Probabilistic Linguistic Meaning

Rick Nouwen (Utrecht University)

2024

---

1 Introduction

This short text introduces the idea that linguistic meaning can be thought of in probabilistic terms, extending older common views on what meaning is and using basic concepts from information theory (Shannon, 1948). The goal of this text is to help students grasp the philosophy behind contemporary computational theories of semantics and pragmatics. This text will not review or introduce these theories, but rather focus solely on the conceptual underpinnings of the very idea that meaning could be probabilistic.

2 Meaning as update

The following advice by the philosopher David Lewis has been hugely influential in our thinking about meaning: *"In order to say what a meaning is, we may first ask what a meaning does and then find something that does that"* (Lewis, 1970). A common answer to what meanings *do* is that they allow language users to navigate a hypothesis space that concerns the interlocutors and the world they inhabit. Humans are not omniscient – there are always things we don't know. For instance, I may be in the dark as to whether or not Sue has any siblings. This creates a hypothesis space:

Sue is an only child	**Sue is not an only child**

A hypothesis space is a set of alternative states of affairs, each of which we entertain as possibly being the *actual* state of affairs. I will refer to these alternatives as (alternative) *possibilities*. Linguistic meaning is the property that sentences have that allows interlocutors to narrow down this space of options that are still compatible with what we know. When Sue says *'I have a brother'* , then (providing I trust Sue) I can **update** the hypothesis space by removing the option that is now no longer compatible with Sue's utterance. As a result, the hypothesis space has been resolved: there is just the single possibility left.

Sue is an only child	**Sue is not an only child**

$\rightsquigarrow$

Sue: "*I have a brother*"

$\rightsquigarrow$

Sue is not an only child

It is often assumed that all linguistic utterances address an implicit (and sometimes explicit) *Question under Discussion* (QUD). For Sue's *'I have a brother'* that could for instance be the question 'Does Sue have siblings?'. Crucially, it is this question that determines what hypothesis space we are considering. That is, the hypothesis space induced by a question under discussion is the set of its possible answers. For a binary QUD like 'Does Sue have siblings?', we get a binary hypothesis space, such as the one above.

It is the conversational context that determines the QUD, not the uttered sentence. For instance, Sue's utterance of 'I have a brother' is also compatible with a situation in which her house burned down and the QUD is whether or not she has a place to stay, where (presumably) we will interpret her utterance as conveying that she can move in with her sibling.

$Context$: *Sue's house just burned down*

$\rightsquigarrow$

Sue has somewhere to stay	**Sue does not have somewhere to stay**

$\rightsquigarrow$

Sue: "*I have a brother*"

$\rightsquigarrow$

Sue has somewhere to stay

QUDs can induce more complex spaces than the simple example above and in these cases utterances may not fully resolve the question. For instance, if the QUD at hand is 'How many Harry Potter books did Sue read?' we get a hypothesis space like this (where the numbers are shorthand for the number of Harry Potter novels Sue read).

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**

If Sue now tells me

φ: I only read the first three Harry Potter books then the hypothesis space is fully resolved, as the only possibility from the above space that is compatible with Sue's utterance is the cell labelled **3**. If, however, Sue tells me

ψ: I read more than one Harry Potter book

then the question is not fully resolved. We can only remove the first two possibilities:

2	**3**	**4**	**5**	**6**	**7**

As such, φ is more informative than ψ. This is because φ is compatible with fewer possibilities than ψ is. As such, φ has to potential to *remove* more possibilities than ψ does. Formally, we can express the relative informativeness of two sentences in terms of logical entailment. If sentence S1 entails S2, then S2 is true in all situations that make S1 true. This means that S2 is compatible with at least the possibilities that S1 is compatible with and possibly with more. This means that the number of cells removed by an utterance of S2 is smaller or equal than the number of cells removed by an utterance of S1. So, if S1 entails S2, S1 is at least as informative as S2.

Let's write $\pi(S, Q)$ for the set of possibilities in hypothesis space Q that sentence S is compatible with. Take Q to be the space induced by the question 'How many Harry Potter books did Sue read?'. Now, $\pi(\varphi, Q) = \{\mathbf{3}\}$ and $\pi(\psi, Q) = \{\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}\}$. We can see that $\varphi \models \psi$ (φ 'entails' ψ), since $\pi(\varphi, Q) \subset \pi(\psi, Q)$. As a consequence, φ is more informative than ψ.

3 Probabilistic update

We often have certain expectations concerning the hypothesis spaces that language allows us to navigate. For instance, the picture to the right is the view from my office of Utrecht's Dom tower.[1] Every day that I go to work, I see this view. So, every day that I arrive at my work, I fully expect the Dom tower to still be there, just like it always has been. But I don't *know* this, since something could have happened overnight, without me realising. Consequently, on an average morning before I've gone to the office (and before I read any news), there is the hypothesis space: {The Dom tower is gone, The Dom tower is fine}. In such a situation, however, I will probably consider the probability that the Dom tower is gone to be much, much lower than the probability that it is fine.

[1]The Dom tower in Utrecht is the tallest church tower of the Netherlands. The tower was built between 1321 and 1382.

So, the possibilities in hypothesis spaces come with subjective probabilities, measures of what we expect to be the case. For instance, in the context just sketched a hypothesis space could look like this:

The Dom tower is gone	**The Dom tower is fine**
0.00001	0.99999

Adding probabilities like this turns hypothesis spaces into *random variables*. That is, a question under discussion induces a sample space, namely the set of possibilities, and a probability function that returns a probability for each possibility in the space. In other words, our probabilistic expectations with respect to a hypothesis space turn that space into a probability distribution.

Now, consider two alternative language statements, (1-a) or (1-b):

(1) a. The Dom tower collapsed.
    b. The Dom tower didn't collapse.

It should be clear that given the background given by the hypothesis space, it would be much more surprising that (1-a) is the case than (1-b). We can express this with Shannon's measure of *surprisal* I:

$$I(x) = -\log(P(x))$$

Since (1-a) identifies the possibility that we assigned a probability of 0.00001, we can associate that utterance with a surprisal value of $-log(0.00001) \approx 16.61$. (Here and in what follows, I use base 2 for logarithms, but nothing substantial rests on what base is used.) For (1-b), the surprisal is much lower, namely $-log(0.99999) \approx$ 0.000014. So, (1-a) comes with a much higher surprisal than (1-b). These surprisal values have an interpretation both from the comprehension and from the production perspective. The high surprisal value of (1-a) means that this sentence is not expected to be true. The low surprisal value of (1-b), however, means that this sentence is not expected to be uttered. Since the Dom tower hasn't fallen down for centuries and since it is not expected to fall down, it seems a waste of energy to assert that (once again) it hasn't collapsed. This in turn triggers a pragmatic effect whenever a sentence like (1-b) *is* uttered. If I wake up one morning an see (1-b) as a headline, I am likely to conclude that my presumptions about the hypothesis space (in particular that it was extremely unlikely that anything would happen to the Dom tower overnight) were wrong. That is, I conclude from it not just that the Dom tower is fine, but also that it was likely not to be fine.

Our expectations do not need to be so biased as in the Dom tower example. Say it's winter and I built a snowman. I wake up the next morning and wonder whether the snowman has survived the night. If I really have no idea, then the possibilities in the hypothesis space would be equiprobable:

The snowman is gone	The snowman is fine
0.5	0.5

As a consequence, whatever way this question under discussion is resolved, the surprisal of that resolution will always be $-log(0.5) = 1$. We can quantify the intuition that our snowman scenario contained less prior uncertainty than our Dom tower scenario by calculating the expected surprisal. This is the information theoretic notion of *entropy*. Let $\mathcal{X} = \langle X, P \rangle$ be a hypothesis space where $X = \{x_1, x_2, \ldots, x_n\}$ is the set of possibilities and P a function that provides probabilities for the possibilities in such a way that they sum up to 1, i.e. $\sum_{x \in X} P(x) = 1$.

$$H(\mathcal{X}) = -\sum_{x \in X} P(x) \log(P(x))$$

Doing the maths results in:

(2) a. H(Dom tower scenario) = - (1e-5$\cdot$ $\log$(1e-5)+1-(1e-5)$\cdot$ $\log$(1-(1e-5))) = 0.00018
b. H(snowman scenario) = - (0.5$\cdot$ $\log$(0.5)+0.5$\cdot$ $\log$(0.5)) = 1

The intuition is the following. Entropy is the expected surprisal. In the Dom tower scenario there is an option with very high surprisal, but that option is extremely unlikely, so it is unlikely that this surprise becomes actual. In other words, the weight of that high surprisal in the calculation of expected surprisal is minute. The low surprisal of the other possibility weighs much more heavily and, as a result, the entropy is low. In the snowman scenario both possibilities have a surprisal of 1 and both have the same probability. On average then, we expect a surprisal of 1.

Note that once a question under discussion has been fully resolved there is just a single possibility left. As a result that possibility will have probability 1. The entropy of that (trivial) hypothesis space will be $1 \cdot \log(1)$ = 0. That is, there is now **no** uncertainty.

If we have complete uncertainty about a hypothesis space then that means that all possibilities are equi-probable. For a set of possibilities X of cardinality n this means that $P(x) = \frac{1}{n}$ for all $x \in X$. Such probability distributions are sometimes called *flat* or *uninformative* and they provide the highest possible entropy for a given set. Any distribution over the same space that is not equi-probable will result in lower entropy, i.e. less uncertainty.

For example, here are two hypothesis spaces for *'How many Harry Potter books did Sue read?'*, together with their entropy. (Checking these should be an easy exercise.) As you can see, adding specific expectations to the hypothesis space (e.g. Sue is more likely to have read few books than many) lowers the uncertainty.

0	1	2	3	4	5	6	7	entropy = 3
1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	entropy = 2.35
0.4	0.3	0.05	0.05	0.05	0.05	0.05	0.05	

4 Semantic update and conditional probability

The probabilities we assigned to possibilities in the previous section were simple subjective probabilities. The effect of an assertive utterance on the context can be expressed in terms of *conditional probabilities*. That is, updating the context on the basis of an accepted assertion means that we update the contextual probability distribution to one conditioned on the assertion being true.

Take once more the QUD that wants to know how many Harry Potter books Sue has read and let's assume all possibilities are equi-probable. We then have, for instance, that $P(\mathbf{3}) = \frac{1}{8}$. We can see the effect of an utterance as a conditional probability. Take the sentence ψ (*'Sue read more than one Harry Potter book'*). The conditional probability $P(\mathbf{3}|\psi)$ expresses the probability that Sue read 3 Harry Potter books, given that we know that she read more than one. Recall that:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Let's apply this. $P(\mathbf{3}|\psi) = \frac{P(\mathbf{3} \wedge \psi)}{P(\psi)}$. Let's first calculate $P(\psi)$. Well, ψ is true in all possibilities except for **0** and **1**. That means that the probability of ψ being true is $\frac{6}{8}$. What is the probability that both **3** is the actual possibility while at the same time ψ being true? That is $\frac{1}{8}$. And, so, $P(\mathbf{3}|\psi) = \frac{1}{6}$.

So, updating with ψ gives us a new probability distribution, namely the distribution $P(x|$ Sue read more than one Harry Potter book) with the following values:

0	**1**	**2**	**3**	**4**	**5**	**6**	7
0	0	1/6	1/6	1/6	1/6	1/6	1/6

And $P(x|$ Sue read the first three Harry Potter books) is the distribution:

0	**1**	**2**	**3**	**4**	**5**	**6**	7
0	0	0	1	0	0	0	0

Things proceed in exactly the same way in cases where the original expectations for the possibilities are not all equi-probable. For instance, if the prior expectation are as follows

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0.4	0.3	0.05	0.05	0.05	0.05	0.01	0.09

it should be easy to check that $P(x|\text{'Sue read more than one Harry Potter book'})$ is the following distribution:

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	0	0.17	0.17	0.17	0.17	0.03	0.30

5 The speaker's perspective

The conditional probability distributions in the previous section were of the form $P(p|u)$: the probability that p is the actual possibility given the fact that utterance u has taken place. This takes the view of the *hearer*, since it tells us how to view the world on the basis of what the speaker has told us. In this section we take the view of the speaker by turning things around. We want to predict $P(u|p)$: the probability that the speaker performs utterance u given the fact that she believes p to be the actual possibility.

It is common to take a Gricean perspective when reasoning about the choices that hearers and speaker have in conversational contexts. This means that there is an assumption that choices made by hearer and speaker in social situations are such that they serve a common communicative goal. One example of this is that speakers are as informative as they can be without compromising truthfulness.

Say that a speaker believes that Sue read all Harry Potter books and say that this speaker is deciding between uttering one the following three sentences:

(3) Sue read none of the Harry Potter books.

(4) Sue read all of the Harry Potter books.

(5) Sue read some of the Harry Potter books.

How does the speaker make this choice? Intuitively, we would want to say that sentence (3) is useless for this speaker, since it does not match her beliefs. Sentence (4) is intuitively a good choice, since it does match her beliefs. Finally, this speaker will judge sentence (5) to be true, but not as useful an utterance as (4), since it is compatible with a lot more situations than the one she believes to be the case. For these reasons, (4) is clearly the best sentence to utter, (3) should not be considered and (5) should be dispreferred.

These intuitions can quite easily be captured probabilistically, if we assume that the *utility* of a sentence is determined by the likelihood that a hearer matches your beliefs once they believe this sentence to be true. In other words, the utility of a sentence depends on the surprisal of the actual possibility, conditioned on the sentence being true. A sentence is useful, if the actual possibility has low surprisal, once we assume the sentence to be true. If on the assumption that the sentence is true the suprisal of the actual possibility is *high*, then the sentence has low utility, since the sentence is not helping us to identify the actual possibility. In other words, we can quantify the utility of a sentence given some possibility believed to be actual, by simply taking the negative surprisal:[2]

$$U(u, p) = log(P(p|u))$$

Let's assume that we have no prior expectations about how many Harry Potter books Sue read:

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

We then get the following conditional probability distributions for the three sentences:

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	sentence (3), none
1	0	0	0	0	0	0	0	

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	sentence (4), all
0	0	0	0	0	0	0	1	

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	sentence (5), some
0	1/7	1/7	1/7	1/7	1/7	1/7	1/7	

Using these to calculate the utility of the three sentences given the belief that Sue read all 7 Harry Potter books, we get the following:

Table 1: Probabilities and utilities for sentences in possiblity **7**

sentence uttered in **7**	$P(\mathbf{7} \mid u)$	$U(u, \mathbf{p})$
(3), none	0	-∞
(4), all	1	0
(5), some	1/7	-2.81

In other words, sentence (4) has the highest utility out of the three sentences and sentence (3) has the lowest possible utility. The next question is how the speaker could use this information to make a decision about the sentence she should utter. Here, there are various possible models of speaker behaviour. One is that the speaker simply chooses the most optimal sentence. That is, the speaker chooses the sentence with highest utility, which in this case is (4), with a utiltiy of 0. Another option, is that the speaker's choice can be captured probabilistically. That is, sentences

[2]This is notoriously confusing terminology. For clarity: surprisal is the negative logarithm of a probability. Since the logarithm of a probability is a negative number, suprisal amounts to a positive number. Negative surprisal will therefore be the positive logarithm of a probability and is therefore a negative number.

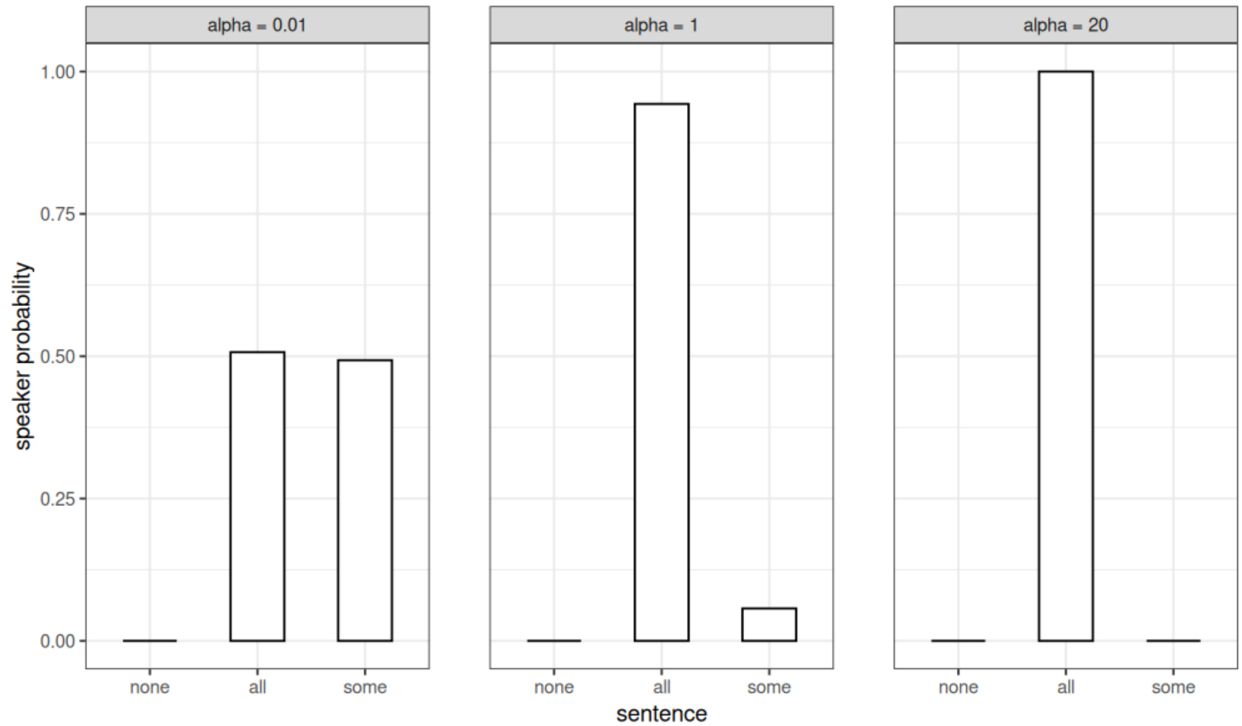

Figure 1: The effect of the temperature parameter α in the softmax on speaker behaviour for a speaker who believes **7** to be actual. Each panel shows the probability that the speaker chooses a specific sentence on the basis of the utilities in table 1, given some value for α.

with high utility are chosen more often than sentences with lower utility. Since information about how exactly utilities map to choices is latent, this kind of probabilistic choice is often modelled using a softmax function with a temperature parameter. The function takes a vector of values $v = v_1 v_2 v_3 \dots v_n$ and returns a probability distribution of the same length.

$$S(v_i) = \frac{e^{\alpha v_i}}{\sum_{j=1}^{j=n} e^{\alpha v_j}}$$

We can use the softmax to model the speakers choice:

$$P_{\text{speaker}}(u|p) = \frac{e^{\alpha U(u,p)}}{\sum_{u'} e^{\alpha U(u',p)}}$$

The three plots in figure 1 show the effects of the softmax.

With values for alpha closer to 0, the softmax diminishes the role that utility plays in the speaker's choice. With higher values, it amplifies that role, making the speaker favour the best option over all others.

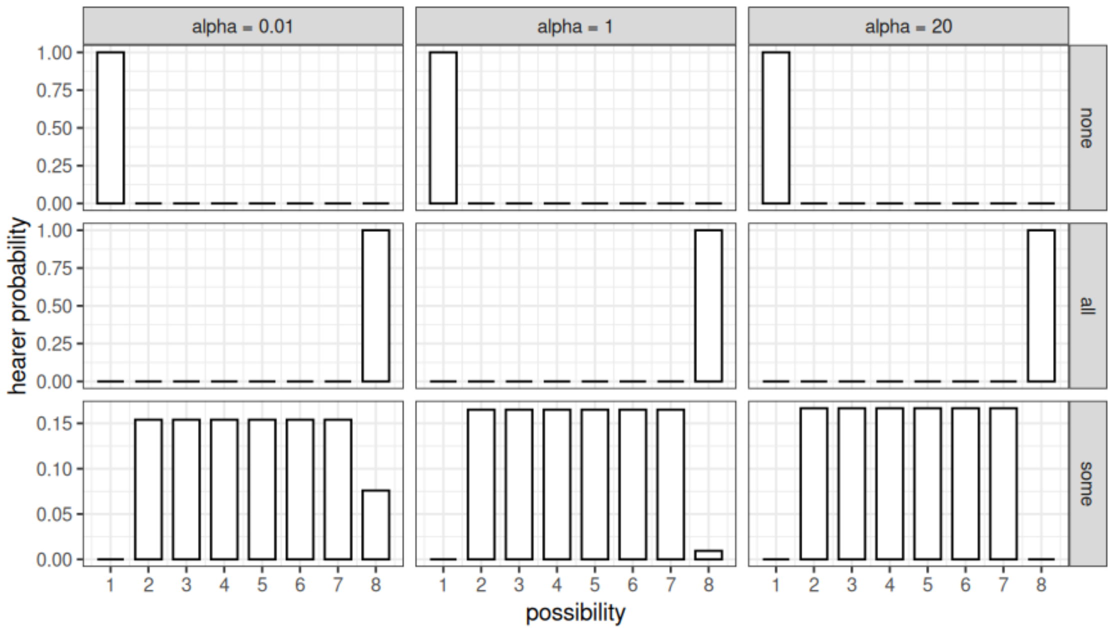

Figure 2: Predictions of the simple hearer model presented in section 6, for the QUD 'how many Harry Potter books did Sue read?', given flat expectations for the possible answers to that QUD.

6 The hearer's perspective

So far, we've seen how to use the surprisal of the content of a sentence as a basis for modeling the choice that a speaker makes when she wants to convey her beliefs about where in the hypothesis space the actual possibility is. How could we now model a cooperative hearer? A hearer receives a certain sentence in the form of some utterance. The hearer's task is to select the possibility that best fits this sentence, given the choice that the speaker made. In other words, the hearer chooses a possibility proportional to the likelihood that the speaker would have uttered the given sentence had she been in that possibility. As a result, we can model hearer choice as a simple application of Bayes' law:

$$P_{\text{hearer}}(p|u) = \frac{P_{\text{speaker}}(u|p)P(p)}{\sum_{p'} P_{\text{speaker}}(u|p')P(p')}$$

Figure 2 shows the result of applying this equation to the Harry Potter example. The plots are the results given a flat prior probability for the possibilities and they show how adjusting assumptions about the speaker (via the value of α) yields subtly different predictions for the hearer. The key prediction made here is that hearers draw a so-called *implicature* from the sentence with *"some"*. If a speaker chooses to use that quantifier, she is unlikely to believe that the corresponding *"all"* sentence is true, since in that case her utility of using *"all"* would have trumped the much lower utility of using *"some"*.

According to the hearer model above, the role of an utterance in a conversation is that it allows a hearer to update a prior distribution over a hypothesis space to a posterior one. In section 4, I showed a semantic way of doing this, namely by simply updating the prior probability $P(x)$ to $P(x|u)$, the conditional probability given the assumption that the utterance is true. The model above constitutes a *pragmatic* update. It models how the prior changes, based on what we now about the likelihood of the speaker performing utterances, given that a certain sentence has been uttered.

If you want to understand how 2 came about: Appendix A shows the calculations that lead up to the plots in figure 2.

7 Further readings

In this final section, I give pointers to the literature where the ideas above emerge and / or are introduced with more depth.

First of all, the idea that the role that (propositional) meaning plays in linguistic communication is one of updating a space of possibilities is mostly attributed to Stalnaker (1978). It was hugely influential in philosophy of language and forms one of the conceptual pillars of certain strands of dynamic semantics (Groenendijk and Stokhof 1991; Veltman 1996, see also Nouwen et al. 2016.) The notion of *question under discussion* was introduced in Roberts (1996, 2012).

Much of the above rests on Paul Grice's view of linguistic communication as a cooperative, social activity (Grice, 1975, 1978) and the Gricean and neo-Gricean theories of pragmatics that were subsequently developed (Horn, 1984; Levinson, 2000), as well as proposals that cast Gricean ideas in game-theoretic terms (Parikh, 1991, 1992; Blutner, 2000; Van Rooij and Franke, 2006). The game-theoretical basis for communication can be traced back to Lewis (1969).

The probabilistic, quantitative use of the notion of update to model semantic and (especially) pragmatic phenomena was mostly developed within game-theoretical approaches to meaning (Benz et al., 2006; Franke, 2009). Later, similar theories were developed in psychology (Frank et al., 2009; Frank and Goodman, 2012), leading to the rational speech act framework (Scontras et al., 2021; Degen, 2023), resulting in a wealth of computational approaches to semantic and pragmatic phenomena (e.g. Bergen et al. 2016; Lassiter and Goodman 2017; Yoon et al. 2020; Nouwen 2024b,a). What I presented above in section 5 and 6 is an information-theoretic rationale for the basic setup of that framework.

References

Benz, A., Jäger, G., and Van Rooij, R. (2006). An introduction to game theory for linguists. In *Game theory and pragmatics*, pages 1–82. Springer.

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9:ACCESS–ACCESS.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of semantics*, 17(3):189–216.

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, 9:519–540.

Frank, M., Goodman, N., Lai, P., and Tenenbaum, J. (2009). Informative communication in word production and word learning. In *Proceedings of the annual meeting of the cognitive science society*, volume 31.

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

Franke, M. (2009). *Signal to act: game theory in pragmatics.* PhD thesis, Universiteit van Amsterdam.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Speech acts*, volume 3, pages 41–58. Academic Press, New York.

Grice, H. P. (1978). Further notes on logic and conversation. In *Pragmatics*, pages 113–127. Brill.

Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and philosophy*, pages 39–100.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In Schiffrin, D., editor, *Meaning Form and Use in Context*, pages 11–42. Georgetown University Press.

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

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* MIT press.

Lewis, D. (1969). *Convention: A philosophical study.* John Wiley & Sons.

Lewis, D. (1970). General semantics. *Synthese*, 22:18–62.

Nouwen, R. (2024a). Meiosis and hyperbole as scalar phenomena. In Liefke, K., Anans, R., Gutzmann, D., and Scheffler, T., editors, *Sinn und Bedeutung 28*. Universität Konstanz.

Nouwen, R. (2024b). The semantics and probabilistic pragmatics of ad-adjectival intensifiers. *Semantics & Pragmatics*, 17.

Nouwen, R., Brasoveanu, A., van Eijck, J., and Visser, A. (2016). Dynamic semantics. *Stanford Encyclopedia of Philosophy*.

Parikh, P. (1991). Communication and strategic inference. *Linguistics and Philosophy*, 14:473–514.

Parikh, P. (1992). A game-theoretic account of implicature. In Moses, Y., editor, *Proceedings of the 4th conference on Theoretical aspects of reasoning about knowledge*, page 85–94. Morgan Kaufmann Publishers, San Francisco, CA.

Roberts, C. (1996). Information structure in discourse: Toward a unified theory of formal pragmatics. *Ohio State University Working Papers in Linguistics*, 49:91–136.

Roberts, C. (2012). Information Structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5:6:1–69.

Scontras, G., Tessler, M. H., and Franke, M. (2021). A practical introduction to the rational speech act modeling framework. *arXiv preprint arXiv:2105.09867*.

Shannon, C. (1948). A mathematical theory of communication. *Bell Labs Technical Journal*, 27:379–423.

Stalnaker, R. C. (1978). *Assertion*. Brill.

Van Rooij, R. and Franke, M. (2006). Optimality-theoretic and game-theoretic approaches to implicature. *Stanford Encyclopedia of Philosophy*.

Veltman, F. (1996). Defaults in update semantics. *Journal of philosophical logic*, 25:221–261.

Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2020). Polite Speech Emerges From Competing Social Goals. *Open Mind*, 4:71–87.

A Calculations

Truth values			
	none	some	all
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	1.00	0.00
4	0.00	1.00	0.00
5	0.00	1.00	0.00
6	0.00	1.00	0.00
7	0.00	1.00	0.00
8	0.00	1.00	1.00

Utility			
	none	some	all
1	0.00	$-\infty$	$-\infty$
2	$-\infty$	-2.81	$-\infty$
3	$-\infty$	-2.81	$-\infty$
4	$-\infty$	-2.81	$-\infty$
5	$-\infty$	-2.81	$-\infty$
6	$-\infty$	-2.81	$-\infty$
7	$-\infty$	-2.81	$-\infty$
8	$-\infty$	-2.81	0.00

$\alpha = 0.01$

Speaker			
	none	some	all
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	1.00	0.00
4	0.00	1.00	0.00
5	0.00	1.00	0.00
6	0.00	1.00	0.00
7	0.00	1.00	0.00
8	0.00	0.49	0.51

Hearer			
	none	some	all
1	1.00	0.00	0.00
2	0.00	0.15	0.00
3	0.00	0.15	0.00
4	0.00	0.15	0.00
5	0.00	0.15	0.00
6	0.00	0.15	0.00
7	0.00	0.15	0.00
8	0.00	0.08	1.00

$\alpha = 1$

Speaker			
	none	some	all
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	1.00	0.00
4	0.00	1.00	0.00
5	0.00	1.00	0.00
6	0.00	1.00	0.00
7	0.00	1.00	0.00
8	0.00	0.06	0.94

Hearer			
	none	some	all
1	1.00	0.00	0.00
2	0.00	0.17	0.00
3	0.00	0.17	0.00
4	0.00	0.17	0.00
5	0.00	0.17	0.00
6	0.00	0.17	0.00
7	0.00	0.17	0.00
8	0.00	0.01	1.00

$\alpha = 20$

Speaker			
	none	some	all
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	1.00	0.00
4	0.00	1.00	0.00
5	0.00	1.00	0.00
6	0.00	1.00	0.00
7	0.00	1.00	0.00
8	0.00	0.00	1.00

Hearer			
	none	some	all
1	1.00	0.00	0.00
2	0.00	0.17	0.00
3	0.00	0.17	0.00
4	0.00	0.17	0.00
5	0.00	0.17	0.00
6	0.00	0.17	0.00
7	0.00	0.17	0.00
8	0.00	0.00	1.00