Probabilistic inference in dynamic semantics

Alex Djalali and Stefan Kaufmann*[†]
**ILLC / Stanford University*
[†]*Northwestern University*

1 Introduction

One of the major strengths of Dynamic Semantics is its ability to model both (i) the constraints governing the introduction, persistence, and accessibility of discourse referents; and (ii) the role played by contextual factors such as beliefs and the common ground in steering the interpretation of utterances. Both (i) and (ii) are crucially involved in the resolution of pronominal anaphoric reference: The interpreter's choice relies on a range of information sources, both linguistic (e.g., embedding structures; grammatical gender) and extralinguistic (e.g., knowledge about what is likely to be true, or to be the speaker's intention). Many of the inferences involved, especially of the latter kind, are uncertain and heavily dependent on the content of the utterance in question. To model these inferences and their interaction with more genuinely linguistic constraints, one would want an integrated framework in which both uncertain beliefs about the world and the linguistic constraints traditionally modeled by Dynamic Semantics can be represented together.

In this paper we introduce such a framework, called **Probabilistic Dynamic Predicate Logic (PDPL).** Its ultimate purpose is to seamlessly integrate what is conventionally split into "world knowledge" on the one hand, and "linguistic knowledge," on the other. The guiding philosophical assumption is that this division is a theoretical artifact which does not correspond to any clear cognitive division that could be drawn on independent grounds. Linguistic knowledge *is* (part of) world knowledge, and both go hand in hand in the reasoning involved in communication.

2 Bridging

With the above vision in mind, in this paper we focus on a relatively small but illuminating application of PDPL: the resolution of pronominal anaphora in *bridging inferences.* Bridging is a way to augment the information that is literally conveyed by what was said, typically by fleshing out a background "scenario" relative to which the utterance is less ambiguous and more coherent that it would be on its own. To the extent that the hearer can be counted on to draw bridging inferences in the utterance situation, such inferences may become part of what is communicated. Often bridging involves establishing coreferential links for anaphoric (in)definites, as shown in (1) (Clark & Haviland 1977; Hawkins 1978; Asher & Lascarides 2003; Levinson 2000; Matsui 2001).

(1) *A bridge collapsed. The strut was rotten.*
  a. 'A bridge collapsed because a strut of the bridge was rotten.'
  b. 'A bridge collapsed because a strut *[of some unmentioned antecedent]* was rotten.'

Here, (1a) is likely to be inferred to connect the sentences in (1). Importantly, though, the anaphoric link identifying the strut as part of the bridge is not required for the truth of (1), and the weaker inference in (1b) may be sufficient, even preferred, in certain contexts.

We argue that standard accounts of bridging fail to provide a satisfactory formal account of the way in which cross-sentential anaphora are resolved in bridging, particularly the role of world knowledge in adjudicating between multiple alternative resolutions of anaphoric coreference. PDPL fills this gap by modeling the information states of interlocutors in a given discourse and keeping track of the probabilities they assign to alternative resolutions.

For a further illustration, consider (2) from Clark & Haviland (1977), and imagine some speaker $\mathcal{A}$ utters (2a) to a hearer $\mathcal{B}$:

(2) a. *John unpacked the picnic. The beer was warm.*
b. *The beer was part of the picnic.*
c. $\exists! x[\text{unpack}(j, x)] \wedge \exists! y[\text{beer}(y) \wedge \text{warm}(y)]$
d. $\exists! x[\text{unpack}(j, x)] \wedge \exists! y[\text{beer-of}(y, z) \wedge \text{warm}(y)] \wedge x = z$

We would not expect $\mathcal{A}$ to explicitly state (2b) in addition to (2a). Speakers can often assume that during a given discourse, hearers will extrapolate from underspecified utterances. In (2), the inference consists in finding an appropriate anaphoric antecedent for *the beer.* Standard pragmatic accounts of bridging inferences, such as Hawkins (1978) and Levinson (2000), assume that in order for successful bridging to occur, $\mathcal{A}$ expects $\mathcal{B}$ to use her general knowledge of language and the world to draw stereotypical connections between the objects and events referred to in the utterance, in such a way that that overall discourse coherence is maintained. Specifically in (2), $\mathcal{A}$ expects $\mathcal{B}$ to share his belief that picnics usually involve beer drinking, whence $\mathcal{B}$ can resolve the anaphoric definite by reinterpreting *beer* as a relational noun of sorts. The resulting augmented logical form of (2a) is (2d): Here, the beer is said to be "beer of" z, and the coreference is established by the assertion that $x = z$.[1]

For the purposes of this paper, we presuppose that the mechanism behind these inferences is in place. What we are concerned with is the following: While the coreference resolution in (2) was straightforward, the standard pragmatic story breaks down when faced with what we refer to as **ambiguous bridging inferences:**

(3) *A car hit a truck. The windshield shattered.*
a. 'A car hit a truck. The windshield of the car shattered.'
b. 'A car hit a truck. The windshield of the truck shattered.'
c. 'A car hit a truck. The windshield *[of some unmentioned antecedent]* shattered.'

Given current theories of bridging, (3a) and (3b) are equally good inferences. Since both cars and trucks have windshields, (3a) is no more more salient an inference than (3b), or vice versa. Nor is there a "stereotypical relationship" between the objects and events mentioned in (3) that could be relied upon in building an unambiguous bridge. However, irrespective of context, we have intuitions that the relative likelihood of (3a) is greater than that of (3c). It is asymmetries like this that we use PDPL to explain.

3 Formal setup

We begin with a non-classical logical language $\mathcal{L}$ whose syntax we take to be that of standard predicate logic. We define a **world model** to be a triple $\langle W, D, I \rangle$, where W is

[1]Needless to say, the variables would be free in standard predicate logic, but we are assuming a dynamic system here.

a non-empty set of possible worlds; D a domain of individuals common to all worlds in W; and I a function from worlds to interpretation functions assigning values to the non-logical constants of $\mathcal{L}$ in the usual way. We assume for simplicity that W and D are finite. Based such a model, the information state of an agent can be represented by two ingredients: a prior probability measure P on subsets of W; and an accessibility relation R in W^2 that is serial, transitive and Euclidean, and such that for all $w, v \in W$, if wRv, then $P(\{v\}) > 0$.[2] The probability $P(\{v\})$ of the singleton proposition $\{v\}$ is written as $P(v)$. The agent's subjective probability at $w \in W$ of a world v is $P_w(v|R) := P(v)/\sum_{u:wRu} P(u)$ if wRv, 0 otherwise. The probability of a sentence $\varphi \in \mathcal{L}$ at world w is the expectation of its truth value: $P_w(\varphi|R) := \sum_{v \in W} P_w(v|R) \times I_v(\varphi)$.

In addition to world knowledge, we are interested in modeling discourse information. Let the set of **possibilities** be $\mathcal{I} = \{\langle w, g\rangle \mid w \in W, g \in D^X, X \subseteq \mathrm{Var}\}$, i.e., pairs of worlds and partial variable assignments. For any possibility $i = \langle w, g\rangle$, we write "$i(\alpha)$" for the denotation of α at i, i.e., $I_w(\alpha)$ if α is a constant, and $g(\alpha)$ if α is a variable. An information state again involves two ingredients, an accessibility relation and a probability distribution. They are derived from R and P in the "initial" state without any discourse information (see below), but may develop in different ways over the course of a discourse.

The first ingredient is an accessibility relation c in $\mathcal{I}^2$ such that for all worlds w, v, $\langle w, \emptyset\rangle c \langle v, \emptyset\rangle$ iff wRv. This relation is serial and transitive if R is, but may become non-Euclidean during the discourse. It is updated in response to incoming linguistic information. Definitions 1 and 2 fix some auxiliary notions; Definition 3 gives the recursive definition.

Definition 1 (Referent activation) For all variables u, a relation $[u] \subseteq \mathcal{I}^2$ is defined as follows: $\langle w, g\rangle[u]\langle w', g'\rangle$ iff (i) $w = w'$; (ii) $u \notin \mathit{dom}(g)$; (iii) $\mathit{dom}(g') = \mathit{dom}(g) \cup \{u\}$; (iv) $g'(v) = g(v)$ for all $v \neq u$.

Definition 2 (Descendants) A pair $\langle i', j'\rangle \in \mathcal{I}^2$ is a *descendant* of $\langle i, j\rangle \in \mathcal{I}^2$ iff for some sequence $x_1, \ldots, x_n$ with $0 \leq n$, $i[x_1]\ldots[x_n]i'$ and $j[x_1]\ldots[x_n]j'$.

Definition 3 (Update)

$$\begin{aligned}
c[Pt_1, \ldots, t_n] &:= \{\langle i, j\rangle \in c : \langle j(t_1), \ldots, j(t_n)\rangle \in j(P)\} \\
c[\neg\varphi] &:= \{\langle i, j\rangle \in c : \langle i, j\rangle \text{ has no descendants in } c[\varphi]\} \\
c[\varphi \wedge \psi] &:= c[\varphi][\psi] \\
c[\exists x] &:= \{\langle i', j'\rangle : \exists\langle i, j\rangle \in c\, [i[x]i' \text{ and } j[x]j']\}
\end{aligned}$$

The semantics of PDPL follows closely that of Groenendijk et al. (1996), except for the fact we deal with sets of pairs of possibilities rather than sets of possibilities. This allows for an easy extension to multi-agent settings and second-order beliefs.

The second ingredient is a probability distribution Pr such that for all w, v,

$$\Pr_{\langle w, \emptyset\rangle}(\langle v, \emptyset\rangle|c) := P_w(v|R).$$

According to Definition 3, all updates except those with $\exists$ are eliminative. For eliminative updates, the probability is shifted by conditioning. In an update with $\exists$, the probabilities of individual possibilities are distributed uniformly over their descendants.

[2] The latter condition is not strictly required, but convenient because it ensures that all relevant conditional probabilities are defined (as long as we restrict ourselves to consistent updates, which we do in this paper).

Definition 4 (Probability update) Probabilities are updated as follows:

1. For eliminative update $[\varphi]$, $\mathrm{Pr}_i(j|c[\varphi]) := \begin{cases} \mathrm{Pr}_i(j|c)/\sum_{k:i(c[\varphi])k}\mathrm{Pr}_i(k|c) \text{ if } i(c[\varphi])j, \\ 0 \text{ otherwise} \end{cases}$

2. If icj, $i[x]i'$ and $j[x]j'$, then $\mathrm{Pr}_{i'}(j'|c[\exists x]) := \mathrm{Pr}_i(j|c)/|D|$

4 Illustration

To illustrate the workings of the theory, we use a highly simplified example which does, however, show how new information can change an interpreter's beliefs about discourse referents. Consider the interpretation of the mini-sequences in (3) in a model

$$\mathbb{M} = \langle W, D, I, P, R\rangle,$$

where

$$W = \{w_{11}, \ldots, w_{14}, w_{21}, \ldots, w_{24}, w_{31}, \ldots, w_{34}\}$$

is a set of twelve worlds, and

$$D = \{a, b, c, d, a', b', c', d'\}$$

a common domain of eight individuals. For clarity, we begin by lumping the worlds together into three classes w_1, w_2, w_3, which we treat as "worlds" until finer distinctions become necessary. The probabilities of these three classes and some of the relevant extensions are given in Table 1. This state of affairs is meant to encode the following beliefs

w	$P(w)$	$I_w(car)$	$I_w(truck)$	$I_w(hit)$
w_1	.3	$\{a, b\}$	$\{c, d\}$	$\{\langle a, d\rangle\}$
w_2	.5	$\{a, b\}$	$\{c, d\}$	$\{\langle b, d\rangle\}$
w_3	.2	$\{a, b\}$	$\{c, d\}$	$\{\langle c, d\rangle\}$

Table 1: Initial state

on the part of the hearer: She knows that there are two cars and two trucks, and she knows which individuals fall under each class. She also knows that there was exactly one accident in which a vehicle hit truck d, but she does not know which of the other three was involved. Car a is owned by a cautious driver who is unlikely to get into that kind of accident. The owner of car b, on the other hand, is a rambunctious young man with a penchant for reckless driving. Truck drivers are generally less likely to hit trucks. Thus given that something hit truck d, it is most likely to have been car b and less likely to have been car a or truck c. The probabilities are given in the table.

The four other individuals are windshields. Each vehicle x has exactly one windshield x'. They are omitted from the table and the calculations for now in the interest of simplicity, which is unproblematic because there is a one-to-one mapping. Indeed, we will not refer to them directly, assuming instead below that a function *windshield_of* maps vehicles to their windshields. Also omitted from the table, but to return below, are the probabilities that the various windshields shattered if they were involved in the accident. Each of the three accident has four possible outcomes, according as none, one, or both of the windshields shattered. For each world w_x, these are represented as $w_{x1}, \ldots, w_{x4}$. The story goes as follows: Car a, the one owned by the careful driver, is small and vulnerable. If it hit the truck, its windshield shattered with probability .9. Car b is a sturdy off-road vehicle with a reinforced windshield whose probability of having shattered is only .45. Finally, the windshields of both trucks are the least likely to have shattered, with a probability of .1. In each of the possible accidents, the shattering of each of the windshields is probabilistically independent of the shattering of the other. This gives rise to the conditional probabilities of the various possible outcomes, given the accident, for each of the twelve worlds as given in Table 2.[3] Notice that these are *conditional* probabilities — the actual prior probability of w_{12} is $.3 \times .81 = .243$. But as mentioned above, we lump the

w	$P(w)$	I_w(shatter)	w	$P(w)$	I_w(shatter)	w	$P(w)$	I_w(shatter)
w_{11}	.09	$\emptyset$	w_{21}	.495	$\emptyset$	w_{31}	.81	$\emptyset$
w_{12}	.81	$\{a'\}$	w_{22}	.405	$\{b'\}$	w_{32}	.09	$\{c'\}$
w_{13}	.09	$\{a', d'\}$	w_{23}	.045	$\{b', d'\}$	w_{33}	.01	$\{c', d'\}$
w_{14}	.01	$\{d'\}$	w_{24}	.055	$\{d'\}$	w_{34}	.09	$\{d'\}$

Table 2: Probabilities of shattering

three classes of worlds together until information about the shattering is processed. Until then, no information cuts through any of the three blocks w_1, w_2, w_3.

The accessibility relation R is initially the entire cross-product $W \times W$. Throughout the updates it will remain serial, transitive and euclidean. In a sense, therefore, the relation does no interesting work and keeping track of it explicitly would needlessly complicate things. We choose instead to track the hearer's beliefs from the perspective of one particular world and one particular assignment for each of the referents (it does not matter which one), recording at each step the probabilities of only those world-assignment pairs that are accessible. We display these various states in a tabular format as well as is possible on the two-dimensional page. Intuitively, each pair of a world and an assignment to n discourse referents is a point in an $n+1$-dimensional space, and the numbers in our tables are the probabilities assigned to such points.

The first update consists in the introduction of the new discourse referent x. According to the rules, the four possible assignments are multiplied out and the probability of each world from Table 1 is distributed uniformly over its descendants. Table 3 shows the resulting probabilities for each pair consisting of a world and a possible assignment for x.

	$g(x)$			
w	a	b	c	d
w_1	.075	.075	.075	.075
w_2	.125	.125	.125	.125
w_3	.05	.05	.05	.05

Table 3: Update with "[$\exists x$]"

	$g(x)$			
w	a	b	c	d
w_1	.15	.15	.0	.0
w_2	.25	.25	.0	.0
w_3	.1	.1	.0	.0

Table 4: Update with "[car(x)]

[3]For instance, in w_{12}, a' shatters and d' does not; the probabilities of these events are .9 and .9, respectively, and since the events are independent, the probability is $.9 \times .9 = .81$.

The next piece of information is that x refers to a car. The probability mass is concentrated on the possibilities in which this is the case, as shown in Table 4.[4]

Next, another discourse referent y is activated and multiplied out over the domain. Table 5 shows the possible pairings of worlds with assignments to x and y with their probabilities.

	$g(x)g(y)$							
w	aa	ab	ac	ad	ba	bb	bc	bd
w_1	.0375	.0375	.0375	.0375	.0375	.0375	.0375	.0375
w_2	.0625	.0625	.0625	.0625	.0625	.0625	.0625	.0625
w_3	.025	.025	.025	.025	.025	.025	.025	.025

Table 5: Update with "[$\exists y$]"

The information that y refers to a truck rules out half of the possible assignments, shifting the probability mass to the remaining possibilities. The result is shown in Table 6.

	$g(x)g(y)$							
w	aa	ab	ac	ad	ba	bb	bc	bd
w_1	.0	.0	.075	.075	.0	.0	.075	.075
w_2	.0	.0	.125	.125	.0	.0	.125	.125
w_3	.0	.0	.05	.05	.0	.0	.05	.05

Table 6: Update with "[truck(y)]"

	$g(x)g(y)$			
w	ac	ad	bc	bd
w_1	.0	.375	.0	.0
w_2	.0	.0	.0	.625
w_3	.0	.0	.0	.0

Table 7: Update with "[hit(x, y)]"

The update with the statement that 'x hit y' rules out all possibilities associated with world w_3, since no car hit a truck there. Worlds w_1 and w_2 survive with exactly one assignment each, as shown in Table 7.

Now it is time to take a more fine-grained look at the various possible outcomes for the windshields. Each of the two surviving "worlds" is really a class of four worlds, its probability being distributed over those four according to the conditional probabilities in Table 2 above. The fine-grained probabilities, along with a reminder of the associated extensions of *shatter,* are given in Table 8.

To give an adequate analysis to the update with *the windshield,* we need to incorporate an account of definites and relational nouns into the formal system. For lack of space, we did not include such an account in the definitions in **section 3.** Here we only describe informally how the expression [$\iota z.windshield\_of(z)$] is to be understood. We emphasize that our treatment is simplified to capture what is crucial for the purposes of the illustration; we do not claim that constitutes an adequate linguistic analysis. The whole noun phrase is to be understood as something like *its windshield* or *the vehicle's windshield.* The referent z introduced in the course of its interpretation therefore is set to a vehicle and must corefer with one of the discourse referents already present. Now, clearly the listener cannot be sure which of the two active referents is intended. However,

[4]Clearly the example would be more interesting if there were some uncertainty about the extension of *car,* but we forgo this option here for simplicity. The probabilities in Table 4 are obtained by dividing the probability of each surviving possibility by the sum of their prior probabilities, i.e., $2\times(.075+.125+.05) = .5$.

		$g(x)g(y)$	
w	$I_w(shatter)$	ad	bd
w_{11}	$\emptyset$	.0338	.0
w_{12}	$\{a'\}$	.3038	.0
w_{13}	$\{a', d'\}$	.0338	.0
w_{14}	$\{d'\}$	.0038	.0
w_{21}	$\emptyset$	.0	.3094
w_{22}	$\{b'\}$	.0	.2531
w_{23}	$\{b', d\}$	.0	.0281
w_{24}	$\{d'\}$	.0	.0344

Table 8: More fine-grained view of the state

the definite indicates to her that the speaker has a particular one in mind which he takes to be uniquely identifiable in the context. The ι operator can be defined to model this kind of asymmetric discourse situation. To accomplish this, the result of the update must differ between the respective belief states of speaker $\mathcal{A}$ and hearer $\mathcal{B}$. A definition of the update with $[\iota x]$ would call for the simultaneous execution of the pair of updates in (4).

(4) a. $c_{\mathcal{A}}[\iota x] := \{\langle i', j' \rangle \in c_{\mathcal{A}}[\exists x] : i'(x) = i'(y) \;\&\; j'(x) = j'(y)\}$ for some active referent y;

b. $c_{\mathcal{B}}[\iota x] := c_{\mathcal{B}}[\exists x]$.

Table 9 shows the result for our concrete example. The different shades of gray delineate two sets of possibilities that form disjoint equivalence classes under the speaker's accessibility relation, but not under the hearer's. Intuitively, the it is common knowledge that the speaker does and the hearer does not know whether z co-refers with x or y.

	$g(x)g(y)g(z)$				
w	ada	add	bdb	bdd	
w_{11}	.0169	.0169	.0	.0	.375
w_{12}	.1519	.1519	.0	.0	
w_{13}	.0169	.0169	.0	.0	
w_{14}	.0019	.0019	.0	.0	
w_{21}	.0	.0	.1547	.1547	.625
w_{22}	.0	.0	.1266	.1266	
w_{23}	.0	.0	.0141	.0141	
w_{24}	.0	.0	.0172	.0172	
	.1875	.1875	.3125	.3125	
	.5		.5		

Table 9: Update with "$[\iota z.windshield\_of(z)]$"

	$g(x)g(y)g(z)$				
w	ada	add	bdb	bdd	
w_{11}	.0	.0	.0	.0	.5217
w_{12}	.4226	.0	.0	.0	
w_{13}	.0470	.0470	.0	.0	
w_{14}	.0	.0052	.0	.0	
w_{21}	.0	.0	.0	.0	.4783
w_{22}	.0	.0	.3522	.0	
w_{23}	.0	.0	.0391	.0391	
w_{24}	.0	.0	.0	.0478	
	.4696	.0522	.3913	.087	
	.8609		.1391		

Table 10: Update with "[shatter(z)]"

Colors: z=x z=y

A couple of remarks are in order about Table 9. As the marginal probabilities at the bottom show, the probability that z refers to the windshield of car b is higher (.3125) than that of car a (.1875). This is to be expected because aside from the activation of z, no new information has been acquired, and it was the case all along that car b was more likely to have been involved in the accident. Furthermore, notice that z is just as likely to refer to the windshield of a car (.3125 + .1875 = .5) as it is to refer to that of the truck. This, too, is to be expected because the probabilities from Table 8 are uniformly distributed over the two descendants, one for a car and for a truck, of each possibility.

Finally, we add the information that z shattered. The result of this update is shown in Table 10. A few things are noteworthy about the difference between tables 9 and 10. First, z is now much more likely to refer to the windshield of x, i.e., the windshield of a car (.4696 + .3913 = .8609) than that of the truck (.1391). Moreover, it is more likely to be car a (.4696) than car b (.3913). As a result, the probabilities of the assignments of x have changed as well: It, too, is now more likely to refer to car a (.4696 + .0522 = .5218). These changes are due to the assumptions, built into the probabilities, that the windshields of trucks are generally unlikely to shatter in an accident and that the one of car a is more vulnerable than that of car b, respectively. Finally, notice also that the relative probabilities of the two classes of worlds w_1 and w_2 are flipped. This, too, is a result of the fact that the vulnerable car a was in the accident in the former whereas the sturdy car b was involved in the latter.

Table 11 summarizes the main changes occuring as a result of the update with the information that the windshield shattered.

		before	after
$z = x$	$ada + bdb$	.5	.8609
$z = y$	$add + bdd$	.5	.1391
$x \mapsto a$	$ada + add$	.375	.5217
$x \mapsto b$	$bdb + bdd$	.625	.4783
$z \mapsto a$	ada	.1875	.4696
$z \mapsto b$	bdb	.3125	.3913
$z \mapsto d$	$add + bdd$	.5	.1391
w_1		.375	.5217
w_2		.625	.4783

Table 11: Changes brought about by the update with "[shatter(z)]"

5 Conclusions

The example in **section 4** showed that Probabilistic Dynamic Predicate Logic provides a framework to model the way in which probabilistic world knowledge can affect beliefs about the discourse and simultaneously be affected by incoming information. The last piece of information processed in the example shifted both beliefs about the facts (w_1 vs. w_2) and beliefs about anaphoric coreference ($z = x$ vs. $z = y$). We believe that the framework has many potential applications in linguistic pragmatics and communication in general, and that it can naturally be embedded in models of multi-agent communication and game theory. A detailed study of these extensions is left for future work.

REFERENCES

Asher, N. and A. Lascarides. 2003. Logics of Conversation. Cambridge University Press.

Clark, H. H. and S. E. Haviland. 1977. Comprehension and the given—new contract. In: Roy O. Freedle (ed.). Discourse Production and Comprehension. Norwood, NJ: Ablex. 1–40.

Groenendijk, J. and M. Stokhof. 1991. Dynamic predicate logic. Linguistics and Philosophy 14: 39–100.

Groenendijk, J., M. Stokhof and F. Veltman. 1996. Coreference and modality. In: S. Lappin (ed.). The Handbook of Contemporary Semantic Theory. Blackwell. 179–213.

Hawkins, J. A. 1978. Definiteness and Indefiniteness. Atlantic Highlands, NJ: Croom Helm London.

Levinson, S. 2000. Presumptive Meaning. Cambridge, MA: The MIT Press.

Matsui, T. 2001. Experimental pragmatics: Towards testing relevance-based predictions about anaphoric bridging inferences. In: V. Akman et al. (eds.). Context 2001, LNAI 2116. 248–260.