A Computational Pragmatics for Weaseling

Leander Vignero

Center for Logic and Philosophy of Science, KU Leuven

Probabilistic expressions (PEs) have long been studied in linguistics and related fields, and are noted for their vagueness (Kent, 1964). Such expressions include words like *possibly* and *probably.* They express degrees of belief in propositions. Put differently, they lexicalize judgments about uncertainty. Their semantics proves highly elusive; the central problem being that, even though PEs lexicalize uncertainty, they rarely correspond to precise probabilities. To make things more complicated, there is hard evidence that the scopes of possible interpretation of PEs are governed by probability distributions themselves (Mosteller and Youtz, 1990). This gives rise to higher-order probability distributions.

The interpretation of vague language is the purview of pragmatics, which can be studied at the intersection of a plethora of fields, including cognitive science, computer science, linguistics, and philosophy of language. In this paper, the usage and understanding of probabilistic expressions is viewed through the lens of recent developments in computational pragmatics.

Specifically, several enriched Rational Speech Act (RSA) frameworks are developed (Frank and Goodman, 2012). The RSA framework is a Bayesian, computational model of communication and should be understood as a low-level implementation of the key premise of Gricean pragmatics (Grice, 1975) and Relevance Theory (Sperber and Wilson, 1996). Put simply, this framework operationalizes the Gricean idea of optimal communication probabilistically. In spite of its relative novelty, the RSA framework has been used to model a vast array of linguistic phenomena: metaphor (Kao et al., 2014), politeness (Yoon et al., 2016, 2017, 2018), scalar implicature (Goodman and Stuhlmüller, 2013), and adjectival interpretation (Lassiter and Goodman, 2017). Moreover, the framework has also been used to offer solutions to philosophical problems, like the sorites paradox (Lassiter and Goodman, 2017).

The setup of this paper is to develop several RSA-style frameworks in order to model the pragmatics of PEs. In total, I propose four models to think about PEs and situations in which they naturally occur. Moreover, I also provide data from simulations to back up the claims I make about the power of my framework for understanding the pragmatics of PEs.

1. The first model is a classic variation of the RSA-model for probabilistic expressions. **It extends the basic framework in such a way that the Bayesian agents are capable of incorporating the nuance of a PE**. This is done by replacing the simple truth-conditional interpretation of PEs by an empirically informed probability distribution. In this model, speakers still attempt to communicate world states to the listeners. The model does this by using a utility function that, simply put, assigns higher utilities to words that have a better chance of communicating the intended message of the speaker. For example: if a speaker is certain that there is a 70% chance of something happening, the model proposes that the speaker uses the utterances that have the highest probabilities of communicating the intended meaning. Listeners will also understand that these utterances where chosen to maximize the probability that they, the listeners, would understand the intended message, and they take this into consideration when interpreting the utterance.

2. The second model is of a more epistemic nature: speakers no longer try to communicate the state of the world, but rather they try to communicate their epistemic appraisal of

reality. **In this framework this means that they try to induce epistemic states in their listeners that lie as close as possible to their own epistemic states**. I draw inspiration from Bayesian epistemology and operationalize the notion of epistemic appraisal as a probability distribution that the speakers want to communicate to their interlocutors. Note that these are actually the aforementioned higher-order probability distributions. This sets us up to operationalize the idea of similarity of epistemic states as the Kullbach-Leibler divergence. The utility function that determines how well an utterance succeeds at communicating an intended meaning, is now redefined in terms of the Kullbach-Leibler divergence. Bergen and Goodman (2015) use similar information-theoretic ideas to model prosody.

3. **The third model incorporates non-epistemic motives in communication**. This is done by modifying the utility function. I borrow ideas from Yoon et al. (2016) to model the balance between epistemic and non-epistemic motives in communication. The main idea is that one can use convex combinations of epistemic and non-epistemic utility functions to generate a total utility function.

4. The **fourth model aims to improve on the second and third model** by addressing its principal drawback: one has to postulate of possible probability distributions that the speaker might try to communicate, which is known to the listeners. In many cases, this will not be a plausible way of modeling real human communication and conversation. The key idea is that we can use the model to create a canonical set of probability distributions that the speaker can choose to communicate.

Such models are not just of interest to linguists; I also provide a brief discussion of phenomena like *plausible deniability* and *avoiding accountability*, which are also studied studied in argumentation theory (e.g., Walton, 1996). and cognitive science (e.g., Lerner and Tetlock, 1999) respectively. This kind of deceitful use of vague verbiage is colloquially known as *weaseling*. I therefore argue that an RSA-style model of PEs can easily provide a theory of how rational agents can and should engage in weaseling. From a high-level perspective, rational weasels should exploit Grice's principle of cooperation to deceive their interlocutors. It is worth noting that Yoon et al. (2016, 2017, 2018) already applied an RSA model to disingenuous speech, specifically to politeness.

The young, yet blossoming research concerning the RSA framework has a strong tradition of providing implementations of the proposed models (Yoon et al., 2016; Scontras et al., 2017) and sometimes provides data from simulations (Lassiter and Goodman, 2017). In this paper, I do not merely intend to continue this tradition but to go beyond *toy models*— highly simplified representations of reality, that aim to illustrate a model. Recent scholarship by Douven and Rott (2018) in formal epistemology has shown that toy models seldom tell the whole story, and that formal and computational theories are best understood and evaluated through larger, more realistic simulations. I also use these simulations to evaluate both the epistemic and computational practicality of the proposed models.

References

Bergen, L. and Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in cognitive science*, 7(2):336–350.

Douven, I. and Rott, H. (2018). From probabilities to categorical beliefs: Going beyond toy models. *Journal of Logic and Computation*, 28(6):1099–1124.

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

Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and semantics 3: Speech arts*, pages 41–58. New York: Academic Press.

Kao, J., Bergen, L., and Goodman, N. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the annual meeting of the Cognitive Science Society*, volume 36.

Kent, S. (1964). Words of estimated probability. *Intelligence Studies*, (8):49–65.

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

Lerner, J. S. and Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, 125(2):255.

Mosteller, F. and Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, pages 2–12.

Scontras, G., Tessler, M. H., and Frank, M. C. (2017). Probabilistic language understanding: An introduction to the rational speech act framework. Technical report, Retrieved 2018-10-30 from https://www.problang.org/.

Sperber, D. and Wilson, D. (1996). *Relevance: Communication and Cognition*. Oxford: Blackwell, 2nd edition.

Walton, D. (1996). Plausible deniability and evasion of burden of proof. *Argumentation*, 10(1):47–58.

Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th annual conference of the cognitive science society*, pages 2771–2776. Cognitive Science Society.

Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2017). I wont lie, it wasnt amazing: Modeling polite indirect speech. In *Proceedings of the thirty-ninth annual conference of the Cognitive Science Society*.

Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2018). Polite speech emerges from competing social goals. *PsyArXiv. November*, 20.