Definitely maybe and possibly even probably: efficient communication of higher-order uncertainty

Michele Herbstritt (michele.herbstritt@gmail.com)
Department of Linguistics, University of Tübingen
Wilhelmstrasse 19, 72070 Tübingen, Germany

Michael Franke (mchfranke@gmail.com)
Department of Linguistics, University of Tübingen
Wilhelmstrasse 19, 72070 Tübingen, Germany

Abstract

Possibility and probability expressions, like *possibly* or *probably*, are frequently assumed to communicate that the probability of a proposition is above a certain threshold. Most previous empirical research on these expressions has focused on cases of known objective chance: if the true objective probability is given, would a speaker use *possibly*, *probably* or one of their kin? Here, we investigate the use of probability expressions when speakers have subjective uncertainty about objective chance, i.e., higher-order uncertainty. Experimental data suggest that speakers' choices of a probability expression is a complex function of their state of higher-order uncertainty. We formulate a computational probabilistic model of pragmatic speaker behavior that explains the experimental data.

Keywords: uncertainty; probability; experimental pragmatics; computational modeling

Introduction

Real-world reasoning, decision-making and communication tasks take place almost invariably under uncertainty. From everyday future planning and small talk about the weather to scientific practices and financial or legal reports, we reason, decide and speak about things that we do not know for sure. It is not surprising that possibility and probability expressions are ubiquitous in communication. To mention only a recent example, a quick search within 'The Litvinenko Inquiry', delivered by Sir Robert Owen to the Home Secretary of the United Kingdom on January 19, 2016 (329 pages),[1] resulted in 84 instances of *probable/probably/(un)likely* and 103 instances of *might/possibly/it is possible that*.

Higher-order uncertainty This paper reports experimental and modeling work on the meaning and use of the English expressions *probably (not)* and *possibly*. These expressions (and variations thereof) have been extensively investigated in linguistics (Kratzer, 1991; Egan & Weatherson, 2011; Lassiter, 2011) and psychology (Beyth-Marom, 1982; Brun & Teigen, 1988; Teigen, 1988; Windschitl & Wells, 1996, 1998). What sets our work apart from most of the literature is that we investigate the use of *probably (not)* and *possibly* in situations of what we call "higher-order uncertainty".

To illustrate, imagine an urn containing 10 balls of two different colors (e.g. red and blue). The proportion of red balls out of 10 expresses the objective chance that a random draw will result in a red ball. Knowing the objective chance means knowing the exact content of the urn, and yet it is not enough to be sure about what will happen, unless the objective chance is equal to 0 or 1. This represents the first layer of uncertainty, where having perfect information about how the world is like is not enough to perfectly predict how the world will be.

A second, higher-order, level of uncertainty comes into the picture when the true objective chance is uncertain. Imagine an agent who has only imperfect information about the urn: the agent knows that it contains 10 balls of two different colors, but the agent is only allowed to draw (and look at) a certain number of balls. The agent might observe that none, some or all of the drawn balls are red and form an uncertain belief about the content of the urn. Clearly, the lower the number of drawn balls, the less precise the agent's belief.

This work is about uncertainty of this higher-order kind. In the remainder of this section we summarize some relevant ideas from the linguistic literature on possibility and probability expressions. The following section reports on an experiment that investigates choices of probability expressions under higher-order uncertainty. Standard regression modeling suggests that higher-order uncertainty may affect lexical choices in complex ways. We therefore turn to a computational model of a pragmatic speaker that aims to predict lexical choice in a natural way. We compare model variants and scrutinize the predictions of the most credible variant.

Previous work The amount of previous linguistic research about possibility expressions such as *might*, *possible*, *possibly* (best known as "epistemic modals") is gigantic.[2] Here we refer to the milestone work by Kratzer (1991), which puts forward a uniform analysis of possibility and probability expressions as quantifiers over possible worlds. It is a purely qualitative analysis, with no reference to probability measures.

Much less research has appeared in linguistics specifically about probability expressions. However, Kratzer's picture has been challenged on many grounds by several authors in recent years. Oversimplifying, there seems to be a consensus about the need of a semantics which incorporates probability measures (Yalcin, 2010; Lassiter, 2011; Moss, 2015). The simplest implementation of this semantics postulates that the

[1] Available online at `https://www.litvinenkoinquiry.org`.

[2] See (Egan & Weatherson, 2011) for an introduction to the contemporary debate.

meaning of a sentence of the form *Probably, p* is that the probability of p exceeds a certain threshold θ.

Recent work considers complex nested cases as well (Moss, 2015; Lassiter & Goodman, 2015):

(1) It might be probable that Alice is wearing green.

(2) Alice is definitely likely to be wearing green.

Suppose that Bob is Alice's friend and has witnessed her wearing green 5 times on 8 consecutive days. On the other hand, Carol is Alice's mum and has observed that Alice was wearing green 500 times on 800 consecutive days. Despite the fact that the objective proportion of green observations is the same for both, only Carol is in the position to utter (2), while Bob should limit himself to (1). These examples raise the question —most often left unanswered— of what kind of uncertainty is in play when speakers use possibility and probability expressions. If the objective proportion is not enough to distinguish between (1) and (2), then what is needed?

This is relevant also for simple, non-nested, uses of possibility and probability expressions, which are way more frequent than nested cases. Consider a more extreme version of the urn scenario. There are 100 balls in the urn, of two different colors. Imagine to draw 8 balls and observe that 5 of them are red. Then (3) is a very appropriate thing to say and (4) is not. But imagine to draw 80 balls and observe that 50 of them are red: the proportion of observed red balls is the same, but, intuitively, (4) is more appropriate than before.

(3) A randomly drawn ball might be red.

(4) A randomly drawn ball will probably be red.

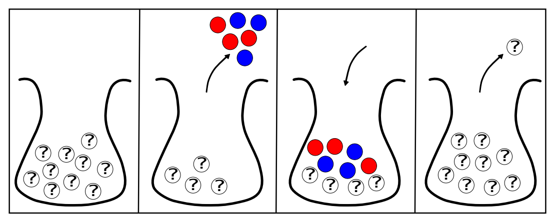
It seems that in the case of uncertainty about objective chance, the maximum likelihood guess about the latter is not all that matters. If 50 out of 80 observed balls are red, we can be much more certain about the objective chance level, than when 5 out of 8 balls are observed. This difference may give rise to complex nested uses of probability expressions, but does it have to? It seems that higher-order uncertainty is frequent in life. If there is an objective chance of rain, we do not know it, but still we do not hear the neighbor say that it will definitely maybe and possibly even probably rain.

There are two options to reconcile higher-order uncertainty with the use of simple probability expressions. One is to assume that an agent who is uncertain about the objective chance of a proposition p, flattens his higher-order belief into a single probabilistic belief about p. Simple probability expressions do not communicate genuine higher-order uncertainty, but simple one-dimensional uncertainty about p. Another possibility is that higher-order uncertainty plays a role in our choice of probability expressions. The question, to be addressed by a computational model below, would then be: how exactly would different layers of uncertainty affect the choice of probability expressions? In any case, since the mapping from higher-order certainty to flattened one-dimensional beliefs is many-to-one, it should be possible to find experimental evidence to decide between these rival options.

Experiment

Participants 50 (self reported) English native speakers with US IP-addresses were recruited via Amazon's Mechanical Turk.

Material The setup of the experiment is the urn scenario introduced above. The urn contains 10 balls of two different colors. Observations of the content of the urn are made by drawing a certain number of balls (referred to as "access") and counting how many of them are red ("observation"). Each trial showed a picture representing the urn scenario with various access and observation. A sample stimulus is shown in Figure 1. A short description that came with every picture reminded participants that they had to imagine drawing a certain number of balls, looking at them, and putting them back in the urn; then, they would draw another ball and make a prediction about its color.



You draw 6 balls and observe that 3 of them are red.

Figure 1: Sample stimulus.

We selected seven proportions of observed red balls, namely $0, 0.25, 0.33, 0.5, 0.67, 0.75, 1$. Each proportion is obtained in two different ways: one encoding what we call "low" uncertainty (access>5) and one what we call "high" uncertainty (access<5). Crossing proportions and uncertainty levels resulted in 14 experimental conditions (Table 1).

Table 1: Experimental conditions.

	0	0.25	0.33	0.5	0.67	0.75	1
high	0/2	1/4	1/3	2/4	2/3	3/4	2/2
low	0/10	2/8	3/9	4/8	6/9	6/8	10/10

The experiment consisted of *expression* and *likelihood trials*. Expression trials probe into the lexical choice of probability expressions. We asked the participants to complete a sentence of the form *The next ball will* [...] *be red* by selecting the most appropriate expression from a drop-down menu containing *certainly not*, *probably not*, *possibly*, *probably*, *certainly*. In likelihod trials participants had to answer the question *How likely do you think it is that a randomly drawn ball will be red?* by adjusting a slider ranging from 0 to 100 with a step of 5. Answers from likelihood trials hint at participants' beliefs about the probability of the crucial proposition *The next draw will be red* in each condition.

Procedure Each participant completed 12 trials, one for each of 12 conditions randomly picked from the 14 total conditions. Half of the trials were expression trials, the other half likelihood trials. Prior to the main experimental phase, participants completed a training phase that contained the following introduction:

> "This experiment is an interactive two player game of chance. The players cooperate to guess the content of an urn. Both players know that the urn always contains 10 balls of different colors (for example, red and blue). But only one player (the *sender*) is allowed to draw a certain number of balls from the urn and look at them. The sender puts the balls back into the urn and gives it a nice shake, then the sender draws a new ball from it. Before looking at it, the sender sends a message to the other player (the *receiver*). The receiver reads the message and tries to guess the exact content of the urn."

The motivation for this was to avoid potential confounds, as much as possible, about the purpose of conversation when choosing between probability expressions. Previous research has established that contextual *questions under discussion*, i.e., ways of classifying what counts as important to a conversation, may affect the use and interpretation of probability expressions (Teigen, 1988; Windschitl & Wells, 1998; Lassiter, 2011; Herbstritt, 2015). Moreover, by introducing the participants to a fictive interactive game context we hoped to suggest that they should reason about the effect of their choices on another agent.

Results & analysis Figure 2 shows results from the expression trials, i.e., the percentages of choices for each message in each condition, grouped by uncertainty level. Figure 3 shows mean answers from the likelihood trials, alongside the posterior beliefs that an ideal Bayesian agent should hold.

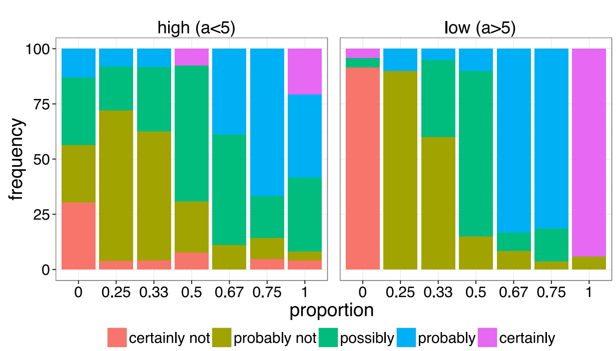


Figure 2: % of message choices.

We want to test whether choice of probability expression is governed (i) only by participants' probabilistic beliefs about the prejacent *The next draw will be red*, or (ii) by the more complex higher-order uncertainty state, expressed as a function of accessed and observed balls. To do so, we formulate and compare two kinds of multinomial logistic regression models. The first and `simple` model seeks to predict the five-level categorical factor `answer` (*certainly not*, *probably not*, *possibly*, *probably*, *certainly*) with a single metric factor `belief` which is the participants' mean answers in the relevant likelihood trials. The second and `complex` model considers metric factors `observation` and `access` as predictors. We also consider an `interaction` model that contains the latter factors' interaction as well.

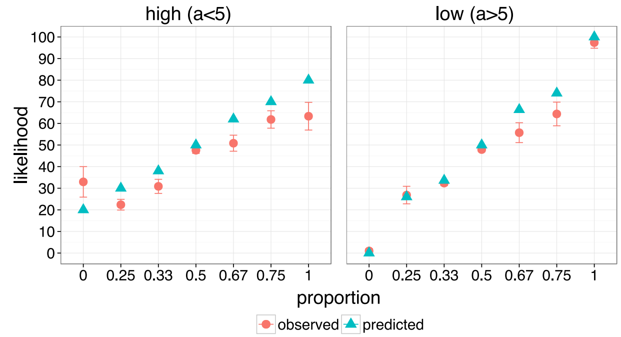


Figure 3: Measured and ideal beliefs.

Table 2: AIC scores.

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

AIC scores of best-fits of these regression models are given in Table 2. Despite added complexity, `complex` seems a better model than `simple`, and `interaction` seems to be the best. This suggests that there might be more to the use of probability expressions than just the level of belief in the prejacent; the result suggests that higher-order uncertainty may affect choice of expressions in more subtle ways.[3]

How do speakers choose expressions under higher-order uncertainty? What exactly is the role of observation and access? Can we predict the speakers' choices by assuming that possibility and probability expressions have a simple semantics but the speakers use them pragmatically? To answer these questions we developed a probabilistic model based on a version of the *Rational Speech Acts* model (henceforth, RSA) proposed by Goodman and Stuhlmüller (2013) (G&S).

Computational model

Probabilistic modeling is arguably better suited for capturing the subtleties of (pragmatic) language use than the standard formal semantics framework. RSA has proven to be a successful modeling tool when it comes to complex usage patterns (or pragmatic effects) on the basis of simple semantics and largely shared insights from rational choice theory and

[3] AIC score of the `simple` model using ideal beliefs instead of measured ones as factor `belief` is equal to 658.98: better than the model using measured beliefs, but still worse than the `complex` model.

information theory.[4] The version of RSA developed by G&S deals with exactly the kind of uncertain beliefs that we are interested in, although for a different purpose.

Basic model The basic setup of the model mirrors the setup of the experiment. The urn contains 10 balls, any number of which can be red. The set of natural numbers $S = \{0, \dots, 10\}$ represents the state space: each $s \in S$ is a possible quantity of red balls in the urn. The set $A = \{0, \dots, 10\}$ contains the possible access values. Given a state s and access value a, the number of red balls observed by the agent is denoted o, and upper-bounded by $\min(s, a)$.

The first level of uncertainty is determined by the state: $s/10$ is the objective chance that a randomly drawn ball will be red. Crucially, the agent does not know this value. Instead, the agent observes o red balls out of a. Based on this observation the agent forms an uncertain belief over the state space, modeled as the Bayes-inverted conditional probability of observing o red balls out of a drawn balls in state s:

$$\text{speaker.bel}(s|o,a) \propto P(o|a,s) * \text{prior}(s) \tag{1}$$

In turn, $P(o|a,s)$ is given by the hypergeometric model of the urn:

$$P(o|a,s) = \text{hypergeometric}(o; a, s, 10) \tag{2}$$

Equation 1 defines the beliefs of a *speaker*, as we aim to model the behavior of message-sending agents. As such, a crucial component of the model is the set of available messages together with their meaning. In the spirit of RSA, we assume a simple literal semantics, expressed as follows in the standard notation of formal semantics:

$[\![\text{certainly}(p)]\!]_s = 1$ iff $P(p) = 1$ in s
$[\![\text{probably}(p)]\!]_s = 1$ iff $P(p) > \theta$ in s
$[\![\text{possibly}(p)]\!]_s = 1$ iff $P(p) > 0$ in s
$[\![\text{probably not}(p)]\!]_s = 1$ iff $P(p) < 1 - \theta$ in s
$[\![\text{certainly not}(p)]\!]_s = 1$ iff $P(p) = 0$ in s

The threshold θ is a free parameter in the model (more about this below). The variable p is to be instantiated with (some sentence equivalent to) *The next ball will be red*.

An important assumption in RSA modeling —loosely based on Paul Grice's Maxim of Quantity, is that the communicative goal of the speaker is to maximize the information transferred to the listener. There are several ways of formalizing the maximization of information. We think of it as bringing the listener's beliefs as close as possible to the speaker's beliefs, i.e. minimizing the (Hellinger) distance between the probability distributions representing those beliefs.[5]

RSA models the behavior of (imperfectly) rational agents. Adopting the terminology of rational choice theory, the speaker's behavior is to soft-maximize the expected utility of the message in the current situation:

$$\text{speaker.prob}(m|o,a) \propto exp(\lambda * \text{EU}(m; o, a)) \tag{3}$$

EU is multiplied by a rationality parameter λ (free in the model) that modulates "how rational" the choice is.[6] EU is defined as negative Hellinger distance (HD) between the speaker's beliefs and the beliefs of a "literal listener":

$$\text{EU}(m; o, a) = -\text{HD}[\text{speaker.bel}(s|o,a), \text{literal.bel}(s|m)] \tag{4}$$

The literal listener is a theoretical construct needed to ground the otherwise infinite process implied in a recursive model of pragmatic reasoning. It is a dummy agent who does not perform any kind of pragmatic reasoning and simply interprets every message m according to the literal semantics:

$$\text{literal.bel}(s|m) = P(s|m \text{ is true}) * \text{prior}(s) \tag{5}$$

Parameters estimation The basic version of the model, M_0 has two free parameters: the threshold θ and the rationality λ. It assumes flat prior beliefs over the state space. We also considered three more complex versions of the model: a model where the prior beliefs are expressed by a symmetric beta-binomial distribution with free shape parameters $\alpha = \beta$ (M_1); a model with two (possibly) different free rationality parameters, λ_{low} and λ_{high}, one for each level of uncertainty (M_2); finally, a model combining these two variations (M_3).

Each model was implemented in JAGS (Plummer, 2003) and the posterior likelihoods of the free parameters given the experimental data were estimated by Bayesian inference via MCMC sampling. We remained uncommitted on the prior distributions over the parameters, except for fixing sensible intervals:

$$\theta \sim \mathcal{U}(0,1) \quad \lambda, \lambda_{\text{low}}, \lambda_{\text{high}} \sim \mathcal{U}(0,20) \quad \alpha \sim \mathcal{U}(0,20)$$

We gathered two chains of 5000 samples for each model after an initial burn-in phase of 5000. Convergence was confirmed via $\hat{R}$ (Gelman & Rubin, 1992). The results were interesting for at least three reasons.

First, the mean value inferred for θ was always equal to 0.55, regardless which model we simulated. This is an important result: our data provide evidence that an objective chance bigger than 55% is enough to consider *probably* as an appropriate expression. This is in line with intuition and previous experimental results. It speaks in favor of the model that data-driven inference recovers this intuitive value for θ without stipulating it from the start.

Second, the estimation of α in M_1 resulted in a mean value of 2.81 (HDI: 1.26-4.78).[7] Notice that $\alpha = 1$ would imply

[4]See recent works by Degen, Tessler, and Goodman (2015), Kao and Goodman (2015) and Hawkins, Stuhlmüller, Degen, and Goodman (2015) among others.

[5]We depart from the informativity criterion adopted by G&S because it does not allow the speaker to send messages that are literally false; however, we want to allow some pragmatic slack to the speaker: it is plausible to think that, for example, chances around 97% or bigger are certain enough for us to say *certainly*, even if this expression is literally true only when the odds are 100%.

[6]As $\lambda \to \infty$, the choice approaches perfect rationality.

[7]All HDIs (*highest density intervals*) reported here comprise 95% of the posterior density, such that no point outside the interval has higher density than any point within.

flat prior beliefs over the state space, which can thus reasonably be excluded given our data (more about this below). This can be perhaps explained by looking again at the introductory text of the experiment: it is written that "[...] the urn always contains 10 balls of different colors" and it is reasonable to assume that this might have caused a number of the participants to neglect the possibility that the urn contained 0 or 10 red balls (which results in non uniform prior beliefs).

Third, the estimation of λ_{low} and λ_{high} in M_2 resulted respectively in mean values of 7.36 (HDI: 6.03-8.79) and 3.49 (HDI: 2.75-4.24). The fact that the difference between λ_{low} and λ_{high} is different from zero with completely non-overlapping HDIs suggests that our data provide evidence for assuming different rationality parameters for different uncertainty level (more about this below). An intuitive explanation is that the more precise the participants' belief, the easier it was for them to behave more rationally.

Finally, the mean values for the free parameters of M_3 and their HDIs are reported in Table 3.

Table 3: Estimated parameters of M_3.

	θ	α	λ_{low}	λ_{high}
mean	0.55	2.81	6.11	3.16
HDI	0.50-0.59	1.04-5.42	4.75-7.45	2.39-3.95

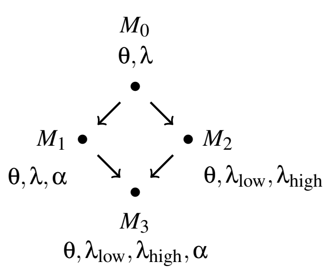


Figure 4: Nesting relations between models.

Model comparison Our four models are nested. M_0 is a special case of both M_1 and M_2, which are special cases of M_3, as depicted in Figure 4. This allows for model comparisons based on Bayes factors (BF) with the Savage-Dickey method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). As expected, both M_1 and M_2 are more credible models given our data. The Bayes factor in favor of M_1 against M_0 is substantial (approximately $BF = 8.68$). More strikingly, the Bayes factor in favor of M_0 against M_2 is so low that it cannot be computed with normal float precision levels, which in turn means that M_2 is definitely more credible than M_0. Moving to M_3, we get a similarly striking result when we compare it to M_1: the Bayes factor in favor of M_1 approaches zero. On the other hand, the Bayes factor in favor of M_3 against M_2 is only equal to 1.40 which is barely enough to distinguish the two models on the basis of the data. Summing up, allowing our model to use two different rationality parameters for the two different uncertainty levels invariably results in a definitely more credible model.

Correlation and model criticism Model comparison based on Bayes factors tells us if/how much a model is better than another, but does not give us an absolute measure of how good a model is. Having picked M_3 as the best model at our disposal, we correlated the experimental observations with the predictions made by the model fitted with the maximum-likelihood estimated values of its free parameters. The results of Pearson's product-moment correlation provide us with a measure of goodness of fit for the model as its best:

$$df = 68;\ r = 0.927;\ 95\%\ \text{ci} : 0.885\text{-}0.954;\ p < 0.001$$

By that (frequently used) measure, predictions of the model look quite good.

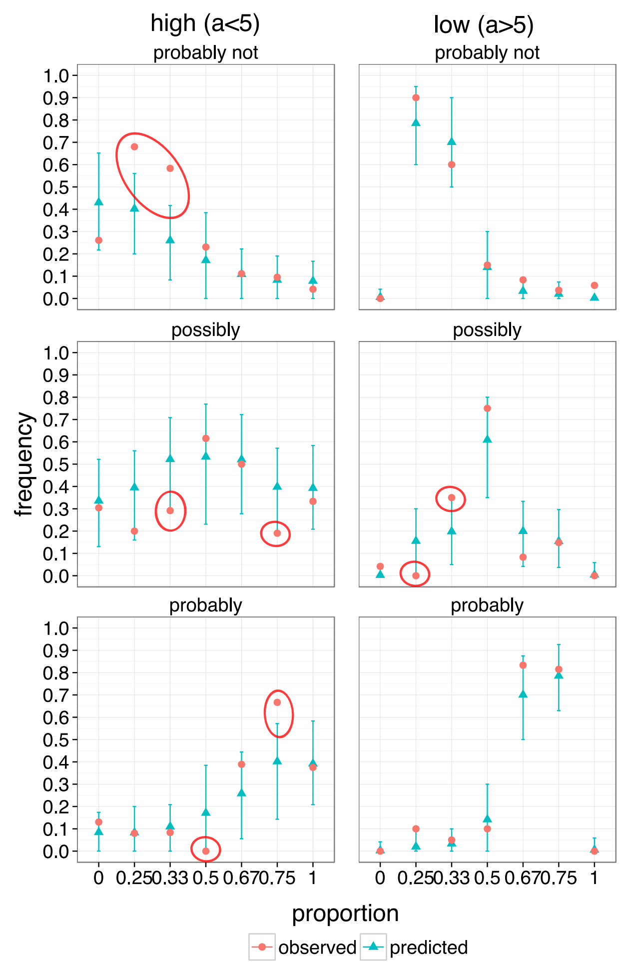


Figure 5: Posterior Predictive Check.

A more stringent criterion for goodness of fit are Posterior Predictive Checks (PPC) (Kruschke, 2014). PPCs look at hypothetical data generated from a model when parameters are randomly drawn from their posterior distribution (i.e., conditional on the data). By visual means, we then check whether any actually observed data point is "surprising" in the sense that it lies outside of the 95% HDI of the hypothetically generated data. Figure 5 shows the mean frequencies of posterior predictive message choices of M_3 (in light blue), together with their 95% HDIs. The pink dots represent observed data.

Ideally, all observations should lie within the posterior HDIs. Visual inspection of the plots allows us to recognize several critical points where the posterior predictions of the model diverge from the observed data (circled in red). We observe that most critical points are found in the high uncertainty condition, plausibly where most noise occurred in the data. It is also exactly at these points where participants' mean answers in the likelihood trials diverged from idealized Bayesian belief update (see Figure 3). This suggests that failures in PPCs might not be the fault of the pragmatic part of the model, but the belief update part of the model.

Conclusion

We presented experimental evidence suggesting that the speakers' use of possibility and probability expressions depends not only on the objective chance of the event in question but also on the speakers' state of higher-order uncertainty. We formulated a computational model, based on a conservative extension of RSA, that explains the experimental data in terms of simple semantics and standard pragmatic reasoning.

Despite its simplicity, the predictions of the model are quite good. Comparing AIC scores for the best-fits of the `interaction` regression model and our theory-driven computational model reveals a striking preference for the latter (AIC: 635.93 vs 270.90).

A possible line of future empirical research will be to investigate speakers' use of nested possibility and probability expressions under higher-order uncertainty: if given the choice, would speakers prefer a nested construction over a simple one to communicate their uncertain beliefs? Moreover, a natural continuation of the work presented here will be to investigate listeners' interpretation of (simple and nested) possibility and probability expressions.

Acknowledgments

The authors would like to thank Judith Degen, Margherita Isella and Anthea Schöller. Financial support by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63) is gratefully acknowledged.

References

Beyth-Marom, R. (1982). How probable is probable? a numerical translation of verbal probability expressions. *Journal of forecasting*, *1*(3), 257–269.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, *41*(3), 390–404.

Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. In D. C. Noelle et al. (Eds.), *Proceedings of CogSci37.*

Egan, A., & Weatherson, B. (2011). *Epistemic modality*. Oxford University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, *7*, 457–472.

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

Hawkins, R. X. D., Stuhlmüller, A., Degen, J., & Goodman, N. D. (2015). Why do you ask? good questions provoke informative answers. In D. C. Noelle et al. (Eds.), *Proceedings of CogSci37.*

Herbstritt, M. (2015). Experimental investigations of probability expressions: a first step in the (probably) right direction. In M. Kaeshammer & P. Schulz (Eds.), *Proceedings of ESSLLI 2015 student session.*

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In D. C. Noelle et al. (Eds.), *Proceedings of CogSci37.*

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

Kruschke, J. (2014). *Doing bayesian data analysis, 2nd edition: A tutorial with r, jags, and stan*. Academic Press.

Lassiter, D. (2011). *Measurement and modality: the scalar basis of modal semantics*. Unpublished doctoral dissertation, NYU Linguistics.

Lassiter, D., & Goodman, N. D. (2015). *Nested and informative epistemics in a graphical models framework.* Talk given at Bridging Logical and Probabilistic Approaches to Language and Cognition, ESSLLI 2015, Barcelona.

Moss, S. (2015). On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, *8*(5), 1–81.

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of DSC3* (Vol. 124, p. 125).

Teigen, K. H. (1988). When are low-probability events judged to be 'probable'? effects of outcome-set characteristics on verbal probability estimates. *Acta Psychologica*, *6*(2), 157–174.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive psychology*, *60*(3), 158–189.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*(4), 343.

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology*, *75*(6), 1411–1423.

Yalcin, S. (2010). Probability operators. *Philosophy Compass*, *5*(11), 916–37.