UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Wonky worlds: Listeners revise world knowledge when utterances are odd

Permalink

https://escholarship.org/uc/item/9wn4w9zk

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors

Degen, Judith
Tessler, Michael Henry
Goodman, Noah D

Publication Date

2015

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Wonky worlds: Listeners revise world knowledge when utterances are odd

Judith Degen, Michael Henry Tessler, Noah D. Goodman
{jdegen,mhtessler,ngoodman}@stanford.edu
Department of Psychology, 450 Serra Mall
Stanford, CA 94305 USA

Abstract

World knowledge enters into pragmatic utterance interpretation in complex ways, and may be defeasible in light of speakers' utterances. Yet there is to date a surprising lack of systematic investigation into the role of world knowledge in pragmatic inference. In this paper, we show that a state-of-the-art model of pragmatic interpretation greatly overestimates the influence of world knowledge on the interpretation of utterances like *Some of the marbles sank*. We extend the model to capture the idea that the listener is uncertain about the background knowledge the speaker is bringing to the conversation. This extension greatly improves model predictions of listeners' interpretation and also makes good qualitative predictions about listeners' judgments of how 'normal' the world is in light of a speaker's statement. Theoretical and methodological implications are discussed.

Keywords: scalar implicature; world knowledge; prior beliefs; experimental pragmatics; computational pragmatics

How often do you think marbles would sink in water? Probably extremely often, if not always. Now imagine reading *Max threw fifteen marbles in the water. Some of the marbles sank.* Have you begun to reconsider your assumptions? Perhaps you now suspect that these marbles are in fact made of hollow plastic or the water is covered with thick algae? That is, maybe these are not just normal marbles in normal water. Here we explore how prior world knowledge enters into pragmatic utterance interpretation, and when this world knowledge is defeasible: some utterances lead listeners to conclude that the world under discussion is abnormal and has appropriately different prior probabilities. We refer to such an abnormal world as a *wonky* world.

The Rational Speech Acts framework (RSA) (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013), and related models (Franke, 2011; Russell, 2012), treat communication as a signaling game (Lewis, 1969) between a speaker and a listener. The listener reasons by Bayesian inference about what the world is like given that a speaker who produced the utterance is trying to be informative (with respect to a naïve listener). Variants of these models have successfully captured listeners' quantitative behavior on a number of pragmatic inference tasks, including ad hoc Quantity implicature (Degen, Franke, & Jäger, 2013), markedness implicature (Bergen, Goodman, & Levy, 2012), scalar implicature (Goodman & Stuhlmüller, 2013), and non-literal language (Kao, Wu, Bergen, & Goodman, 2014). A defining feature of Bayesian reasoning is that prior beliefs affect inferences that will be drawn. Bayesian models of language interpretation, accordingly, predict that prior beliefs about the world should affect the listener's interpretation of an utterance. While this impact of prior knowledge has been noted, and included in models, it hasn't been systematically studied.

Generalizing our opening example, consider *Some of the X sank*, where *X* is a plural noun such as *marbles*, *feathers*, or *balloons*, and *the X* refers to a contextually established group of objects from category X. When the prior probability, θ_X, of an X[1] sinking is not extreme (e.g., a feather sinking), RSA leads to the standard scalar implicature: the posterior probability that all of the X sank, after hearing the utterance, is much lower than its prior probability (i.e., *Some of the feathers sank* yields that not all of them did). This is because a rational speaker would have been expected to produce the more informative *All of the X sank*, had it been true. As we will show below, RSA makes two strong predictions about the effect of the prior: (1) As θ_X approaches 1, the interpretation probability that all X sank approaches 1, that is, the scalar implicature disappears. This prediction follows because the extreme prior overwhelms the effect of the utterance. (2) For moderate to high prior probability (roughly $0.5 < \theta_X < 1$) and a large total number of objects (more than about 10), the posterior expectation of the number of X that sank should be approximately the same as the prior expectation—that is, the utterance shouldn't affect the expected number of X that sank. This prediction follows from the weak semantics of *some* and the isolated effect of the alternative *all*: *Some of the X sank* only restricts the interpretation (i.e., the number of marbles that sank) to be greater than zero; competition with *All of the X sank* results in the scalar implicature that can at most rule out the state in which all of the X sank. Thus, a sufficiently strong prior will dominate the inference about exactly how many X sank.

However, intuition is at odds with these predictions: as Geurts (2010) has observed, for events with very high prior probability of occurrence (e.g., marbles sinking), an utterance like *Some of the marbles sank* seems to yield strong implicatures; thus, contrary to RSA predictions, the subjective probability that all of the marbles sank is intuitively close to 0.

In Exp. 1 we collect prior probabilities for a variety of events (e.g., sinking) and categories (e.g., marbles). In Exps. 2a and 2b we collect corresponding posterior interpretations after observing utterances containing quantifiers. These experimental results confirm the intuition of relatively strong implicature—hence prediction (1) of RSA is incorrect—and show that the prior has a muted effect on posterior expectation—hence prediction (2) of RSA is incorrect. Given the previous success of RSA models, this constitutes a striking puzzle. To address this puzzle we pursue the intuition raised at the very beginning of this paper: that sometimes, the speaker's utterance will lead the listener to infer that the world under discussion is wonky and she should therefore use less extreme prior beliefs in the computation of speaker meaning. We introduce a variant of RSA, *wonky RSA (wRSA)*, in which the listener can revise her beliefs about the domain under discussion. We show that this extension resolves the puzzle of the prior's muted effects. In Exp. 3 we explore participants' intuitions about whether the world is normal or wonky in the scenarios of Exps. 2a and 2b and find that wRSA predicts listeners' wonkiness judgments.

---

[1] We will use 'X' interchangeably to refer to both the category and the members of the category.

Experiment 1: prior elicitation

Exp. 1 measured listeners' prior beliefs about how many objects exhibit a certain effect (e.g., how many marbles sink).

Method

Participants, procedure and materials We recruited 60 participants over Amazon's Mechanical Turk.

On each trial,[2] participants read a one-sentence description of an event like *John threw 15 marbles into a pool.* They were then asked to provide a judgment of an effect, e.g. *How many of the marbles do you think sank?*, on a sliding scale from 0 to 15. Each item had a similar form: the first sentence introduced the objects at issue (e.g., marbles). The question always had the form *How many of the X Yed?*, where *X* was the head of the direct object noun phrase introduced in the first sentence (e.g., *marbles*, *cups*, *balloons*) and *Yed* was a verb phrase denoting an effect that the objects underwent (e.g., *sank*, *broke*, *stuck to the wall*). Each verb phrase occurred with three different objects, e.g., *sank* occurred with *marbles*, *cups*, and *balloons*. Items were constructed to intuitively cover the range of probabilities as much as possible, while also somewhat oversampling the upper range of probabilities to have more fine-grained coverage of this region that is of most interest for testing the RSA model. Judgments were obtained for 90 items, of which each participant saw a random selection of 30 items.

Results

Data from one participant, who gave only one response throughout the experiment, were excluded. Each item received between 12 and 29 ratings. Distributions of ratings for each item were smoothed using nonparametric density estimation for ordinal categorical variables (Li & Racine, 2003) with the `np` package in R (Hayfield & Racine, 2008). As intended, items covered a wide range of probabilities. See Figure 1 for a histogram of expected values of each smoothed prior distribution.

In the next section, we use these empirically obtained smoothed prior beliefs to derive RSA predictions for the interpretation of utterances like *Some of the marbles sank*, before empirically measuring participants' interpretations.

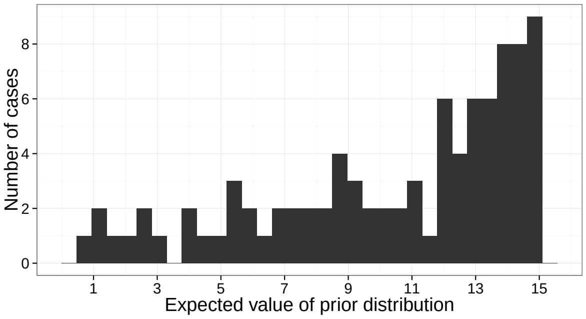

Figure 1: Histogram of expected values of each empirically elicited and smoothed prior distribution.

Effect of the world prior in RSA

The basic Rational Speech Acts model defines a pragmatic listener $P_{L_1}(s|u)$, who reasons about a speaker $P_{S_1}(u|s)$, who in turn reasons about a literal listener $P_{L_0}(s|u)$. Each listener does Bayesian inference about the world state, given either the literal truth of utterance u or the speaker's choice of u; the speaker is a softmax-optimal decision maker, with the goal of being informative about the state s. RSA is defined by:

$$P_{L_0}(s|u) \propto \delta_{[\![u]\!](s)} \cdot P(s) \tag{1}$$

$$P_{S_1}(u|s) \propto \exp(\lambda \ln P_{L_0}(s|u)) \tag{2}$$

$$P_{L_1}(s|u) \propto P_{S_1}(u|s) \cdot P(s) \tag{3}$$

Here $[\![u]\!] : S \to \text{Boolean}$ is a truth-function specifying the literal meaning of each utterance.

For concreteness, assume the set of states of the world $S = \{s_0, s_1, s_2, \ldots, s_{15}\}$, where the subscript indicates the number of objects (e.g., marbles) that exhibit an effect (e.g., sinking). Further assume that the set of utterances *All/None/Some of the marbles sank* is denoted $U = \{u_{\text{all}}, u_{\text{none}}, u_{\text{some}}\}$ and each has its usual literal meaning: $[\![u_{\text{none}}]\!] = \{s_i | i = 0\}$, $[\![u_{\text{some}}]\!] = \{s_i | i > 0\}$, $[\![u_{\text{all}}]\!] = \{s_i | i = 15\}$.

In Figure 2 we show the predictions of RSA (dark blue dots) for the items from Exp. 1 in two different ways: the left panel shows the posterior expected number of affected objects as a function of the prior expectation; the right panel shows the posterior probability of the state in which all objects are affected, as a function of the prior probability of that state.[3] We see that the prior has a strong effect, which can be summarized by the two predictions described in the Introduction: (1) $P(s_{15}|u_{\text{some}}) \to 1$ as $P(s_{15}) \to 1$. (2) $\mathbb{E}[P(s|u_{\text{some}})] \simeq \mathbb{E}[P(s)]$ over the upper half of its range. We next turn to an empirical test of these predictions, or rather, of the intuition that they may be incorrect.

---

[2] See this experiment at `http://cocolab.stanford.edu/cogsci2015/wonky/prior/sinking-marbles-prior.html`

[3] That the individual model predictions look somewhat noisy is due to the different shapes of the prior distributions, such that for the same expected value of the distribution, the distribution itself can take different shapes, which will be treated slightly differently by the model.

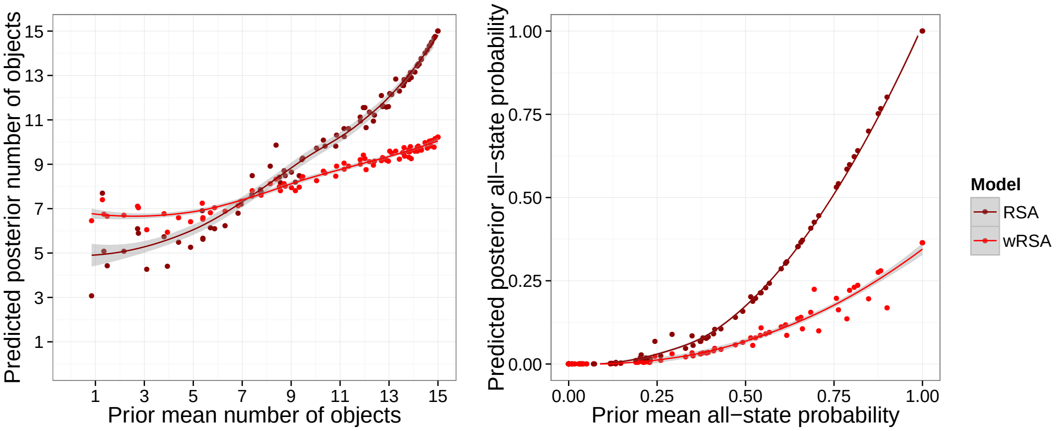

Figure 2: For each item, RSA and wRSA model predicted number of objects ($\mathbb{E}[P(s|u_{\text{some}})]$ as a function of $\mathbb{E}[P(s)]$, left) and model predicted all-state probability ($P(s_{15}|u_{\text{some}})$ as a function of $P(s_{15})$, right) after observing *Some of the X Yed*.

Experiment 2a and 2b: comprehension

Exps. 2a and 2b[4] measured participants' posterior beliefs $P(s|u)$ about how many objects exhibited a certain effect (e.g., marbles sinking), after observing an utterance. The only difference between the experiments was the dependent measure. We used different dependent measures for two reasons. First, this allowed for directly and independently estimating the two values that the predictions above are concerned with: $\mathbb{E}[P(s|u_{\text{some}})]$ and $P(s_{15}|u_{\text{some}})$. Second, this allows for evaluating the generalizability of the empirical result, a growing concern in experimental pragmatics.[5]

Method

Participants, procedure and materials For each experiment we recruited 120 participants over Mechanical Turk.

Participants read the same descriptions as in Exp. 1. They additionally saw an utterance produced by a knowledgeable speaker about the event, e.g. *John, who observed what happened, said: "Some of the marbles sank"*. In Exp. 2a (just as in Exp. 1), they then provided a judgment of an effect, e.g. *How many of the marbles do you think sank?*, on a sliding scale from 0 to 15. In Exp. 2b they instead rated on sliding scales with endpoints labeled "definitely not" and "definitely", how likely they thought 0%, 1-50%, 51-99%, or 100% of the objects exhibited the effect.

Each participant saw 10 *some* trials and 20 filler trials, of which 10 contained the quantifiers *all* or *none*, and the rest were utterances that did not address the number of objects that displayed the effect. These 10 additional fillers were intended to establish whether participants were using information about the prior in the first place. Of these, half were generic short fillers that were intended to communicate the prior, e.g., *Typical*. The rest were longer sentences that addressed a different aspect of the described scenario, e.g. *What a stupid thing to do*. The utterances were randomly paired with 30 random items for each participant.

Results and discussion

Data from eight participants in Exp. 2b were excluded from the analysis because these participants assigned less than .8 probability to the interpretation corresponding to the correct literal interpretation on literal *all* and *none* trials.[6]

The main question of interest was whether participants' judgments of how many objects exhibited the effect after hearing an utterance with *some* followed the predictions of the basic RSA model laid out in the previous section. Mean empirical $\mathbb{E}[P(s|u)]$ and $P(s_{15}|u)$ are shown in Figure 3 for each item. There was a small effect of the prior. For utterances of *Some of the X Yed*, the mean number of objects judged to exhibit the effect increased with increasing expectation of the prior distribution (β=.18, *SE*=.02, *t*=7.4, $p<.0001$). Similarly, the probability of all 15 objects exhibiting the effect increased with increasing prior probability of doing so (β=.06, *SE*=.01, *t*=5.0, $p<.0001$). However, the size of these effects is astronomically smaller than predicted by RSA (for comparison, see dark lines in Figure 2).

One possible explanation for this highly attenuated effect of the prior is that participants simply do not bring world knowledge to bear on the interpretation of utterances. However, this possibility is ruled out by examining participants' performance in the filler conditions: in both Exps. 2a and 2b, the filler conditions closely tracked the prior (see Figure 3).

---

[4] See these experiments at http://cocolab.stanford.edu/cogsci2015/wonky/expectation/sinking-marbles.html and http://cocolab.stanford.edu/cogsci2015/wonky/stateprobs/sinking-marbles-nullutterance.html

[5] For a discussion of the role of dependent measures in experimental pragmatics, see e.g., Benz and Gotzner (2014); Degen and Goodman (2014).

[6] In general, this task yielded noisier results than the task in Exp. 2a (as can be seen in the average lower probability of the all-state after observing *all*, in the right panel of Figure 3) because participants used the sliders in different ways. For example, for cases where intuitively, the all-state was true, some participants assigned non-zero probability to only the all-state, while others were reluctant to do so and always assigned some probability to the 51-99% state.

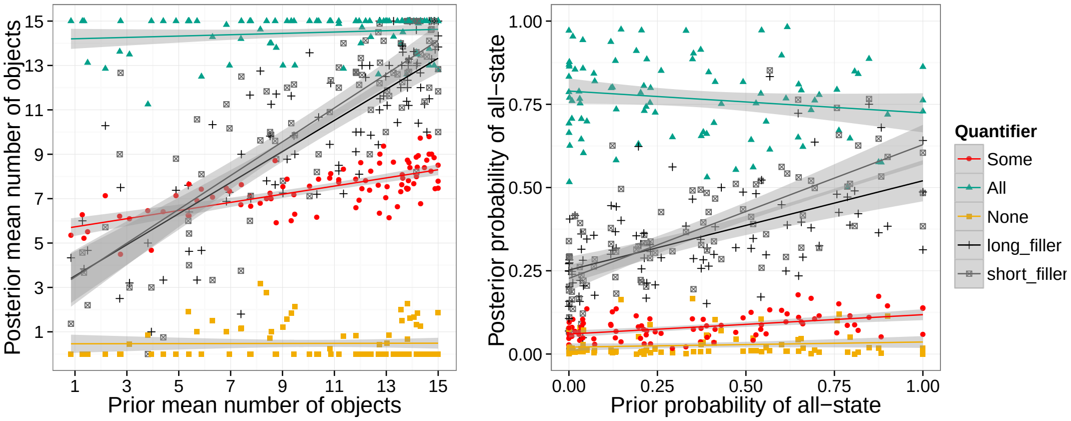

Figure 3: For each item and quantifier, (left) empirical mean number of objects as a function of the prior mean number of objects (i.e., $\mathbb{E}[P(s|u_{\text{some}})]$ vs. $\mathbb{E}[P(s)]$ from Exp. 2a and Exp. 1); and (right) empirical mean all-state probability as a function of the prior mean all-state probability (i.e., $P(s_{15}|u_{\text{some}})$ vs. $P(s_{15})$ from Exp. 2b and Exp. 1). Implicatures from *some* to *not all* were generally strong, as evidenced in the low all-state probabilities after observing *some*.

Exps. 2a and 2b demonstrate that there is an effect of listeners' prior beliefs on the interpretation of utterances with *some*. However, this effect is quantitatively much smaller than predicted by RSA, and qualitatively does not match the predictions identified above: the implicature is not canceled for extreme priors and the posterior expectation diverges from the prior expectation. In the next section, we extend the RSA model to formalize a listener who may decide that her initial beliefs about the domain are not shared by the speaker and responds by revising her priors.

Effect of the world prior in 'wonky RSA'

To capture the idea that the pragmatic listener is unsure what background knowledge the speaker is bringing to the conversation, we extend the basic RSA model by using a "lifted variable" (Goodman & Lassiter, 2014; Lassiter & Goodman, 2013; Bergen et al., 2012; Kao et al., 2014) corresponding to the choice of state prior. That is, we posit that the prior, now $P(s|w)$, depends on a "wonkiness" variable w, which determines if it is the "usual" prior for this domain or a more generic back-off prior that we take to be uniform:

$$P(s|w) \propto \begin{cases} 1 & \text{if } w \\ P_{\text{usual}}(s) & \text{if not } w \end{cases}$$

This inferred prior is used in both the literal and pragmatic listeners, indicating that it is taken to be common ground. However, the w variable is reasoned about only by the pragmatic listener, which captures the idea that it is an inference the pragmatic listener makes about which assumptions are appropriate to the conversation. Using the notation of the earlier modeling section:

$$P_{L_0}(s|u,w) \propto [\![u]\!](s) \cdot P(s|w) \tag{4}$$
$$P_{S_1}(u|s,w) \propto \exp(\lambda \ln P_{L_0}(s|u,w)) \tag{5}$$
$$P_{L_1}(s,w|u) \propto P_{S_1}(u|s,w) \cdot P(s|w) \cdot P(w) \tag{6}$$

We refer to this model as wRSA. Notice that the choice of w that the listener makes will depend on $P_{S_1}(u|s,w)$: if a given utterance can't be explained by the usual prior because it is unlikely under any plausible world state s, the pragmatic listener infers that the world is wonky and backs off to the uniform prior. That is, if the utterance is odd, the listener will revise her opinion about appropriate world knowledge.

To make predictions for Exp. 2 from wRSA we use the smoothed empirical priors from Exp. 1 as $P_{\text{usual}}(s)$ for each item. The wonkiness prior $P(w)$ and the speaker optimality λ are fit to optimize mean squared error (MSE) with Exp. 2 data. The optimal parameters ($\lambda = 2$, $P(w) = 0.5$) resulted in an MSE of 2.15 (compared to 14.53 for RSA) for the expected number of objects, and 0.01 (compared to 0.07 for RSA) for the all-state probability. The better fit of wRSA compared to RSA can be seen in the comparison of Figure 2 and Figure 3: in both cases, wRSA (light blue lines) predicts a much attenuated effect of the prior compared to regular RSA (dark blue lines), in line with the empirical data. Furthermore, wRSA does not make either of the problematic predictions identified earlier for regular RSA.

These results are encouraging: wRSA is able to account for the qualitative and quantitative departures of participants' behavior from RSA, with respect to the effect of the prior. Is this because listeners are actually inferring that the world is unusual from an utterance like *Some of the marbles sank*? The wRSA model makes predictions about the probability that a given world is wonky after observing an utterance; see Figure 4 for predicted wonkiness probabilities for u_{all}, u_{none}, and u_{some} using the optimal $P(w)$ and λ parameters from fitting wRSA to the Exp. 2 data. Note the U-shaped curve, in which the world is judged wonky if u_{some} is used in worlds with extreme priors. We can test these predictions directly by simply asking participants whether the situation is normal.

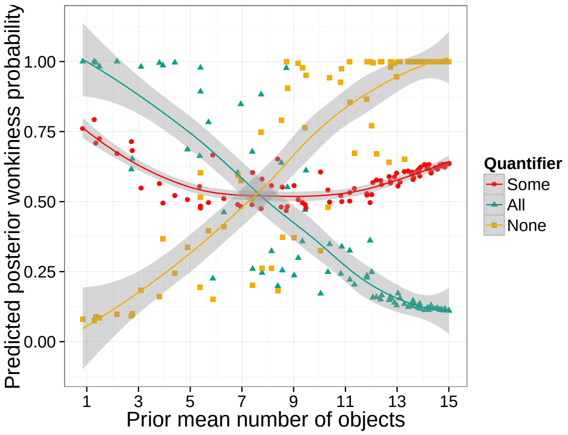

Figure 4: For each item, predicted wonkiness probability after observing an utterance ($u_{\text{all}}, u_{\text{none}}, u_{\text{some}}$), as a function of the prior expected number of affected objects.

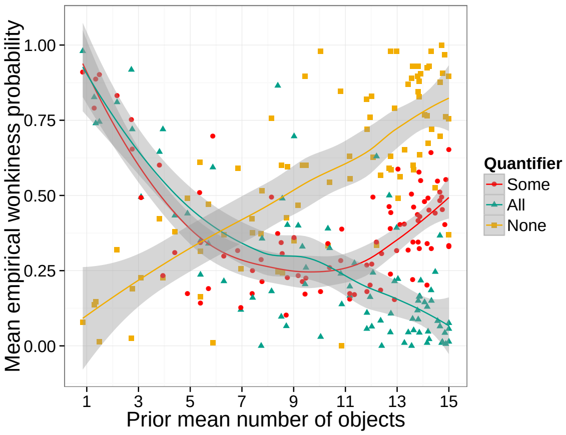

Figure 5: For each item, mean empirical wonkiness probability after observing an utterance ($u_{\text{all}}, u_{\text{none}}, u_{\text{some}}$), as a function of the prior expected number of affected objects.

Experiment 3: wonkiness

Exp. 3[7] measured participants' beliefs in world wonkiness after observing the scenarios and utterances from Exps. 2.

Participants, procedure and materials We recruited 60 participants over Mechanical Turk.

The procedure and materials were identical to those of Exps. 2a and 2b, with the exception of the dependent measure. Rather than providing estimates of what they believed the world was like, participants were asked to indicate how likely it was that the objects (e.g., the marbles) involved in the scenario were normal objects, by adjusting a slider that ranged from *definitely not normal* to *definitely normal*.

Results The extreme ends of the sliders were coded as 1 (*definitely not normal*, i.e., wonky) and 0 (*definitely normal*, i.e., not wonky). We interpret the slider values as probability of world wonkiness. Mean wonkiness probability ratings are shown in Figure 5 and closely mimic wRSA's predictions (see Figure 4). For u_{all} and u_{none}, increasing prior expectation of objects exhibiting the effect resulted in a fairly linear decrease and increase in the probability of wonkiness, respectively. For u_{some}, the pattern is somewhat more intricate: probability of wonkiness initially decreases sharply, but rises again in the upper range of the prior expected value.

Qualitatively, the model captures both the linear increase and decrease in wonkiness probability for u_{all} and u_{none}, respectively. Importantly, it also captures the asymmetric U-shaped wonkiness probability curve displayed by u_{some}. Intuitively, this shape can be explained as follows: for very low probability events, it is surprising to learn that such an event took place (which is what is communicated by u_{some}), so wonkiness is high. For medium probability events, learning that this event took place is not very surprising, so wonkiness is relatively low. For high probability events, u_{some} may be literally true, but it is not useful in the sense of providing the listener new information. For comparison to the comprehension data fit, the model's MSE for empirical wonkiness probability predictions, using the best parameters from fitting the model to the comprehension data, was 0.07.

Discussion and conclusion

We have shown that listeners' world knowledge, in the form of prior beliefs, enters into the computation of speaker meaning in a systematic but subtle way. The effect of the prior on interpretation was much smaller, and qualitatively different, than predicted by a standard Bayesian model of quantifier interpretation (RSA). This suggests that in certain situations, listeners revise their assumptions about relevant priors as part of the computation of speaker meaning. Indeed, in the cases where the largest deviations from RSA obtained, participants also judged the world to be unusual. Extending RSA with a lifted wonkiness variable that captures precisely whether listeners think the world is unusual, and allows them to back off to a uniform prior (i.e., ignore entirely their previously held beliefs about the world), provided a good fit to the empirical wonkiness judgments and dramatically improved the fit to participants' comprehension data. This model constitutes the first attempt to explicitly model the quantitative effect of world knowledge and its defeasibility on pragmatic utterance interpretation and raises many interesting questions.

In one sense the revision of beliefs in the wRSA listener is standard Bayesian belief updating with respect to a complex prior; however it is not the simple belief update of a flat or hierarchical prior, because the different aspects of prior belief (i.e. $P(w)$ and $P(s|w)$) interact in complex ways with the listener's assumptions about the speaker. As a result, an odd utterance can lead the listener to update their own view of w; this in turn impacts both their own prior over states and what prior they believe the speaker believes they are using—an odd utterance leads the listener to re-evaluate common ground.

---

[7] See the experiment at `http://cocolab.stanford.edu/cogsci2015/wonky/wonkiness/sinking-marbles-normal.html`

This is reminiscent of linguistic theories of presupposition accommodation (Lewis, 1979; Stalnaker, 1973, 1998). It will be interesting to further explore the relation of the wRSA approach to presupposition.

Throughout this paper we discussed wonkiness as an attribute of the *world*, yet empirically we elicited wonkiness judgments about *objects* involved in events. This raises the question of what exactly listeners are revising their prior beliefs about: objects, events, the speaker's beliefs, or the way the speaker uses language? Relatedly, we used a uniform prior distribution as the alternative prior when the listener believes the world is wonky. One could imagine various more flexible alternatives. For instance, listeners may make minimal adjustments to their prior knowledge, or alternatively, may prefer extreme priors that rationalize the utterance once they have discounted the usual priors. Future research should investigate the options listeners consider when their world knowledge must be revised to accommodate an utterance.

This work also has methodological implications: researchers working in the field of experimental semantics and pragmatics would be well served to take into account the effect of 'odd' items, prior beliefs, and interactions between the two.[8] In particular, if the attempt to design uniform stimuli across conditions yields odd utterances in some conditions, we predict that participants will respond by revising their prior beliefs in ways that can be unpredictable. That is, we expect unpredictable interaction effects between stimuli and conditions. This is likely to inflate or compress potential effects of an experimental manipulation.

Concluding, this work exemplifies the importance and utility of exploring the detailed quantitative predictions of formal models of language understanding. Exploring the prior knowledge effects predicted by RSA led us to understand better the influence of world knowledge and its defeasibility on pragmatic interpretation. Listeners have many resources open to them when confronted with an odd utterance, and reconstruing the situation appears to be a favorite.

Acknowledgments

This work was supported by ONR grant N00014-13-1-0788 and a James S. McDonnell Foundation Scholar Award to NDG, an SNF Early Postdoc.Mobility Award to JD, and an NSF Graduate Fellowship to MHT.

References

Benz, A., & Gotzner, N. (2014). Embedded implicatures revisited: Issues with the Truth-Value Judgment Paradigm. In J. Degen, M. Franke, & N. D. Goodman (Eds.), *Proceedings of the formal & experimental pragmatics workshop, european summer school for language, logic and information (esslli)* (pp. 1–6). Tübingen.

Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. In *Proceedings of the thirty-fourth annual conference of the cognitive science society.*

Degen, J., Franke, M., & Jäger, G. (2013). Cost-Based Pragmatic Inference about Referential Expressions. In *Proceedings of the 35th annual conference of the cognitive science society.*

Degen, J., & Goodman, N. D. (2014). Lost your marbles? The puzzle of dependent measures in experimental pragmatics. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society.*

Fine, A. B., Jaeger, T. F., Farmer, T. F., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10).

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

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*(1), 1–82.

Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge Univ Press.

Goodman, N. D., & Lassiter, D. (2014). Probabilistic Semantics and Pragmatics: Uncertainty in Language and Thought. *Handbook of Contemporary Semantics*.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–84.

Hayfield, T., & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, *27*(5).

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Kao, J., Wu, J., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 12002–12007.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of salt 23* (pp. 587–610).

Lewis, D. (1969). *Convention. A Philosophical Study.* Harvard University Press.

Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, *8*(1), 339–359.

Li, Q., & Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, *86*, 266–292.

Russell, B. (2012). *Probabilistic Reasoning and the Computation of Scalar Implicatures*. Unpublished doctoral dissertation, Brown University.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, *2*(4), 447–457.

Stalnaker, R. (1998). On the Representation of Context. *Journal of Logic, Language and Information*, *7*(1), 3–19.

---

[8] For an in depth discussion of this issue in syntactic processing, see, e.g., Jaeger (2010); Fine, Jaeger, Farmer, and Qian (2013).