Use and interpretation of probability expressions under high-order uncertainty

Introduction. What is the meaning of possibility and probability expressions (PPE) such as *possibly*, and *probably*? According to the classic work by Kratzer (1991) these can be analyzed as quantifiers over possible worlds. Recent works (Yalcin, 2010, Lassiter, 2011, Moss, 2015) agree on the necessity of adding probability measures to the semantics. For example: $[\![\text{probably(p)}]\!]_w = 1$ *iff* $P_w(p) > \theta$, where P_w is a probability measure and θ a threshold. Most previous empirical research (Beyth-Marom, 1982; Teigen, 1988; Windschitl and Wells, 1998) focuses on cases of known objective chance: if a speaker knows the odds of event p, would she use *possibly(p)*, *probably(p)*, or other PPE? And, what is the probability associated to p when a listener interprets *probably(p)*? However, everyday situations often do not involve knowledge of the objective chance. One says that it will probably rain tomorrow, but one has only a vague idea of the odds, based on incomplete information. We investigate the use and interpretation of PPE when the agents have subjective uncertainty about the objective chance, i.e. higher-order uncertainty.

Production study. We manipulated higher-order uncertainty in the setting of an urn with 10 colored balls (red and blue) with partial access. Drawing 4 balls and observing that 3 are red and drawing 8 and observing that 6 are red amount to the same observed proportion ($= 0.75$) but the associated uncertainty about the exact content of the urn is different. Does higher-order uncertainty play a role in the use of PPE? To answer this question we ran an experiment on Mechanical Turk ($N = 50$). In each trial the participants saw a picture representing an urn scenario and completed a sentence of the form *The next ball will* [.] *be red* choosing among *certainly not*, *probably not*, *possibly*, *probably*, *certainly*. Participants also estimated the odds of a randomly drawn ball being red. Results are visualized in Figure 1. We ran multinomial logistic regressions to test whether choices of PPE depend only on participants' beliefs about the prejacent (`simple` model), or on the structured high-order uncertainty state, as a function of drawn and observed balls (`interaction`). AIC scores of best-fits of the models are given in Table 1. Despite added complexity, `interaction` is better than `simple`, supporting our intuition that higher-order uncertainty affects choices of PPE.

Computational model. To investigate the use of PPE under high-order uncertainty and pin down the role of access and observation, we developed a probabilistic model based on the *Rational Speech Acts* model by Goodman and Stuhlmüller (2013). The overall assumption is that PPE have simple threshold semantics which refers to objective chance and the speakers use them pragmatically. As usual, the model is grounded in the definition of a listener who interprets the messages literally. Skipping the technical details, we model the use of PPE as the behavior of a pragmatic speaker who makes an observation of the urn, forms a rational Bayesian belief and chooses an utterance by reasoning on the literal listener. The model has a number of free parameters, which we estimated by Bayesian inference. Most notably, the mean inferred value for the threshold θ equals 0.55: it speaks in favor of the model that data-driven inference recovers this value, in line with intuition and previous research, without stipulating it from the start. The correlation between data and predictions of the fitted model is quite high (Table 2). The AIC score for the best-fit of the pragmatic model equals 270.90, much better than 635.93 of the best regression model.

Interpretation study. Applying Bayes' rule to the pragmatic speaker function (with the appropriate prior) we model a pragmatic listener who reasons about the speaker's choice and infers the distribution over the high-order uncertainty state and the state of the world. To test the predictions we ran an experiment on MT (preliminary data, $N = 58$). In each trial the participants read one of the five messages of the production study and we measured their intuitions about how many balls were drawn by the sender of the message (access) and how many of them were red (observation). For each message, participants also estimated the number of red balls contained in the urn. These results are visualized in Figure 2 with the predictions of the fitted model. The overall tendency of the data is captured reasonably well. The data obtained for the access/observation measure are harder to visualize and more noisy. Correlation results are reported in Table 3.

Conclusion. Although this is work in progress, data and modeling preliminarily suggest that subjects are able to reason about high-order uncertainty in line with a Bayesian model of rational belief formation and that they condition their choice of PPE on their full uncertain belief.

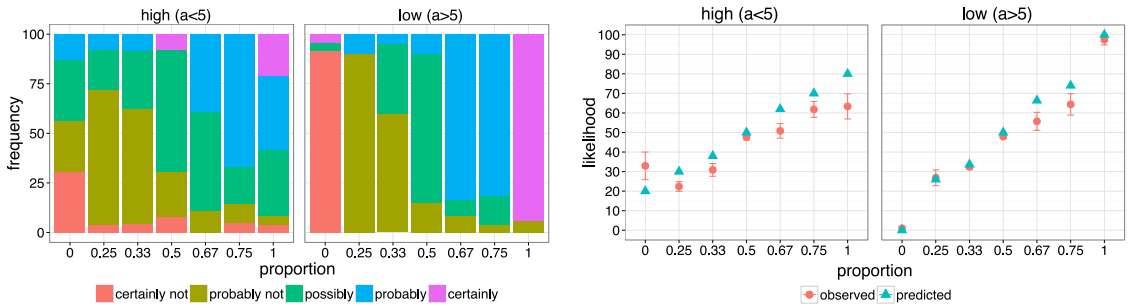

Figure 1: Results of the production study: message choices (left) and estimated objective chance (right). Uncertainty levels *high* and *low* depend on the number of drawn balls (i.e. access, respectively, lower than 5 and higher than 5). Proportion levels express the number of observed red balls among the drawn ones.

	`simple`	`complex`	`ineraction`
AIC	662.78	646.69	635.93

Table 1: AIC scores of best-fits of regressions models. The `simple` model predicts the categorical factor `answer` (*certainly not*, *probably not*, *possibly*, *probably*, *certainly*) with a single metric factor `belief` which is the participants' mean estimate in each condition. The `complex` model considers metric factors `observation` and `access` as predictors. The `interaction` model contains the latter factors' interaction as well.

df	*r*	*95% ci*	*p*
68	0.927	0.885-0.954	< 0.001

Table 2: Pearsons product-moment correlation test between production data and MLE best-fit of the speaker model.

	S	*rho*	*p*
state data	5502.2	0.801	< 0.001
acc,obs data	32101	0.834	< 0.001

Table 3: Spearman's rank correlation test between interpretation data and MLE best-fit of the listener model.

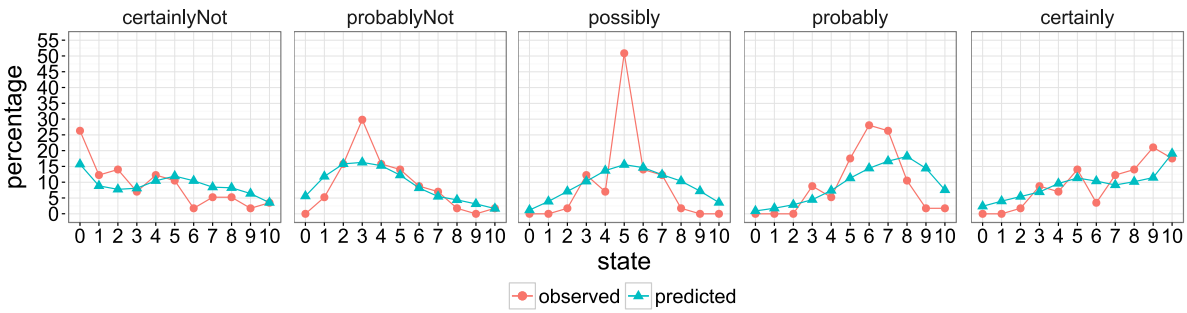

Figure 2: Results of the interpretation study: participants' choices of the number of red balls in the urn for each message, compared with predictions.

References

Beyth-Marom, R. (1982). "How probable is probable? A numerical translation of verbal probability expressions". In: *Journal of forecasting* 1.3, pp. 257–269.

Goodman, Noah D and Andreas Stuhlmüller (2013). "Knowledge and implicature: Modeling language understanding as social cognition". In: *Topics in cognitive science* 5.1, pp. 173–184.

Kratzer, A. (1991). "Modality". In: *Semantics: An international handbook of contemporary research*. Ed. by A. von Stechow and D. Wunderlich. Berlin: de Gruyter, pp. 639–650.

Lassiter, D. (2011). "Measurement and Modality: the scalar basis of modal semantics". PhD thesis. NYU Linguistics.

Moss, S. (2015). "On the semantics and pragmatics of epistemic vocabulary". In: *Semantics and Pragmatics* 8.5, pp. 1–81.

Teigen, K. H. (1988). "When are low-probability events judged to be 'probable'? Effects of outcome-set characteristics on verbal probability estimates". In: *Acta Psychologica* 6.2, pp. 157–174.

Windschitl, P. D. and G. L. Wells (1998). "The alternative-outcomes effect". In: *Journal of Personality and Social Psychology* 75.6, pp. 1411–1423.

Yalcin, S. (2010). "Probability operators". In: *Philosophy Compass* 5.11, pp. 916–37.