as intuitive physics and theory of mind) using probabilistic programming tools by viewing commonsense knowledge as a set of interrelated definitions (or concepts) in a probabilistic programming language. Probabilistic programs specify sampling procedures over possible program executions, each execution fixes (some of) the variables in that describe the world, hence each collection of concepts describes how to sample from a rich space of possible worlds. The inference (or conditioning, or query) operator describes how to use these distributions on worlds for reasoning. We have been able to explain many inferences that humans draw from sparse evidence using these tools. Here we posit that this collection of concepts also forms the set of primitive elements available for **lexical semantics**: word meanings can be built from the pieces of core cognition.

However, the connection between core concepts and lexical semantics is not direct: first, because language must flexibly adapt to the context of communication, the connection between lexical representation and interpreted meaning is mediated by pragmatic inference; second, sentence meanings act as constraints on possible worlds, via their truth-value, whereas core concepts describe generation of possible worlds. We propose to explore a model of language that formalizes and explains these differences, again using tools of probabilistic programming.

In recent work we have made preliminary progress at combining underspecified semantic representations with a general purpose pragmatic inference mechanism. This potentially allows us to account for subtle aspects of language use with relatively simple semantic denotations. In particular, we view semantic representation as indivisible

[1]This phrase is chosen to connote the *conceptual core* of human thinking, in contrast to the distal processes of low-level perception, attention, etc. There is no particular connection intended to the developmental claims of *core knowledge*, except for a concern with concepts and high-level cognition.

from core cognition, and the effective meaning of words as emerging from a pragmatic inference process.

Our approach is similar in spirit to cognitive semantics, in that we attempt to ground semantics in mental representation, but we draw on the highly successful tools of Bayesian cognitive science to formalize these ideas. Similarly, our approach draws heavily on the progress made in formal model-theoretic semantics [61, 34], borrowing insights about how syntax drives semantic composition, but we compose elements of stochastic logics rather than deterministic ones. Finally, our approach is related to recent progress in robotics [81] but is more systematic and systematically connected to cognitive and linguistic theory.

We propose to study lexical semantics in the setting of an architecture for language understanding that is described in section 3. We first provide, in section 2, background on the probabilistic programming language Church and on using Church to describe core cognition. In section 4 we describe several case studies of lexical semantics in this framework.

2 Background

Probabilities describe degrees of belief, and probabilistic inference describes rational reasoning under uncertainty. It is no wonder, then, that probabilistic models have exploded onto the scene of modern artificial intelligence, cognitive science, and applied statistics: these are all sciences of inference under uncertainty. But as probabilistic models have become more sophisticated, the tools to formally describe them and to perform probabilistic inference have wrestled with new complexity. Just as programming more than the simplest algorithms requires tools for abstraction and composition, complex probabilistic modeling requires new progress in model representation—*probabilistic* programming languages. These languages provide compositional means for describing complex probability distributions; implementations of these languages provide *generic* inference engines: tools for performing efficient probabilistic inference over an arbitrary program.

By providing a uniform and universal representation for probabilistic models, probabilistic programming provides a framework for unifying disparate Bayesian models of human cognition. Indeed, while Bayesian models have been extremely influential in cognitive science [e.g. 82], it is only recently that we have the tools to view the Bayesian approach as a general framework for mental representation. We next give a brief introduction to probabilistic programming, then an indication of how these tools can be used for modeling human cognition.

2.0.1 Probabilistic Programming Languages and Church

In their simplest form, probabilistic programming languages extend a well-specified deterministic programming language with primitive constructs for random choice. This is a relatively old idea, with foundational work by Giry, Kozen, Jones, Moggi, Saheb-Djahromi, Plotkin, and others [see e.g. 38]. Yet it has seen a resurgence thanks to new tools for probabilistic inference and new complexity of probabilistic modeling applications. There are a number of recent probabilistic programming languages [e.g. 73, 66, 60, 68, 67, 45, 65, 59, 44] embodying different tradeoffs in expressivity, efficiency, and perspicuity. We will focus on the probabilistic programming language Church [32], which has the benefits of being close to the core mathematical foundation (stochastic lambda calculus) yet having sufficient expressivity to easily represent abstract structures needed in cognitive modeling.

Church extends (the purely functional subset of) Scheme [1] with elementary random primitives, such as `flip` (a bernoulli), `multinomial`, and `gaussian`. In addition, Church includes language constructs that simplify modeling. For instance, `mem`, a higher-order procedure that memoizes its input function, is useful for describing persistent random properties and lazy model construction. If we view the semantics of the underlying deterministic language as a map from programs to executions of the program, the semantics of the probabilistic language is a map from programs to distributions over executions. When the program halts with probability one, this induces a proper distribution over return values. Indeed, *any* computable distribution can be represented as the distribution induced by a Church program in this way (see [25, §6], [2, §11], and citations therein).

Probabilistic programs extend probabilistic graphical models [47], aka Bayes nets, one of the most important ideas of modern AI. Indeed, graphical models can be seen as flow diagrams for probabilistic programs—and just as flow diagrams for deterministic programs are useful but not powerful enough to represent general computation, graphical models are a useful but incomplete approach to probabilistic modeling. For an example of this, we need look no further than the fundamental operation for inference, probabilistic conditioning, which forms a posterior distribution from the prior distribution. Conditioning is typically viewed as a special operation that happens *to* a probabilistic model (capturing observations or assumptions), not one that can be expressed *as* a model. However, because probabilistic programs allow *stochastic recursion*, conditioning can be defined as an ordinary probabilistic function (Fig. 1, Top). Notice that, since conditioning is an ordinary function, conditioning can be nested inside other calls to the conditioning operator. This is a pattern that we will use in defining our language models in section 3.

In Church, conditioning is specified by the more convenient `query` syntax (Fig. 1, Bottom). A Church query first gives a set of stochastic function definitions, which set

```
(define (rejection-query thunk condition)
  (let ((val (thunk)))
    (if (condition val)
        val
        (rejection-query thunk condition))))
```

```
(query
  ...defines...
  query-expression
  condition-expression)
```

```
(define (thunk)
  ...defines...
  (pair query-expression
     condition-expression))
(define condition rest)
```

Figure 1: (Top) Defining conditional inference in Church as a stochastic recursion: rejection sampling represents the conditional probability of the thunk conditioned on the condition predicate being true. We typically use special query syntax (Bottom, left), which can be desugared into a query thunk (Bottom, right).

up the "language" of the query, then gives the query expression (whose value we will return) and finally the condition expression, which must return true. We also allow `factor` statements which provide another way of constraining a query: the statement `(factor expr)` multiplies the value of `expr` (assumed to be a real number) into the probability of the current execution. This provides a convenient way to add a soft constraint on the distribution of executions (via a side-effecting operation).

2.0.2 Core Cognition: Intuitive Theories and Conceptual Structure

Church programs can be used to express a wide variety of cognitive models, capturing concepts from core cognitive domains and core concepts from many specific areas of knowledge. For a complete tutorial on using Church to model human cognition, including many examples, see `http://www.stanford.edu/~ngoodman/ProbMods.html`.

To illustrate, we consider the problem of capturing commonsense knowledge about the game tug-of-war. Figure 2a, gives Church functions specifying key concepts for this domain (though certainly not exhausting possible knowledge about the game). This "conceptual library" of probabilistic functions can be used to reason about many patterns of evidence, via different queries, without needing to specify ahead of time the set of people, the teams, or the matches. A Church program provides a description of how to go about simulating a possible world in the domain (here, randomly choosing strengths, laziness, etc., and computing the winner of each match). To reason from evidence or hypotheses, this simulation process is directed: a query describes what assumptions to fix and what to simulate (for instance, fixing match outcomes and simulating strengths).

(a)

```
;;Strength is a persistent (memoized) property of each person:
(define strength (mem (lambda (person) (gaussian 1.0 1.0))))

;;Laziness varies from match to match:
(define (lazy person) (flip 0.3))

;;When a person is lazy they pull less:
(define (pulling person) (if (lazy person)
                             (/ (strength person) 2)
                             (strength person)))

;;Total pulling of the team is the sum:
(define (total-pulling team) (sum (map pulling team)))

;;The winning team pulls hardest:
(define (winner team1 team2)
  (if (< (total-pulling team1) (total-pulling team2))
      team2
      team1))
```

(b)

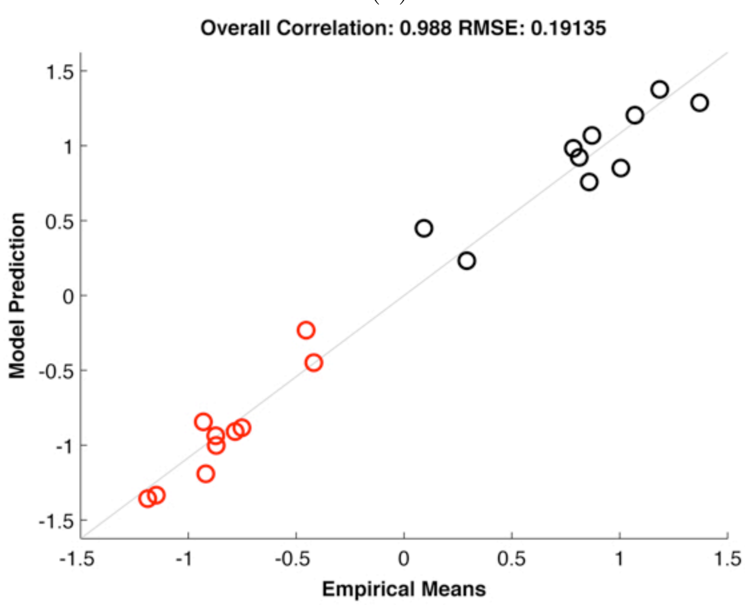

Figure 2: Modeling intuitive concepts in the tug-of-war domain. (a) A Church model to capture core concepts. While this model is simple, probabilistic queries can explain human reasoning from diverse evidence with high quantitative accuracy. (b) Comparison of the model to human judgements in Experiment 1 of [27].

The predictions of this simple model match human intuitions quite well, Figure 2b. In [27] we presented people with the results of a set of matches in a "ping-pong tournament", and asked for judgements about the strength of one of the players. We found a correlation of 0.98 between model and human judgements. This compelling fit suggests that the definitions in Figure 2a capture important aspects of the concepts people have for reasoning about team games. However, connecting these concepts to natural language would provide a much more natural probe of people's intuitions. For instance, we would like to ask *Is Bob strong?* rather than asking for an arbitrary rating of Bob's strength; but how does the adjective *strong* in this positive form relate to the conceptual degree of `strength`? The connection must be indirect since, as discussed in section 4.0.7, a statement like *Bob is strong* can be interpreted very differently in different contexts. More generally, how may we use concepts defined in a Church model for a given domain to build the semantics of natural language for this domain?

3 The Architecture of Natural Language

In this section we give an overview of our framework for modeling natural language understanding. This framework provides the architecture into which lexical semantics fits; it is thus the enabling theory for the entire project. We first describe the role of literal meanings in updating beliefs, then describe the pragmatic enrichment of meaning through inference, finally we extend this framework to handle context-specific semantic denotations via free variables.

3.0.3 L0 Model: Literal meaning

The basic assumption of our probabilistic model of language interpretation is that sentences can be used to update the listener's belief distribution. Because probabilistic belief update is performed by *conditioning* a prior distribution to get a posterior distribution, the meaning of a sentence should be a conditioning expression—an expression that evaluates to true or false. If we assume further that a listener's prior knowledge about the world is given by her concept definitions, then the literal interpretation of language can be given by the Church function:

```
;;The basic listener: how is the world, given that the utterance is true?
(define (L0 utterance QUD)
  (query
    ...core concept definitions...
    (eval QUD)
    (eval (meaning utterance))))
```

This describes a listener `L0` who hears `utterance`, and is interested in the import for a given question-under-discussion[2] (`QUD`). The (meta-)operator `eval` evaluates a given expression in the current environment. Here this means evaluating both the `QUD` expression and the meaning of the utterance in an environment in which the core concept definitions are visible. The result is a posterior distribution for each possible `QUD`—hence a posterior distribution over possible worlds.

Meaning Composition We assume for simplicity that the utterance has already been parsed into a syntactic tree. We don't address this syntactic parsing problem in this project, but instead assume it can be effectively handled by existing tools. Following standard practice of formal semantics [34] we construct the meaning of the sentence recursively along the syntactic tree:

```
;;Meanings are constructed by recursive composition along the (syntactic) tree:
(define (meaning utterance)
  (if (lexical-item? utterance)
      (lexicon utterance)
      (pair (meaning (left utterance)) (meaning (right utterance)))))
```

Here the predicate `lexical-item?` determines if the remaining utterance is a single lexical item (entry in the lexicon), the function `left` returns the left subtree of the utterance (which we assume is the operator—possibly through a syntactic transformation), and `right` returns the right subtree. By combining meanings with the `pair` operator, we are implying that composition happens by function application: when the meaning is evaluated (in `L0`) the left subtree meaning will be applied to the right subtree meaning. Of course, a more systematic treatment of syntax and composition, such as CCG [78], could fill in the `meaning` function.

Lexical Meanings The function `lexicon` looks up the meaning of a word, returning an expression that can be evaluated in the current environment. Because this environment contains core concept definitions, the meanings of words can (and will) refer to non-linguistic mental representations: meanings of words are about the possible

[2]Conceptually, we think of the listener as forming a distribution over possible worlds. However worlds are unwieldy to represent (being infinitely large), so we represent this distribution instead by a function that can give the distribution over answers to every question that can be asked in a world. This QUD is a notion used widely in formal models of discourse [83, 28, 29, 69, 70, 9].

worlds that a listener considers. The lexicon is described in detail in section 4. (In later sections use the standard notation ⟦*word*⟧ for `(lexicon` *word*`)`.)

3.0.4 L1 Model: Pragmatic enrichment

The literal meaning, as encoded by `lexicon` and interpreted by `L0`, forms the stable contribution of words to meaning across contexts; however meaning can often be strengthened or changed in particular communicative contexts by pragmatic inference. Here we model pragmatic enrichment of the literal meaning following [22, 31, 79]. Critically, because `query` is an ordinary function that may be nested in itself, we are able to model a listener reasoning about a speaker, who reasons about a literal listener. This model formalizes the idea that a listener is trying to infer what the speaker intended, while a speaker is trying to make the listener form a particular belief.

The reflective speaker makes a speech act in order to lead the listener to infer a particular value of the question under discussion:

```
;;The speaker: what should I say, so that the listener forms the right
    interpretation?
(define (S1 val QUD)
  (query
    (define utterance (language-prior))
    utterance
    (equal? val (L0 utterance QUD))))
```

This function describes a soft-max optimal speaker whose goal is for the literal listener to arrive at a given interpretation. The `language-prior` forms the *a priori* (non-contextual and non-semantic) distribution over linguistic forms, which may be modeled with a probabilistic context free grammar or similar model. This prior inserts a *cost* for each utterance: using a less likely utterance will be dispreferred *a priori*.

The pragmatic listener can now be modeled as a Bayesian agent inferring the value of the question under discussion, given that the reflective speaker has bothered to make a particular speech act:

```
;;The pragmatic listener: what value does the QUD have, given that a speaker chose
    this utterance to express it?
(define (L1 utterance QUD)
  (query
    ...core concept definitions...
    (define val (eval QUD))
    val
    (equal? utterance (S1 val QUD))))
```

This model gives rise to standard scalar implicatures (e.g. *some* implies *not all*) and has been shown to predict human judgements with high quantitative accuracy in several language understanding tasks [22, 31]. Ongoing research aims to further explore the ability of this model to predict human behavior in reference games and simple language understanding tasks.

3.0.5 L1-sv Model: Free semantic variables

```
(define (L0 utterance QUD sv)
  (query
    ...core concept definitions...
    (eval QUD)
    (eval (meaning utterance))))

(define (S1 val QUD sv)
  (query
    (define utterance (language-prior))
    utterance
    (equal? val (L0 utterance QUD sv))))

(define (L1-sv utterance QUD)
  (query
    ...core concept definitions...
    (define sv (semvar-prior))
    (define val (eval QUD))
    val
    (equal? utterance (S1 val QUD sv))))
```

Figure 3: A probabilistic model of natural language understanding incorporating a literal listener `L0`, a reflective speaker `S1`, and a pragmatic listener `L1-sv` who reasons about the question under discussion and the value of free semantic variables.

Under standard linguistic analysis, the literal meaning of a sentence frequently leaves some aspects underspecified; to assign a complete meaning to the sentence, pragmatic inference is required to fill in the value of these *free variables*. This occurs, for example, in sentences such as *He is drinking a martini*, which cannot receive a determinate meaning until the intended referent of *He* is identified (see section 4.0.6). It is believed that the same holds for *Sam is tall*, since literal meaning underdetermines the height required to count as *tall* (see section 4.0.7); or again for *Moose have horns*, where the semantics does not determine what proportion of moose must have horns in order for this sentence to be true (see section 4.0.8). This type of context-dependence is widespread in language, but a general and precise framework for understanding how speakers and listeners make use of and resolve underspecified meanings has not previously been proposed.

In our approach, semantic context-dependence is connected with pragmatic inference by instantiating semantic variables at the `L1` level and passing them down to lower levels. The complete model is shown in Figure 3. In this model, a reflective listener `L1-sv` evaluates candidate resolutions of the free semantic variables `sv` jointly with possible values for the `QUD`. As we describe in section 4, this results in an interaction that produces powerful context-sensitive usage of words despite relatively impoverished semantic representations.

For a fully compositional theory of natural language semantics, a mechanism is needed to account for the fact that expressions with free semantic variables (such as the pronoun *he*) can occur in arbitrarily embedded positions in a sentence. Notice that when the `QUD` and the meaning of the utterance are evaluated in `L0` of this model, they are evaluated in an environment in which the semantic variables (`sv`) are bound variables (because these are arguments to the `L0` function). These variables are thus free with respect to the lexical entries, but not with respect to the full language model—they have been filled in by the pragmatic listener model. Using an environment to bind the free variables in this way is similar to the approach of relativizing the meaning function to a fixed set of parameters [57, 58, 7]. An alternative approach would be to adopt the assumptions about compositionality associated with *variable-free semantics* [80, 77, 36, 8]. In this approach, the standard mechanism of function application is augmented, making it possible to lift a free variable out of deeply embedded syntactic positions, to the top level of literal meaning.

4 Lexical Semantics: Case Studies in Flexible Meaning

The language architecture described above predicts the interpretation of an utterance given its semantic content and context; fixing the semantic content of particular (classes of) words makes this abstract architecture concrete by grounding language use into the concepts of core cognition. In this section we describe our proposal to study several classes of words as representative case studies in which concepts, uncertainty, and inference interact in important ways. We place particular emphasis on the cases in which free semantic variables are needed to allow contextual flexibility in semantic interpretation.

4.0.6 Indexicals and referring expressions

Perhaps the simplest case where semantic free variables are needed is in the meanings of referring expressions. For instance, the meaning of *He is drinking a martini* can be taken to be (approximately): $\mathrm{male}(x) \wedge \mathrm{drinking}(x, y) \wedge \mathrm{martini}(y)$, where x and y are free semantic variables that range over individual entities. The semantic contribution of *he* to this meaning is two-fold: it introduces a free variable, and it constrains that free variable to be a male. However, in order to compose properly with the surrounding sentence, the return value of *he* must be only the variable x. We can add the constraint, without affecting the return value, by using a Church factor statement:

$[\![he]\!]$=`(begin (factor (if (male x) 1.0 0.1)) x)`.
This definition adds the constraint as a side-effect, then goes on to return the value of the free variable `x`. Notice that the constraint that *he* be male is a soft constraint: in a situation where no male referent for x is plausible, it can be interpreted as a non-male. (For instance, in a situation in which we are discussing the best way to cook a burger the sentence *Wait until he's just brown, then flip him* can be interpreted as referring to the burger.) Interpretation of the pronoun is guided by the semantics, but driven by probabilistic inference of the sentence meaning (similar to the conclusion of Kehler et al. [41]).

The constraints in this approach are determined by the conceptual content, but so are the potential referents: the possible values of x are not entities in the world, but entities in the possible world represented by the listener `L1-sv`. This is particularly striking in the case of indexicals such as *I/here/now/that*. The free variable for *now* ranges over times, that for *I* over people, and each must have meta-access to the speech act itself. For instance,

$[\![I]\!]$=`(begin (factor (if (eq? x (speaker-of utterance)) 1.0 0.1)) x)`;
where we have made use of the fact that the utterance is bound in the environment in which $[\![I]\!]$ is evaluated (being an argument to `L0`), and assumed that the `speaker-of` function returns the individual who performed a speech act.

Horn's *division of pragmatic labor* [35] presents a more subtle case for the pragmatic fixing of reference. This principle dictates that, in the absence of distinguishing semantic content, simple utterances should be interpreted as un-marked (i.e. simple/probable/good) meanings and complex utterances as marked (i.e. deviant/unlikely/bad) meanings. For instance *I got the car started* is interpreted as doing something unusual to start the car, since the simpler phrase *I started the car* is available to be interpreted as simply turning the ignition. To formalize this, imagine cheap/costly utterances *u1*/*u2* and *a priori* likely/unlikely interpretations of the QUD `val1`/`val2`. We would like the less costly expression to be interpreted as the more likely interpretation: $[\![u1]\!]$=`val1`, $[\![u2]\!]$=`val2`. This is a simple signaling game [23], and this solution is a Nash equilibrium; however, there are many equally good Nash equilibria for this game. A number of attempts have been made to explain why this particular equilibrium should be chosen, but all require ad-hoc stipulations on the equilibrium concept. Indeed, the simple model `L1`, with no semantic free variables, is unable to arrive at the correct interpretations. We have shown in [10] that the `L1-sv` model does arrive at the correct interpretations in this case. Many questions and difficulties remain to integrate this result with lexical semantics more generally. In particular, how does the division of labor play out in situations with similar but not equivalent semantic meaning, and what semantic work can we rely on this effect to do?

We propose to use the `L1-sv` framework with constrained entity variables to study the details of how referring expressions can be flexibly interpreted. As illustrated above, complex interpretation patterns can be expected as context, concepts, and inference interact. We will explore the ways that context allows semantic constraints to be violated, the conceptual grounding of indexicals, the effect of alternative referring expressions, and the use of division of pragmatic labor in simplifying semantic denotations.

4.0.7 Scalar Adjectives

Many words resist simple truth-functions that could be used to tell with certainty when they hold. For instance, exactly how tall must a person be to be a *tall person*? Vagueness of meanings has been the subject of much discussion in philosophy and linguistics, and is of critical importance in the psychological literature on graded concepts [71, 63, 4, 40, 21, 33]. Vagueness is particularly clear in the meanings of gradable adjectives, such as *tall/short*, *wide/narrow*, *happy/sad*, *wet/dry*, and *full/empty*. These adjectives serve to express measurement along a scale (*Sam is six feet tall*); they are grammatically gradable (*Sam is very tall*), are typically vague, and are highly context-sensitive [e.g. 39, 20, 17, 85, 86, 43, 42].

Vagueness can be seen from the existence of *borderline cases*: individuals for whom it is unclear whether an adjective applies. A seven-foot-tall man is obviously tall, and a five-foot-tall man is obviously not; but does a man who is $6'2''$ count as "tall"? When asked a question of this type, speakers typically express uncertainty, and show disagreement in responses to a forced-choice [12, 74, 84, 3]. The vagueness of adjectives appears to be related to statistical properties of a *comparison class* [11, 74, 76], which can be implicit in the context or provided explicitly as in: *Sam is rich for a janitor/politician* or *Michael Jordan is tall for a man/basketball player*.

Our approach to gradable adjectives starts with a scalar theory of their lexical semantics. We adopt a degree semantics in which adjectives relate individuals to a threshold value[3]: $[\![A]\!] = \lambda x[\mu_A(x) > \theta_A]$, where θ_A is a free threshold variable on A's scale and $\mu_A(x)$ is the degree of x on this scale. We will further assume that the function $\mu_A(x)$ is a concept defined in core cognition. For example, if `height` is a persistent property of an individual (here drawn from a fixed gaussian for simplicity), the basic meaning of *tall* is:

```
;;core concept definitions:
(define height (mem (λ (x) (gaussian 6.0 1.0))))
...

;;lexicon:
⟦tall⟧ = (λ (x) (> (height x) θ))
```

[3] Or rather: $[\![A]\!]$= `(λ (x) (> (`μ_A` x) `θ`)`, but we have used the more familiar mathematical notation where it is clearer.

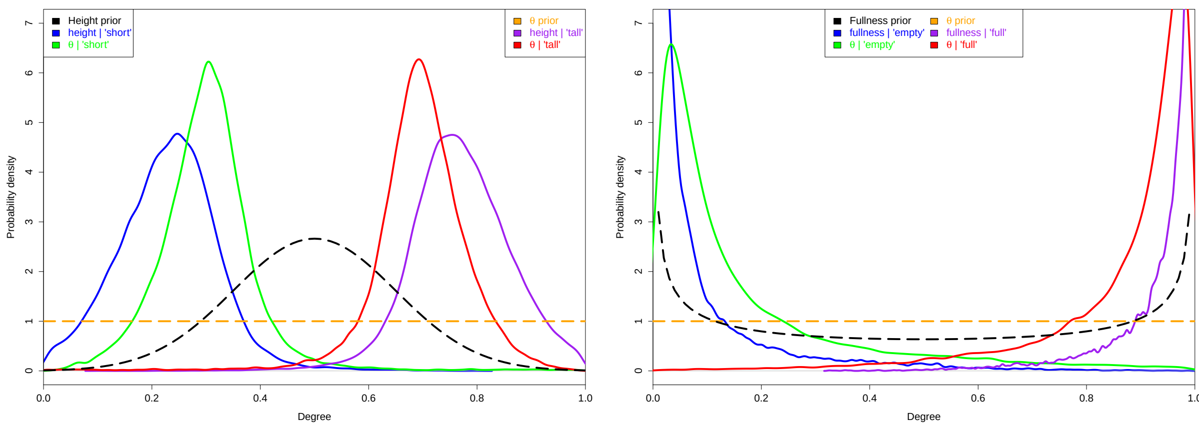

Figure 4: Predictions of the `L1-sv` model for a threshold semantics for adjectives, with prior distribution over degrees appropriate to *tall/short* and *full/empty*. These simulations used Markov Chain Monte Carlo to draw 30,000 samples from the joint posterior on degree and θ, with $\alpha = 4$, the utterance prior of $u \propto \text{length}(u)$, a burn-in of 5000 samples, and a lag of 100. Plots show the kernel density of the relevant variables. The alternative utterances considered are to say nothing or to use the positive (e.g. *tall*) or negative (e.g. *short*) adjective.

Crucially, θ is left as a free variable in the semantic representation. The meaning of *Sam is tall*, then, is simply that Sam's height is at least θ; the task of inferring θ in order to derive a meaning for this sentence will be preformed by the full interpretation model, `L1-sv`. This approach to the interpretation of vague adjectives captures the insights of previous probabilistic accounts [18, 74, 24, 51], but improves on them in several ways, notably in providing a clear lexical semantics from which the probabilities follow.

To illustrate, imagine a situation in which a speaker is attempting to communicate Sam's height to a listener who does not know how tall Sam is, but knows that Sam is an adult male. The `QUD` is `(height 'sam)`. We assume that listener and speaker share the common prior knowledge that heights of people are approximately normally distributed. The meaning of the sentence *Sam is tall*, as constructed by the `meaning` function, will be simply `(> (height 'sam) theta)`. The listener `L1-sv` will then do a joint inference of Sam's height and the threshold variable `theta`.

If the value of `theta` is very low (e.g., one foot), the utterance *Sam is tall* is extremely uninformative: the listener already has a strong belief that Sam is more than one foot tall. On the other hand, if `theta` is extremely high, the utterance will be extremely informative since the prior probability that Sam's height is greater than `theta` is very low. This means that *Sam is tall* was much more likely to be uttered by `S1` if the value of `theta` is high. But this pressure to infer high `theta`, and hence large heights, is balanced by the low prior probability of very large heights. The posterior distribution on heights given the utterance (shown in Figure 4) reflects this balancing process. In effect, interpretations are preferred which make Sam significantly taller than average, but not implausibly tall. Our model thus gives precise form to an intuition about the meaning of scalar adjectives which has been stated repeatedly in the linguistic and philosophical literature [e.g. 11, 19, 42].

There is an immediate explanation in this approach of the context-dependence of scalar adjectives: different comparison classes have different prior distributions, which affects the joint inference carried out by `L1-sv`. This can explain not only quantitative shifting of the threshold depending on context (e.g. tall man vs. tall basketball player), it can also explain qualitative differences between different types of adjectives. For instance, absolute adjectives, like *full/closed*, also refer to a degree along some scale, but behave differently than relative adjectives (like *tall*): absolute adjectives compare a degree to a (fairly) fixed extreme point [72, 43, 42][4]. For instance, a closed door is not one that's more closed than average, it's *closed*. Prior distributions which have significant probability mass near the edges of a scale result in very different interpretation than those with thin tails (Figure 4 left vs right): the interpreted meaning is strongly peaked near the extreme point of the degree. The difference between relative and absolute adjectives can thus be explained by qualitative differences in the prior distributions which they invoke. These differences in prior distribution are a consequence of non-linguistic knowledge represented in concepts of the domain.

Sorites paradox The much-discussed *sorites paradox* is a key puzzle of gradable adjectives, for example:

(1) Sorites paradox:
  a. A man who is 7 feet tall is tall.
  b. A man who is 0.01 inches less tall than someone who is tall is also tall.
  c. Therefore, a man who is 4 feet tall is tall.

This argument is logically valid, and the premises 1a and 1b appear to be reasonable; but the conclusion 1c is clearly false. An account of vagueness should explain both the logical puzzle (which premise is incorrect?) and the psychological problem (why do people find the premises so compelling, while also maintaining that the conclusion is completely implausible? [Cf. 19]).

Our approach to adjective interpretation offers a new account of the sorites paradox. The argument is logically unsound because the second premise 1b is not *strictly true*:

[4]There are also differences in modification patterns between relative and absolute adjectives. For example, we can modify *full* as *completely full* but we cannot usually say not *completely tall*.

if x's height is very close to the cutoff, θ, then x may count as "tall" while someone just slightly shorter does not. However, the second premise is *highly plausible* because the probability that θ and x's height will be so close is small. Indeed, if we choose the height h of x and the threshold θ from the joint posterior distribution that `L1-sv` defines given the utterance x *is tall*, the probability that $h - \theta > 0.01$ (i.e. premise 1b holds) is approximately 95%. This explains the psychological puzzle described above: why is the second premise so compelling, if it is not strictly true? The answer is that this premise has high probability (a fact which does not justify its repeated use as a premise in logical arguments [49]).

Absolute adjectives are different: the second premise is less intuitively plausible for an adjective like *closed* ("A door that is 0.1 inches less closed than a closed door is also closed") [42]. We suggest that this is essentially because the posterior variance of the threshold is lower for absolute adjectives. In simulations using a prior peaked slightly near the boundary (as in Figure 4 right), the second premise has much lower probability (65%). Our model thus suggests that the difference in sorites susceptibility between relative and absolute adjectives may be a difference in the degree of uncertainty after hearing the adjective statement, rather than a qualitative difference between kinds of uncertainty.

4.0.8 Generics and Quantifiers

Generic sentences are a ubiquitous way of communicating about the properties of categories [13]:

(2) a. Birds fly.
  b. Ducks lay eggs.
  c. Mosquitoes carry West Nile Virus.

The meaning of generic sentences has been a longstanding puzzle to psychologists, philosophers, and linguists alike [13, 48, 15, 16, 64, 55, 26, 14]. A first guess might be that "Birds fly" means that *all* birds fly; but this is not right, since the existence of penguins does not make 2a false. Weakening the meaning to *most* or *usually* is not sufficient either, in light of 2b-2c: fewer than half of ducks ever lay eggs (the female ones who survive and reproduce), and only a small proportion of mosquitoes carry West Nile Virus. Assuming a maximally weak meaning (*some*) seems like the only truth-functional meaning consistent with 2a-2c; but this is problematic as well: experimental results indicate that, in the absence of prior knowledge, people usually infer from a generic sentence like 2a that the property is highly prevalent [14]. In contrast to these semantic complications, generics seem to be one of the *simplest* linguistic constructions on many other dimensions: they require no explicitly marked operator (contra quantifiers like *some/all*), they are acquired early and used abundantly in child-directed speech [26], and they have high frequency in everyday speech.

What is needed, it seems, is a meaning for generic sentences that makes them semantically simple—for instance by explaining the subtleties of usage through pragmatic inference. Something along the following lines: *Birds fly* is true if the rate of flying among birds is greater than what you would expect if you did not know anything about birds; *Ducks lay eggs* is true if the rate of egg-laying among ducks is greater than that of animals in general; and so on. This kind of sensitivity to prior expectations is reminiscent of the analysis of gradable adjectives in section 4.0.7 above.

We propose to treat generic statements in a way that parallels scalar adjectives, with the scale being *probability*: the generic statement *Kind Property* imposes a (free variable) threshold on the probability of the property holding for members of the kind:

$[\![\textit{Kind Property}]\!] = P(\mathrm{Prop}(x) \mid x \sim \mathrm{Kind}) > \theta$.

That is, the probability that an object drawn from the distribution over objects *Kind* has property *Prop* is greater than an underspecified threshold θ. If we wish to encode this purely in terms of the sampling semantics of Church (i.e. not requiring a reflection operator that exposes the probability directly), we could write:

$[\![\textit{Kind Property}]\!]$=`(all (repeat theta (λ () (Prop (Kind))))).`

That is: draw `theta` samples by sampling a `Kind` object and evaluating `Prop`, return true if all samples are true. (The two versions of the semantics, explicit probability threshold and repeated sampling, are equivalent in expectation up to a transformation of the free variable.)

This probability threshold semantics for generics results in a strong interpretation when there is no prior knowledge: Figure 5a shows that with uniform prior over property rate the generic is only endorsed when the property is usually true for this kind (i.e. a meaning close to *all*, though tolerating exceptions). However, background knowledge about the property can radically alter interpretation of the generic. Figure 5b shows that if the property is believed *a priori* to be rare, then the generic can be endorsed even if the property fairly infrequent for this kind. This is reminiscent of *mosquitos have West Nile virus*, assuming background knowledge that carrying a disease is rare even when possible. Figure 5c shows the predictions for a case like *birds fly*, where naive reasoners assume biological characteristics of animals to be *homogeneous* [see 62]: an animal kind will have either a very high or very low rate of most properties. In this situation the interpretation of the generic is very strong—again close to *all*. Finally, what of cases like *Ducks lay eggs*? Figure 5d shows an appropriate prior for egg-laying: at most half of animals in a kind do it. The prediction is an interpretation requiring around half of animals in the kind to lay eggs.

These preliminary results are encouraging; we propose to build a more complete model of generics using these techniques, and evaluate it against the significant empirical literature on generic usage. Having verified this approach to generics, we will have a sharp tool for exploring people's intuitions about kinds and properties.

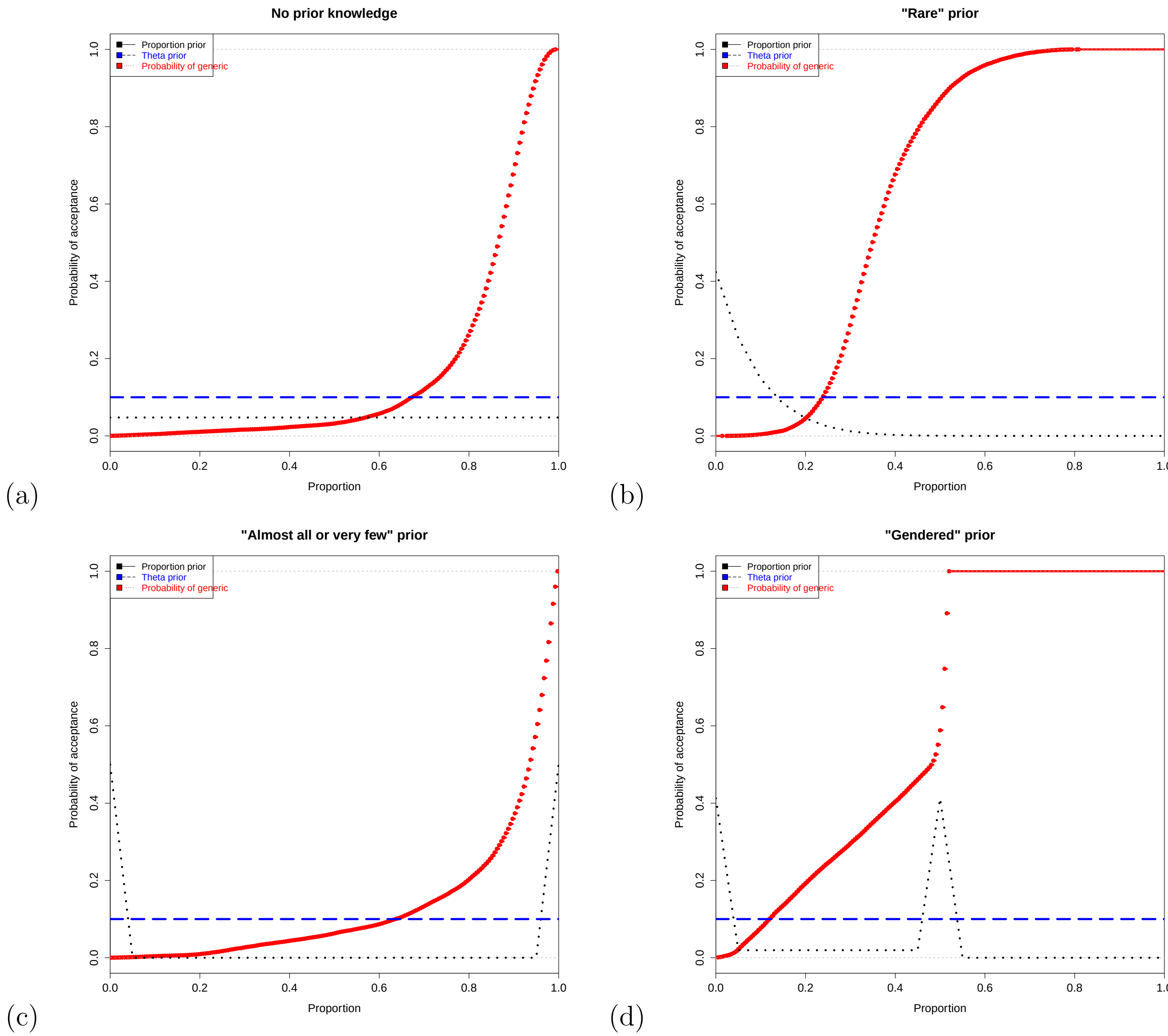

Figure 5: Prediction of the generics model for four different prior expectations about frequencies of the property.

Quantifiers The standard linguistic analysis of quantifiers is as deterministic logical operators on sets. However, the above analysis for generics suggests an interesting analogy: Modifiers for scalar adjectives can strengthen meaning, e.g. *very tall*, and even squeeze out much vagueness, e.g. *completely closed*. What if quantifiers are analogous modifiers for the generic, which we have given a scalar interpretation? That is, perhaps *all* plays a similar role in *all dogs have fur* that *completely* does in *the door is completely closed*. This would explain why generics appear as the un-marked form of quantification, and would explain why even strong quantifiers like *all* are found experimentally to have some slack [75]. However a number of potential problems with this semantics for quantifiers must be explored. For instance, how does this modifier semantics explain scope ambiguity? What predictions follow about vagueness and slack in quantifier meaning, and are these empirically viable?

4.0.9 Modal Verbs and Adverbs

Modals are a class of linguistic expressions whose meanings are deeply bound up with reasoning about beliefs and desires, for instance:

(3) a. Mary <u>wants</u> to have a birthday party.
b. I <u>believe</u> it is <u>likely</u> that she will.
c. Mary doesn't <u>want</u> John to <u>know</u> about her party.

Modal language offers an opportunity to draw close connections between semantics and the intuitive theories of belief, desire, observation, etc. (that is Theory of Mind). Recent research has formulated a rich probabilistic framework for theory of mind [5, 30, 6, 37]; at the same time, semantic research on modality has begun to move toward representations based on related tools from probability and decision theory [56, 87, 88, 89, 50, 53, 52, 46]. The latter line of research emphasizes the graded nature of belief and desire, but has not drawn systematic connections with the relevant cognitive science research—rather than inheriting these structures from theories of mental representation, they posit them as part of semantics. In our approach the meanings of linguistic expressions of uncertainty, desire, and causation are not *sui generis*, but are defined in terms of the concepts that agents use to reason about their own and others' actions and motivations.

The key technical idea is that the intuitive theory of mind provides for each person a distribution over worlds. Rather than reifying worlds directly, we can represent beliefs as mappings from expressions (the question of interest, `QOI`) to values. Thus Bob's beliefs, `(beliefs 'bob)`, will be a function from expressions to values, and `((beliefs 'bob) '(sky-is 'blue))` represents (the probability that) Bob believes the sky is blue. The belief function for a rational Bayesian agent can be written:

```
(define (beliefs person)
```

```
(λ (QOI)
  (query
    ...intuitive theories...
    (eval QOI)
    (observations person))))
```

That is, the agent conditions on their observations in the context of their intuitive theories (prior knowledge) to form a posterior distribution over the question of interest. We do not assume that all agents have rational beliefs, however, only that their beliefs can be represented as a similar distribution.

Using this core notion of belief, we can understand the semantics of *believes* as: $[\![believes]\!]=$ `(λ (person expr) (all (repeat theta (λ () ((beliefs person) expr)))))`.
That is, *Bob believes expr* is a scalar construction, much like generics in section 4.0.8, that requires Bob's degree of belief in `expr` to be above a (semantics free variable) threshold. (Here we have implemented the threshold by taking `theta` samples from Bob's belief distribution and requiring that `expr` be true for each.) When the literal listener, `L0`, conditions on a belief statement, it places a constraint on the belief distribution of the agent in question.

The intuitive theory of mind is not propositional in any important way: it describes beliefs as a generative distribution on worlds (given observations, etc). However, the semantics of *X believes Y* constructs a constraint on this belief distribution out of the proposition *Y*. In this way, belief *language* is propositional, while belief *representations* themselves are non-propositional—providing an interesting take on the question of propositional attitudes in theory of mind.

We propose to explore and extend this semantics of belief terms, and to integrate it with our previous work on epistemic modals (*plausible/likely/certain/...*) [54]. We further plan to explore modals of desire (e.g. *wants*), and the embedding of these modals in each other—capturing the meanings of statements like 3a-c, above.

References

[1] H. Abelson and G. Sussman. *Structure and Interpretation of Computer Programs.* MIT Press, 1996.

[2] N.L. Ackerman, C.E. Freer, and D.M. Roy. Noncomputable conditional distributions. In *Logic in Computer Science (LICS), 2011 26th Annual IEEE Symposium on*, pages 107–116. IEEE, 2011.

[3] S. Alxatib and F.J. Pelletier. The psychology of vagueness: Borderline cases and contradictions. *Mind & Language*, 26(3):287–326, 2011.

[4] S.L. Armstrong, L.R. Gleitman, and H. Gleitman. What some concepts might not be. *Cognition*, 13(3):263–308, 1983.

[5] C.L. Baker, R. Saxe, and J.B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.

[6] C.L. Baker, R.R. Saxe, and J.B. Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, 2011.

[7] C. Barker. The dynamics of vagueness. *Linguistics and Philosophy*, 25(1):1–36, 2002.

[8] C. Barker and P.I. Jacobson, editors. *Direct compositionality*. Oxford University Press, 2007.

[9] David Beaver and Brady Clark. *Sense and sensitivity: How focus determines meaning*. Wiley-Blackwell, 2008.

[10] L. Bergen, N.D. Goodman, and R. Levy. That's what she (could have) said: How alternative utterances affect language use, 2012. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.

[11] M. Bierwisch. The semantics of gradation. In M. Bierwisch and E. Lang, editors, *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*, pages 71–261. Springer-Verlag, 1989.

[12] N. Bonini, D. Osherson, R. Viale, and T. Williamson. On the psychology of vague predicates. *Mind & language*, 14(4):377–393, 1999.

[13] G.N. Carlson. *Reference to kinds in English*. PhD thesis, 1977.

[14] A. Cimpian, A.C. Brandone, and S.A. Gelman. Generic statements require little evidence for acceptance but have powerful implications. *Cognitive science*, 34(8): 1452–1482, 2010.

[15] A. Cohen. Generics, frequency adverbs and probability. *Linguistics and Philosophy*, 22:221–253, 1999.

[16] A. Cohen. *Think generic! The Meaning and Use of Generic Sentences*. CSLI, 1999.

[17] M.J. Cresswell. The semantics of degree. In B. Partee, editor, *Montague grammar*, pages 261–292. Academic Press, 1976.

[18] D. Edgington. Vagueness by degrees. In R. Keefe and P. Smith, editors, *Vagueness: A Reader*, pages 294–316. MIT Press, 1997.

[19] D.G. Fara. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, 20:45–81, 2000.

[20] K. Fine. Vagueness, truth and logic. *Synthese*, 30(3):265–300, 1975.

[21] J. Fodor and E. Lepore. The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition*, 58(2):253–270, 1996.

[22] M.C. Frank and N.D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

[23] M. Franke. *Signal to act: Game theory in pragmatics.* PhD thesis, Institute for Logic, Language and Computation, 2009.

[24] J. Frazee and D. Beaver. Vagueness is rational under uncertainty. *Proceedings of the 17th Amsterdam Colloquium*, 2010.

[25] Cameron E. Freer and Daniel M. Roy. Computable de Finetti measures. *Annals of Pure and Applied Logic*, 163(5):530–546, 2012. doi: 10.1016/j.apal.2011.06.011. URL `http://www.sciencedirect.com/science/article/pii/S0168007211000868`.

[26] S.A. Gelman. Learning words for kinds: Generic noun phrases in acquisition. *Weaving a lexicon*, pages 445–484, 2004.

[27] T. Gerstenberg and N. D. Goodman. Ping pong in Church: Productive use of concepts in human probabilistic inference. In *Proceedings of the 34th annual conference of the cognitive science society*, 2012.

[28] J. Ginzburg. Resolving questions, I. *Linguistics and Philosophy*, 18(5):459–527, 1995.

[29] J. Ginzburg. Resolving questions, II. *Linguistics and Philosophy*, 18(6):567–609, 1995.

[30] N. D. Goodman, C. L. Baker, and J. B. Tenenbaum. Cause and intent: Social reasoning in causal learning. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2009.

[31] N.D. Goodman and A. Stuhlmüller. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 2013.

[32] N.D. Goodman, V.K. Mansinghka, D.M. Roy, K. Bonawitz, and J.B. Tenenbaum. Church: A language for generative models. In *In UAI*, 2008.

[33] J.A. Hampton. Typicality, graded membership, and vagueness. *Cognitive Science*, 31(3):355–384, 2007.

[34] I. Heim and A. Kratzer. *Semantics in generative grammar*. Blackwell, 1998.

[35] L. Horn. Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In D. Schiffrin, editor, *Meaning, form, and use in context*, volume 42. Georgetown University Press, 1984.

[36] P. Jacobson. Towards a variable-free semantics. *Linguistics and Philosophy*, 22 (2):117–185, 1999.

[37] A. Jern, C.G. Lucas, and C. Kemp. Evaluating the inverse decision-making approach to preference learning. In *Advances in Neural Information Processing Systems: Proceedings of the 2011 conference*, 2011.

[38] C. Jones and G.D. Plotkin. A probabilistic powerdomain of evaluations. In *Logic in Computer Science (LICS), 1989 4th Annual IEEE Symposium on*, pages 186–195, Jun 1989. doi: 10.1109/LICS.1989.39173.

[39] H. Kamp. Two theories about adjectives. In E. Keenan, editor, *Formal semantics of natural language*, pages 123–155. Cambridge University Press, 1975.

[40] H. Kamp and B. Partee. Prototype theory and compositionality. *Cognition*, 57 (2):129–191, 1995.

[41] A. Kehler, L. Kertz, H. Rohde, and J.L. Elman. Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44, 2008.

[42] C. Kennedy. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, 2007.

[43] C. Kennedy and L. McNally. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81(2):345–381, 2005.

[44] A Kimmig, B Demoen, L De Raedt, V S Costa, and R Rocha. On the implementation of the probabilistic logic programming language ProbLog. *Theory and Practice of Logic Programming*, 11(2-3):235–262, 2011.

[45] Oleg Kiselyov and Chung-chieh Shan. Embedded probabilistic programming. In *Domain-Specific Languages*, pages 360–384, 2009.

[46] P. Klecha. Positive and conditional semantics for gradable modals, 2012. *Proceedings of Sinn und Bedeutung 16*.

[47] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[48] M. Krifka, F.J. Pelletier, G.N. Carlson, A. ter Meulen, G. Chierchia, and G. Link. Genericity: An introduction. *The generic book*, pages 1–124, 1995.

[49] H.E. Kyburg. *Probability and the logic of rational belief*. Wesleyan University Press Middletown, 1961.

[50] D. Lassiter. Gradable epistemic modals, probability, and scale structure. In Li and Lutz, editors, *Semantics and Linguistic Theory (SALT) 20*, pages 197–215, Ithaca, NY, 2010. CLC Publications.

[51] D. Lassiter. Vagueness as probabilistic linguistic knowledge. In R. Nouwen, R. van Rooij, U. Sauerland, and H.-C. Schmitz, editors, *Vagueness in Communication*, pages 127–150. Springer, 2011.

[52] D. Lassiter. Nouwen's puzzle and a scalar semantics for obligations, needs, and desires. In Neil Ashton, Anca Chereches, , and David Lutz, editors, *Semantics and Linguistic Theory 21*. CLC Publications, 2011.

[53] D. Lassiter. *Measurement and Modality: The Scalar Basis of Modal Semantics*. PhD thesis, New York University, 2011.

[54] D. Lassiter and N.D. Goodman. How many kinds of reasoning? Inference, probability, and natural language semantics, 2012. In *Proceedings of the* 34^{th} *Annual Conference of the Cognitive Science Society*.

[55] S.J. Leslie. Generics: Cognition and acquisition. *Philosophical Review*, 117(1): 1–47, 2008.

[56] D. Levinson. Probabilistic model-theoretic semantics for *want*. *Semantics and Linguistic Theory 13*, pages 222–239, 2003.

[57] D. Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1): 339–359, 1979. ISSN 0022-3611. doi: 10.1007/BF00258436.

[58] D. Lewis. Index, context and content. In *Philosophical Papers, volume 1: Papers in Philosophical Logic*. Cambridge University Press, 1997.

[59] A. McCallum, K. Schultz, and S Singh. Factorie: Probabilistic programming via imperatively defined factor graphs. In *Neural Information Processing Systems Conference (NIPS)*, 2009.

[60] Brian Milch, Bhaskara Marthi, Stuart Russell, David Sontag, Daniel L. Ong, and Andrey Kolobov. BLOG: Probabilistic models with unknown objects. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1352–1359, 2005.

[61] R. Montague. The proper treatment of quantification in ordinary english. In K.J.J. HIntikka, J.M.E. Moravcsik, and P. Suppes, editors, *Approaches to Natural Language*, volume 49, pages 221–242. Dordrecht, 1973.

[62] R.E. Nisbett, D.H. Krantz, C. Jepson, and Z. Kunda. The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90(4):339–363, 1983.

[63] D.N. Osherson and E.E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58, 1981.

[64] F.J. Pelletier and N. Asher. Generics and defaults. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier, 1997.

[65] A Pfeffer. Figaro: An object-oriented probabilistic programming language. *Charles River Analytics Technical Report*, 2009.

[66] Avi Pfeffer. IBAL: A probabilistic rational programming language. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 733–740. Morgan Kaufmann Publ., 2001.

[67] D Poole. The independent choice logic and beyond. *Probabilistic inductive logic programming*, pages 222–243, 2008.

[68] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62:107–136, 2006.

[69] Craige Roberts. Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, pages 91–136, 1996.

[70] Craige Roberts. Context in dynamic interpretation. In L. Horn and G. Ward, editors, *Handbook of Contemporary Pragmatic Theory*, pages 197–220. 2004.

[71] E. Rosch. Principles of categorization. *Concepts: core readings*, 1978.

[72] Carmen Rotstein and Yoad Winter. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, 12(3):259–288, 2004.

[73] T. Sato and Y. Kameya. PRISM: A symbolic-statistical modeling language. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 1997.

[74] L. A. Schmidt, N. D. Goodman, D. Barner, and J. B. Tenenbaum. How tall is *Tall*? Compositionality, statistics, and gradable adjectives. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2009.

[75] S.A. Sloman. Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, 35(1):1–33, 1998.

[76] S. Solt. Notes on the comparison class. In U. Sauerland R. Nouwen, R. van Rooij and H.-C. Schmitz, editors, *Vagueness in Communication*, pages 127–150. Springer, 2011.

[77] M. Steedman. *The syntactic process*. MIT press, 2001.

[78] Mark Steedman and Jason Baldridge. Combinatory categorial grammar. *Non-Transformational Syntax Oxford: Blackwell*, pages 181–224, 2011.

[79] A. Stuhlmüller and N.D. Goodman. A dynamic programming algorithm for inference in recursive probabilistic programs. *arXiv preprint arXiv:1206.3555*, 2012.

[80] A. Szabolcsi. Bound variables in syntax: Are there any? In J. Groenendijk et al., editor, *Proceedings of the 6th Amsterdam Colloquium*, pages 331–353, 1987.

[81] Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R Walter, Ashis Gopal Banerjee, Seth Teller, and Nicholas Roy. Approaching the symbol grounding problem with probabilistic graphical models. *AI Magazine*, 32(4):64, 2011.

[82] J.B. Tenenbaum, C. Kemp, T.L. Griffiths, and N.D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011.

[83] Jan van Kuppevelt. Discourse structure, topicality and questioning. *Journal of linguistics*, 31(01):109–147, 1995.

[84] S. Verheyen, J.A. Hampton, and G. Storms. A probabilistic threshold model: Analyzing semantic categorization data with the rasch model. *Acta psychologica*, 135(2):216–225, 2010.

[85] A. von Stechow. Comparing semantic theories of comparison. *Journal of Semantics*, 3(1):1–77, 1984.

[86] T. Williamson. *Vagueness*. Routledge, 1994.

[87] S. Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007. doi: 10.1093/mind/fzm983.

[88] S. Yalcin. Probability Operators. *Philosophy Compass*, 5(11):916–937, 2010.

[89] S. Yalcin. Bayesian expressivism. In *Proceedings of the Aristotelian Society*, 2012.