# Semantic Learning in a Probabilistic Type Theory with Records

**Staffan Larsson**      **Jean-Philippe Bernardy**      **Robin Cooper**

CLASP, Dept. of Philosophy, Linguistics and Theory of Science

University of Gothenburg, Box 200, SE 40530 Sweden

sl@ling.gu.se      jean-philippe.bernardy@gu.se      cooper@ling.gu.se

## Abstract

We propose a probabilistic account of semantic learning from interaction formulated in terms of probabilistic type theory with records, building on Cooper et al. (2014, 2015); Larsson and Cooper (2021). Starting from a probabilistic type theoretic formulations of naive Bayes classifiers, we illustrate our account of semantic learning with a simple language game (the fruit recognition game).

## 1 Introduction

A probabilistic type theory was presented by Cooper et al. (2014) and Cooper et al. (2015), which extends Cooper's Type Theory with Records (TTR, Cooper, 2012a; Cooper and Ginzburg, 2015; Cooper, in prep). This theory, Probabilistic Type Theory with Records (ProbTTR) assigns probability values, rather than Boolean truth-values, to type judgements.

TTR has been used previously for natural language semantics (see, for example, Cooper (2005) and Cooper (2012a)), and to analyze semantic coordination and learning (for example, Larsson and Cooper (2009); Cooper and Larsson (2009)). It has also been applied to the analysis of interaction in dialogue (for example, Ginzburg (2012) and Breitholtz (2020)), and in modelling robotic states and spatial cognition (for example, Dobnik et al. (2013)). We believe that a probabilistic version of TTR could be useful in all these domains.

Two main considerations motivated recasting TTR in probabilistic terms. First, a probabilistic type theory offers a natural framework for capturing the gradience of semantic judgements. This allows it to serve as the basis for an account of vagueness in interpretation, as shown by Fernández and Larsson (2014). Second, such a theory lends itself to developing a model of semantic learning that can be straightforwardly integrated into more general probabilistic explanations of learning and inference. It is the latter goal that we pursue here.

In this paper we build on the account of probabilistic inference and classification in ProbTTR introduced by Larsson and Cooper (2021). There, a ProbTTR version of a random variable, not present in the work of Cooper et al. (2015), was introduced. It was also shown how probabilistic classification of perceptual evidence can be combined with probabilistic reasoning. By proposing a Bayesian account of semantic learning formulated in terms of probabilistic type theory, we connect probabilistic semantic learning to the modeling of perceptual meaning as classifiers.

In the following, we first provide a brief overview of TTR and Probabilistic TTR. Section 3 reviews the account of semantic classification presented by Larsson and Cooper (2021). Section 4 details a frequentist account of semantic learning in ProbTTR, and Section 5 provides an example of semantic learning. Section 6 concludes and discusses related and future work.

## 2 Background

This section reviews the background needed to follow the rest of the paper: TTR, Probabilistic TTR fundamentals, and Bayes nets and Naive Bayes classifiers.

### 2.1 TTR: A brief introduction

We will be formulating our account in a Type Theory with Records (TTR). We can here only give a brief and partial introduction to TTR; see also Cooper (2005), Cooper (2012b) and Cooper (in prep). To begin with, $s : T$ is a judgment that some $s$ is of type $T$. One *basic type* in TTR is *Ind*, the type of an individual; another basic type is *Real*, the type of real numbers.

Next, we introduce *records* and *record types*. If $a_1 : T_1, a_2 : T_2(a_1), \ldots, a_n :$

$T_n(a_1, a_2, \ldots, a_{n-1})$, where $T(a_1, \ldots, a_n)$ represents a type $T$ which depends on the objects $a_1, \ldots, a_n$, the record to the left in Figure 1 is of the record type to the right.

In Figure 1, $\ell_1, \ldots \ell_n$ are *labels* which can be used elsewhere to refer to the values associated with them. A sample record and record type is shown in Figure 2.

Types constructed with predicates may be *dependent*. This is represented by the fact that arguments to the predicate may be represented by labels used on the left of the ':' elsewhere in the record type. In Figure 2, the type of $c_{man}$ is dependent on ref (as is $c_{run}$).

If $r$ is a record and $\ell$ is a label in $r$, we can use a *path* $r.\ell$ to refer to the value of $\ell$ in $r$. Similarly, if $T$ is a record type and $\ell$ is a label in $T$, $T.\ell$ refers to the type of $\ell$ in $T$. Records (and record types) can be nested, so that the value of a label is itself a record (or record type). As can be seen in Figure 2, types can be constructed from predicates, e.g., "run" or "man". Such types are called *ptypes* and can intuitively be thought of as types of situations. Such types of situations can be construed as propositions, following the "propositions as types" principle.

## 2.2 Probabilistic TTR fundamentals

The core of ProbTTR is the notion of a probabilistic judgement, where a situation $s$ is judged to be of a type $T$ with some probability.

(1) $p(s : T) = r$, where $r \in [0,1]$

Such a judgement expresses a subjective probability in that it encodes an agent's take on the likelihood that a situation is of that type.

A *probabilistic Austinian proposition* is an object (a record) that corresponds to, or encodes, a probabilistic judgement. Probabilistic Austinian propositions are records of the type in (2).

$$(2) \quad \begin{bmatrix} \text{sit} & : & \textit{Sit} \\ \text{sit-type} & : & \textit{Type} \\ \text{prob} & : & [0,1] \end{bmatrix}$$

A probabilistic Austinian proposition $\varphi$ of this type corresponds to the judgement that $\varphi$.sit is of type $\varphi$.sit-type with probability $\varphi$.prob.

(3) $p(\varphi.\text{sit}:\varphi.\text{sit-type})= \varphi.\text{prob}$

We assume that agents track observed situations and their types, modelled as probabilistic Austinian propositions.

We use $p(T_1||T_2)$ to represent the probability that any situation $s$ is of type $T_1$, given that $s$ is of type $T_2$. Note that $p(T_1||T_2)$, is different from $p(T_1|T_2)$, the probability of there being something of type $T_1$ given that there is something of type $T_2$. We can refer to the former as the *bound variable* (or perhaps *universal*) conditional probability[1], and the latter as the *existential* conditional probability.

## 2.3 Bayesian nets and the Naive Bayes classifier

A Bayesian Network is a Directed Acyclic Graph (DAG). The nodes of the DAG are random variables, each of whose values is the probability of one of the set of possible states that the variable denotes. Its directed edges express dependency relations among the variables. When the values of all the variables are specified, the graph describes a complete joint probability distribution for its random variables. Bayesian Networks provide graphical models for probabilistic learning and inference (Pearl (1990); Halpern (2003)).

A standard Naive Bayes model is a special case of a Bayesian network. More precisely, it is a Bayesian network with a single class variable $C$ that influences a set of evidence variables $E_1, \ldots, E_n$ (the evidence), which do not depend on each other. Figure 3 illustrates the relation between evidence types and class types in a Naive Bayes classifier.

A Naive Bayes classifier computes the marginal probability of a class, given the evidence:

(4)

$$p(c) = \sum_{e_1, \ldots, e_n} p(c \mid e_1, \ldots, e_n)p(e_1) \ldots p(e_n)$$

where $c$ is the value of $C$, $e_i$ is the value of $E_i$ $(1 \leq i \leq n)$ and

(5) $p(c \mid e_1, \ldots, e_n) =$

$$\frac{p(c)p(e_1 \mid c) \ldots p(e_n \mid c)}{\sum_{C=c'} p(c')p(e_1 \mid c') \ldots p(e_n \mid c')}$$

## 2.4 Random variables in TTR

Larsson and Cooper (2021) introduce a type theoretic counterpart of a random variable in Bayesian

---

[1] In Bayesian jargon, such conditional probabilities are often referred to as *likelihoods*.

$$\begin{bmatrix} \ell_1 & = & a_1 \\ \ell_2 & = & a_2 \\ \dots & & \\ \ell_n & = & a_n \\ \dots & & \end{bmatrix} : \begin{bmatrix} \ell_1 & : & T_1 \\ \ell_2 & : & T_2(l_1) \\ \dots & & \\ \ell_n & : & T_n(\ell_1, l_2, \dots, l_{n-1}) \end{bmatrix}$$

Figure 1: Schema of record and record type

$$\begin{bmatrix} \text{ref} & = & \text{obj}_{123} \\ \text{c}_{\text{man}} & = & \text{prf}_{\text{man}} \\ \text{c}_{\text{run}} & = & \text{prf}_{\text{run}} \end{bmatrix} : \begin{bmatrix} \text{ref} & : & \text{Ind} \\ \text{c}_{\text{man}} & : & \text{man(ref)} \\ \text{c}_{\text{run}} & : & \text{run(ref)} \end{bmatrix}$$
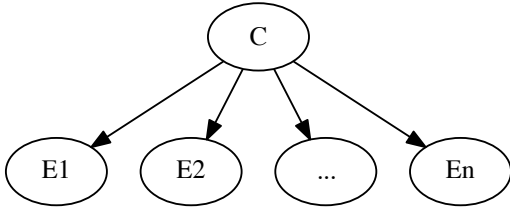
Figure 2: Sample record and record type



Figure 3: Evidence and Class in a Naive Bayes classifier

inference. To represent a single (discrete) random variable with a range of possible (mutually exclusive) values, ProbTTR uses a *variable type* $V$ whose range is a set of *value types* $\mathfrak{R}(V) = \{A_1, \dots, A_n\}$ such that the following conditions hold.

(6) a. All value types for a variable type $V$ are subtypes of $V$, formally $A_j \sqsubseteq V$ for $1 \le j \le n$

   b. All value types for a given variable type $V$ are disjoint, formally $A_j \perp A_i$ for all $i, j$ such that $1 \le i \ne j \le n$

   c. The probability of a situation $s$ being of a variable type $V$ is either 0 or 1, which is also the sum of the probabilities of $s$ being of any of the variable value types , formally for any $s$, $p(s : V) \in \{0, 1\}$ and $p(s : V) = \sum_{T \in \mathfrak{R}(V)} p(s : T)$

(6)(c) encodes a conceptual difference between the probability that something has a property (such as colour, $p(s:\text{Colour})$), and the probability that it has a certain value of a variable (e.g. $p(s:\text{Green})$). If the probability distribution over different values (colours) sums to 1, then the probability that the object in question has a colour is 1. The probability that an object has colour is either 0 or 1. We assume that certain ontological/conceptual type judgements of the form "physical objects have colour" are categorical, and so have Boolean values.

### 2.5 Representing probability distributions

For a situation $s$, a probability distribution over the $m$ value types $A_j \in \mathfrak{R}(\mathbb{A}), 1 \le j \le m$ belonging to a variable type $\mathbb{A}$ can be written (as above) as a set of probabilistic Austinian propositions, e.g.

(7) $\left\{ \begin{bmatrix} \text{sit} = s \\ \text{sit-type} = A_j \\ \text{prob} = p(s : A_j) \end{bmatrix} \mid A_j \in \mathfrak{R}(\mathbb{A}) \right\}$

However, we will also have use for a vector representation of probability distributions, which is also more compact. If we assume $\mathfrak{R}(\mathbb{A})$ is an ordered set $\{A_1, \dots A_m\}$, we can define probability distribution $d_{\mathbb{A}}(s)$:

(8) $d_{\mathbb{A}}(s) = \langle p_1, \dots, p_m \rangle$ where $p_j = p(s : A_j)$ for $A_j \in \mathfrak{R}(\mathbb{A}), 1 \le i \le m$

### 2.6 A ProbTTR Naive Bayes classifier

Corresponding to the evidence, class variables, and their value types, we associate with a ProbTTR Naive Bayes classifier $\kappa$:

(9) a. a collection of $n$ evidence variable types $\mathbb{E}_1^\kappa, \dots, \mathbb{E}_n^\kappa$

   b. $n$ associated sets of evidence value types $\mathfrak{R}(\mathbb{E}_1^\kappa), \dots, \mathfrak{R}(\mathbb{E}_n^\kappa)$

c. a class variable type $\mathbb{C}^\kappa$, e.g. *Fruit*, and

d. an associated set of class value types $\mathfrak{R}(\mathbb{C}^\kappa)$

To classify a situation $s$ using a classifier $\kappa$, the evidence is acquired by observing and classifying $s$ with respect to the evidence types.

Larsson and Cooper (2021) define a ProbTTR Bayes classifier $\kappa$ as a function from a situation $s$ (of the meet type of the evidence variable types $\mathbb{E}_1^\kappa, \ldots, \mathbb{E}_n^\kappa$) to a set of probabilistic Austinian propositions that define a probability distribution over the values of the class variable type $\mathbb{C}^\kappa$, given probability distributions over the values of each evidence variable type $\mathbb{E}_1^\kappa, \ldots, \mathbb{E}_n^\kappa$. Formally, a ProbTTR NaiveBayes classifier is a function

(10) $\kappa : \mathbb{E}_1^\kappa \wedge \ldots \wedge \mathbb{E}_n^\kappa \to$

$$\text{Set}(\begin{bmatrix} \text{sit} & : & Sit \\ \text{sit-type} & : & Type \\ \text{prob} & : & [0,1] \end{bmatrix})$$

such that if[2] $s : \mathbb{E}_1^\kappa \wedge \ldots \wedge \mathbb{E}_n^\kappa$, then

(11) $\kappa(s) = \{\begin{bmatrix} \text{sit} = s \\ \text{sit-type} = C \\ \text{prob} = p^\kappa(s : C) \end{bmatrix} \mid C \in \mathfrak{R}(\mathbb{C}^\kappa)\}$

### 2.7 The fruit recognition game

Larsson and Cooper (2021) illustrate semantic classification using a Naive Bayes classifier in ProbTTR using the *fruit recognition game*. Later in this paper, we will build on this example to illustrate mentor-driven semantic learning.

In this game a teacher shows fruits to a learning agent. The agent makes a guess, the teacher provides the correct answer, and the agent learns from these observations.

We will use shorthands *Apple* and *Pear* for the types corresponding to an object being an apple or a pear, respectively[3]. Furthermore, we will assume that the objects in the fruit recognition game have one of two shapes (a-shape or p-shape, corresponding to types *Ashape* and *Pshape*) and one of two colours (green or red, corresponding to types *Green* and *Red*).

---

[2]Recall that for any $s$, $p(s : V) \in \{0, 1\}$ for any variable type $V$. Therefore, any type judgement regarding a variable type, such as that involved in the classifier function, can be regarded as categorical.

[3]For details, see Larsson and Cooper (2021).

The class variable type is *Fruit*, with value types $\mathfrak{R}(Fruit) = \{Apple, Pear\}$. The evidence variable types are (i) *Col*(our), with value types $\mathfrak{R}(Col) = \{Green, Red\}$, and (ii) Shape, with value types $\mathfrak{R}(Shape) = \{Ashape, Pshape\}$. Figure 4 shows the evidence and class types of the fruit recognition game in a simple Bayesian Network.
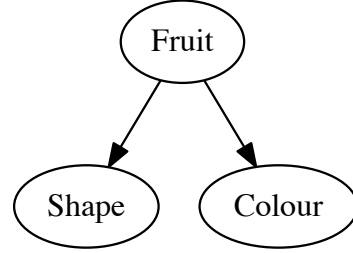


Figure 4: Bayesian Network for the fruit recognition game

For a situation $s$ the classifier FruitC$(s)$ returns a probability distribution over the value types in $\mathfrak{R}(Fruit)$.

(12) FruitC$(s) =$
$$\{\begin{bmatrix} \text{sit} = s \\ \text{sit-type} = F \\ \text{prob} = p^{\text{FruitC}}(s : F) \end{bmatrix} \mid F \in \mathfrak{R}(Fruit)\}$$

## 3 Semantic classification using conditional probabilities

In this section, we follow Larsson and Cooper (2021) in showing how semantic classification (i.e., estimating a probability distribution over class value types) works under the assumption that we can compute conditional probabilities $p(C_j || E_1 \ldots E_n)$ of a class value types $C_j$ given evidence value types $E_1 \ldots E_n$.

In general, for $C_j \in \mathfrak{R}(\mathbb{C}^\kappa)$, we have

(13) $p^\kappa(s : C_j) =$

$$\sum_{\substack{E_1 \in \mathfrak{R}(\mathbb{E}_1^\kappa) \\ E_n \in \ddot{\mathfrak{R}}(\mathbb{E}_n^\kappa)}} p^\kappa(C_j || E_1 \ldots E_n) p(s : E_1) \ldots p(s : E_n)$$

Correspondingly, in the fruit recognition game, for each $F \in \mathfrak{R}(Fruit)$ we have

(14) $p^{FruitC}(s:F) =$

$$\sum_{\substack{L \in \Re(Col) \\ S \in \Re(Shape)}} p(F||L \wedge S)p(s:L)p(s:S)$$

Therefore, to determine the probability that a situation is of the apple type, we sum over the various evidence type values for apple.

(15) $p^{\text{FruitC}}(s:Apple) =$
$p(Apple||Green \wedge Ashape)p(s:Green)p(s:Ashape) +$
$p(Apple||Green \wedge Pshape)p(s:Green)p(s:Pshape) +$
$p(Apple||Red \wedge Ashape)p(s:Red)p(s:Ashape) +$
$p(Apple||Red \wedge Pshape)p(s:Red)p(s:Pshape)$

Conditional probabilities for the fruit classifier are derived from previous judgements of the form $p(F||C \wedge S)$. The example values in the matrix in (16) illustrate a joint probability distribution for the Bayesian Network in Figure 4.

(16)

| Apple/Pear | Ashape | Pshape |
|---|---|---|
| Green | 0.93/0.07 | 0.63/0.37 |
| Red | 0.56/0.44 | 0.13/0.87 |

For each square with *Apple/Pear* type values, the conditional probabilities of the fruit being an apple and of its being a pear sum to 1.

The non-conditional probabilities in (15) are derived from the agents' take on the particular situation being classified; let us call it $s_5$.

(17)

| T = | Ashape | Pshape | Green | Red |
|---|---|---|---|---|
| p($s_5$:T) | 0.90 | 0.10 | 0.80 | 0.20 |

This means we have e.g.

(18) $d_{Shape}(s_5) = \langle 0.90, 0.10 \rangle$

Larsson and Cooper (2021) suggest regarding these probabilities as resulting from probabilistic classification of real-valued (non-symbolic) visual input, where a classifier assigns to each image a probability that the image shows a situation of the respective type. Such a classifier can be implemented in a number of different ways, e.g. as a neural network, as long as it outputs a probability distribution.

With these numbers in place, we can compute the probability that the fruit shown in $s_5$ is an apple:

(19) $p^{\text{FruitC}}(s_5: Apple) =$
$0.93 * 0.80 * 0.90 + 0.63 * 0.80 * 0.10 +$
$0.56 * 0.20 * 0.90 + 0.13 * 0.20 * 0.10 =$
$0.67 + 0.05 + 0.10 + 0.00 =$
$0.82$

## 4  Frequentist semantic learning

For the model of semantic classification that uses conditional probabilities, a central question is of course how to estimate conditional probabilities, of the form $p(C||E_1 \wedge \ldots \wedge E_n)$ (where $C \in \Re(\mathbb{C})$, $E_i \in \Re(\mathbb{E}_i), 1 \leq i \leq n$). Using Bayes rule and marginalising over the class value types, we get for a Naive Bayes classifier:

(20) $\hat{p}^{\kappa}(C||E_1 \wedge \ldots \wedge E_n) =$

$$\frac{p(C)p(E_1||C)\ldots p(E_n||C)}{\sum_{C' \in \Re(\mathbb{C}^{\kappa})} p(C')p(E_1||C')\ldots p(E_n||C')}$$

For all combinations of evidence value types $E_1, \ldots, E_n$ and class value types $C$, we need (a) the conditional probability of the evidence value types given the class value type, $p(E_i||C)$, and (b) the prior of the class value type, $p(C')$.

### 4.1  Computing conditional probabilities

Following a frequentist[4] methodology, conditional probabilities can be estimated by counting previous instances of $C$ and $E_i$:

$$p(E_i|C) = \frac{|E_i \& C|}{|C|}$$

This relies on previous judgements being categorical rather than probabilistic. However, it appears reasonable to assume that agents sometimes make non-categorical judgements, assigning a probability other than 0 or 1 to a situation being of a certain type, and we want to explore the idea of using

---

[4]Regarding the tension between Bayesian and frequentist modelling, one might argue that no practically useful model is purely Bayesian, since the moment that you introduce data, you will extract frequencies from it. While theoretical models may be purely bayesian, in all machine learning models there is an element of frequentism. However, being too naively frequentist yields models which generalise poorly. To take an extreme example, a purely frequentist 5-gram model will assign 0-probability to any 5-gram which does not occur at all in the data, which is clearly wrong. Bayesian reasoning is ultimately a mathematical recipe to construct models which explicitly capture the dependencies between various random variables (hidden or not). In the paper we show two ways to improve the model, with varying levels of complexity and robustness.

such non-categorical past judgements as a basis for future (probabilistic) judgements.

Cooper et al. (2015) sketch a solution with a frequentist flavour (but also with some differences to regular frequentist learning acccounts), based on the idea that an agent makes judgements based on a finite string of probabilistic Austinian propositions, the *judgement history* $\mathfrak{J}$. When an agent $A$ encounters a new situation $s$ and wants to know if it is of type $T$ or not, $A$ uses probabilistic reasoning to determine $p(s : T)$ on the basis of $A$'s previous judgements $\mathfrak{J}$.

So the history of judgements $\mathfrak{J}$ does not contain definite judgements, but rather probabilistic ones. How are these probabilities to be understood? We assume that each such probability corresponds to the judging agent's estimate of the probability that a member of the linguistic community would judge $s$ to be of type $T$. That is, we assume that agents make the (semantic) judgements that they estimate that other agents would also make (on average). This can be intuitively justified by the assumption that agents they take language (including meanings) to be public (shared in a community). Hence, each probabilistic judgement in the history can be considered to correspond to a large number $N$ of independent categorical judgements.

How do we motivate this? After all, language is categorical in nature at least insofar as a speaker makes or does not make an utterance $U$ to describe some situation $s$, thus categorising $s$ as (categorically) correctly described by $U$. However, the categorical nature of language does not imply that agents cannot entertain non-categorical judgements, only that once they speak their judgements, they become categorical[5]. When it comes to computing the probabilities needed for probabilistic classifiers, this means that $round(p(s : C)p(s : E_i)N)$ of them are considered to be of type $(C \wedge E_i)$. (The motivation for rounding to integers using the $round$ function is that if we talk about discrete events, there must be an integer number of them.) On this basis, we can compute likelihoods and probabilities as a ratio of the frequencies of occurrences, summed over all judgements in the history:

(21)

$$p(E_i || C) =$$
$$\lim_{N \to \infty} \frac{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} round(p(s : C)p(s : E_i)N)}{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} round(p(s : C)N)} =$$
$$\frac{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} p(s : C)p(s : E_i)}{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} p(s : C)}$$

Formula (21) tells us that we can consider probabilities in the history of judgments as fractions of events; and this is justified by interpreting them as fractions of language-community speakers making the corresponding categorical judgement. In this sense, we are providing a frequentist interpretation of epistemic probability[6]

One might ask regarding (21), is it possible to multiply the probabilities associated with variables that may be dependent, without taking into account their conditional probabilities? Yes, in this case, it is. We are multiplying probabilistic judgements that have already been made rather than hypothetical judgements[7].

For the purposes of this paper, we will assume that the probabilities needed are indeed encoded directly in $\mathfrak{J}$, but of course in general this might

---

[6]Assuming we have $N$ situations in total, some integer number will be classified as each category. We can then sum these integer numbers. We can get a probability by taking the limit of the ratio with N tending to infinity.

[7]This can perhaps be better understood by analogy to counting in standard probability theory. Suppose that we have a corpus of English sentences where all nouns are annotated for part of speech, and for whether the noun has a subject/object role (or neither). We estimate the conditional probability that a noun is a subject by counting the number of nouns that are also subjects, and divide this sum by the total number of nouns.

$$p(\text{Subject}|\text{Noun}) = \frac{|\text{Subject\&Noun}|}{|\text{Noun}|}$$

Categorical judgements can be regarded as probabilistic judgments with probability 1.0, so that judging a word $w$ to be noun is to judge the probability of $w$ being a nount to be 1.0. Assuming a word $w$ has ben judged as being a subject and a noun, we can describe this probabilistically as p($w$ is Subject)=1.0, p($w$ is Noun)=1.0, and we conclude that p($w$ is Subject\&Noun)=p($w$ is Subject)p($w$ is Noun)=1.0*1.0=1.0 without involving p(Subject|Noun), the conditional probability of something being a subject given that it is a noun (which is in fact what we are trying to compute). In doing so, we are *not* taking the probability that a word is a subject to be independent of its being a noun. In fact, we are assuming the opposite. We are trying to compute the conditional probability of a word being a subject given that it is a noun. But when counting a word $w$ as being both a subject and a noun, we do not invoke this conditional probability as part of the enumeration. To do so it would be both circular and unnecessary, as we have already judged $w$ to be a subject, and to be a noun.

---

[5]We are here ignoring for the moment some complications, including that hearers may assign probabilities to speakers having made an utterance $U$ based on perceptual, semantic and pragmatic confidences. We hope to return to these points in future work.

not be the case. Cooper et al. (2015) explains how probabilities of complex types (such as meet types, join types, function types and record types) can be computed from simpler types.

## 4.2 Computing priors

In addition to conditional probabilities, (20) requires the prior probabilities of the class value types $C \in \mathfrak{R}(\mathbb{C})$. We use $\mathfrak{p}_\mathfrak{J}(T)$ to denote the prior probability that an arbitrary situation is of type $T$ given $\mathfrak{J}$. However, it is important to note that the prior probability for a value type $A$ is not the same as the probability $p(A)$ that there is something of type $A$. Rather, it is the probability that some arbitrary situation $s$ (of which we have no other relevant information) is of type $A$. To see this, imagine that $p(s_1 : A) = 0.8$ and $p(s_2 : A) = 0.2$, and that there are no judgements concerning other situations in $\mathfrak{J}$. In this case, $p(A)$, the probability that there is something of type $A$[8], is $0.8 + 0.2 - (0.8 * 0.2) = 0.84$. However, the prior probability that an arbitrary situation is of type $A$, $\mathfrak{p}_\mathfrak{J}(A)$, is $(0.8 + 0.2)/2 = 0.5$.

Following this, we define the prior probability of an arbitrary situation being of a type $T$, $\mathfrak{p}_\mathfrak{J}(T)$, thus:

(22)

$$\mathfrak{p}_\mathfrak{J}(T) = \frac{\sum_{j \in \mathfrak{J}_T} j.\text{prob}}{\text{P}(\mathfrak{J})} \text{ if P}(\mathfrak{J}) > 0, \text{ otherwise } 0$$

where $\mathfrak{J}_T$ is the set of all judgements concerning $T$:

(23)
$$\mathfrak{J}_T = \{j \mid j \in \mathfrak{J}, j.\text{sit-type} = T\}$$

and $\text{P}(\mathfrak{J})$ is the cardinality of situations in $\mathfrak{J}$, i.e. the total number of situations in $\mathfrak{J}$[9]:

(24) $\text{P}(\mathfrak{J}) = |\{s | \exists j \in \mathfrak{J}, j.\text{sit} = s\}|$

Accordingly, we replace (20) with (25), where $p(C)$ is replaced with $\mathfrak{p}_\mathfrak{J}(C)$:

(25) $\hat{p}^\kappa(C || E_1 \wedge \ldots \wedge E_n) =$

$$\frac{\mathfrak{p}_\mathfrak{J}(C)p(E_1||C)\ldots p(E_n||C)}{\sum_{C' \in \mathfrak{R}(\mathbb{C}^\kappa)} \mathfrak{p}_\mathfrak{J}(C')p(E_1||C')\ldots p(E_n||C')}$$

---

[8] See Cooper et al. (2015) for details.
[9] This replaces an earlier definition in Cooper et al. (2015).

## 5 Example: frequentist semantic learning in the fruit recognition game

The conditional probabilities in (16) are generated from $\mathfrak{J}$ by a learning component. Let's assume that $\mathfrak{J}$ is as in Figure 5, based on previous rounds of the game.

The recorded judgements concerning the types *Apple* and *Pear* are here assumed to be derived not only from the agent's own perception of the fruits in question, but also (and perhaps primarily) from a tutor's explicit judgements, possibly in combination with an estimation of the likelihood that the teacher is competent at judging apples and pears under whatever conditions (light etc.) held at the time of judgement.

In our example, $p(F||L \wedge S)$ comes from previous experience as encoded in $\mathfrak{J}$. We estimate this probability with Bayes' rule, as in (26).

(26) $p(F||L \wedge S) =$

$$\frac{\mathfrak{p}_\mathfrak{J}(F)p(L||F)p(S||F)}{\sum_{F' \in \mathfrak{R}(Fruit)} \mathfrak{p}_\mathfrak{J}(F')p(L||F')p(S||F')}$$

To compute this we need the following for all $F \in \{Apple, Pear\}$:

(27)  a. for all $L \in \{Green, Red\}, p(L||F)$

   b. for all $S \in \{Ashape, Pshape\}, p(S||F)$

   c. $\mathfrak{p}_\mathfrak{J}(F)$

We use (21) to compute conditional probabilities, so that for example

(28) $p(Green||Apple) =$

$$\frac{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} p(s : Apple)p(s : Green)}{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} p(s : Apple)} =$$

$$\frac{0.9 * 1.0 + 0.7 * 0.5 + 1.0 * 0.9 + 0.0 * 0.1}{1.0 + 0.5 + 0.9 + 0.1} = 0.86$$

We also use (22) to compute priors, so that for example

(29)

$$\mathfrak{p}_\mathfrak{J}(Apple) = \frac{\sum_{j \in \mathfrak{J}, j.\text{sit}=s} p(s : Apple)}{\text{P}(\mathfrak{J})} =$$

$$\frac{1.0 + 0.5 + 0.9 + 0.1}{4} = \frac{2.50}{4} = 0.63$$

| $j$.p | $j$.sit-type$\in \mathfrak{R}(Fruit)$ | | $j$.sit-type$\in \mathfrak{R}(Col)$ | | $j$.sit-type$\in \mathfrak{R}(Shape)$ | |
|---|---|---|---|---|---|---|
| $j$.sit | *Apple* | *Pear* | *Green* | *Red* | *Ashape* | *Pshape* |
| $s_1$ | 1.0 | 0.0 | 0.9 | 0.1 | 0.7 | 0.3 |
| $s_2$ | 0.5 | 0.5 | 0.7 | 0.3 | 0.6 | 0.4 |
| $s_3$ | 0.9 | 0.1 | 1.0 | 0.0 | 1.0 | 0.0 |
| $s_4$ | 0.1 | 0.9 | 0.0 | 1.0 | 0.0 | 1.0 |

Figure 5: Conditional probabilities in the fruit recognition game

Based on this, we compute the conditional probabilities shown in (16) and used in classification in Section 2.7, for example

(30) $p(Apple || Green \wedge Ashape) =$

$$\frac{\mathfrak{p}_\mathfrak{J}(Apple)p(Green||Apple)p(Ashape||Apple)}{\sum_{F \in \{Apple, Pear\}} \mathfrak{p}_\mathfrak{J}(F)p(Green||F)p(Ashape||F)} =$$

$$\frac{0.41}{0.41+0.03} = 0.93$$

Based on the judgement above in (19), our agent may venture the guess that the fruit in question in $s_5$ is an apple, to which the tutor may respond "Very good!". This in turn could trigger extending $\mathfrak{J}$ to include probabilistic judgements concerning the classification of $s_5$ as being of types *Apple*, *Pear*, *Green*, *Red*, *Ashape* and *Pshape*, to be used in future rounds of the game.

## 6 Conclusion

Cooper et al. (2014) and Cooper et al. (2015), and more recently Larsson and Cooper (2021), presented a probabilistic formulation of a rich type theory with records, and used it as the foundation for a compositional semantics in which a probabilistic judgement that a situation is of a certain type plays a central role. The basic types and type judgements at the foundation of the type system correspond to perceptual judgements concerning objects and events in the world, rather than to entities in a model, and set theoretic constructions defined on them. This approach grounds meaning in observational judgements concerning the likelihood of situations holding in the world. We have proposed a Bayesian account of semantic learning formulated in terms of ProbTTR, thereby connecting probabilistic semantic learning to other phenomena studied in TTR and ProbTTR, including the modeling of perceptual meaning as classifiers (Larsson, 2013; Larsson and Cooper, 2021).

Our treatment of learning relies on the idea that an agent keeps a record of their previous judgements concerning the likelihood of a classification and sums the probabilities of these judgements. The agent computes conditional probabilities and priors for current judgements on the basis of this record. We have illustrated this view of learning with the fruit recognition game. This simplified example provides a sketch of how an agent can acquire a set of predicates through mentor vetted (supervised) classifier learning.

With respect to semantic learning, this paper follows in the general footsteps of van Eijck and Lappin (2012), who propose a probabilistic theory of language semantics which includes a sketch of semantic learning. It appears that our model is an instance of the strategies outlined by van Eijck and Lappin. Where they only sketch a strategy, we have shown in detail how learning from examples can be modelled.

As part of the Rational Speech Act Theory, Goodman and Lassiter (2015); Lassiter and Goodman (2017) provide an account of semantic update of an agent's view of the world, which can possibly be regarded as a form of semantic learning. Even though they apply it to a single event, their account can be generalised to several events in a natural way. Indeed, Bernardy et al. (2018, 2019) have implemented such a generalisation. What sets the present work apart, in addition to being formulated in ProbTTR, is that each individual event is not categorical, but itself probabilistic. We have achieved this by incorporating elements of frequentist thinking in an otherwise Bayesian account. Conversely, the approaches previously mentioned manage to remain in a purely Bayesian framework, but they do not generalise to probabilistic events in a straightforward manner.

Future work includes exploring and adapting other learning methods to ProbTTR, including a linear transformation model and related neural network and deep learning models, and continuing to apply ProbTTR to a variety of problems in natural language semantics.

## Acknowledgements

## References

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, and Shalom Lappin. 2018. A compositional Bayesian semantics for natural language. In *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pages 1–10, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin, and Aleksandre Maskharashvili. 2019. Bayesian inference semantics: A modelling system and a test suite. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 263–272, Minneapolis, Minnesota. Association for Computational Linguistics.

Ellen Breitholtz. 2020. *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Brill, Leiden, The Netherlands.

Robin Cooper. 2005. Records and record types in semantic theory. *Journal of Logic and Computation*, 15(2):99–112.

Robin Cooper. 2012a. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

Robin Cooper. 2012b. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

Robin Cooper. in prep. From perception to communication: An analysis of meaning and action using a theory of types with records (TTR). Draft available from https://sites.google.com/site/typetheorywithrecords/drafts.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2014. A probabilistic rich type theory for semantic interpretation. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 72–79. Gothenburg, Association of Computational Linguistics.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology 10*, pages 1–43.

Robin Cooper and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory, Second Edition*, pages 375–407. Wiley-Blackwell, Oxford and Malden.

Robin Cooper and Staffan Larsson. 2009. Compositional and ontological semantics in learning from corrective feedback and explicit definition. In *Proceedings of DiaHolmia: 2009 Workshop on the Semantics and Pragmatics of Dialogue*, pages 59–66. Department of Speech, Music and Hearing, KTH.

Simon Dobnik, Robin Cooper, and Staffan Larsson. 2013. Modelling language, action, and perception in Type Theory with Records. In Denys Duchier and Yannick Parmentier, editors, *Constraint Solving and Language Processing - 7th International Workshop on Constraint Solving and Language Processing, CSLP 2012, Orleans, France, September 13-14, 2012. Revised Selected Papers*, number 8114 in Publications on Logic, Language and Information (FoLLI). Springer, Berlin, Heidelberg.

Raquel Fernández and Staffan Larsson. 2014. Vagueness and learning: A type-theoretic approach. In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (*SEM 2014)*.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

N. Goodman and D. Lassiter. 2015. Probabilistic semantics and pragmatics: Uncertainty in language and thought. In S. Lappin and C. Fox, editors, *The Handbook of Contemporary Semantic Theory, Second Edition*, pages 655–686. Wiley-Blackwell, Malden, Oxford.

J. Halpern. 2003. *Reasoning About Uncertainty*. MIT Press, Cambridge MA.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*.

Staffan Larsson and Robin Cooper. 2009. Towards a formal view of corrective feedback. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 1–9. EACL.

Staffan Larsson and Robin Cooper. 2021. Bayesian classification and inference in a probabilistic type theory with records. In *Proceedings of NALOMA 2021*.

Daniel Lassiter and Noah D. Goodman. 2017. Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 194(10):3801–3836.

J. Pearl. 1990. Bayesian decision methods. In G. Shafer and J. Pearl, editors, *Readings in Uncertain Reasoning*, pages 345–352. Morgan Kaufmann.

Jan van Eijck and Shalom Lappin. 2012. Probabilistic semantics for natural language. In Z. Christoff, P. Galeazzi, N. Gierasimszuk, A. Marcoci, and S. Smets, editors, *Logic and Interactive Rationality (LIRA), Volume 2*, pages 17–35. ILLC, University of Amsterdam.