# Still searching for principles: a response to Goodman et al.

We appreciate Goodman et al’s efforts to address concerns raised in Marcus and Davis (2013), but we remain unconvinced.

## Multiple models

To refresh the reader’s memory, we argued that multiple, equally plausible, Bayesian models could be constructed for the tasks under consideration, and Bayesian theories do not constrain which model applies in any given case. Without a prior theory on how to choose proper models, we suggested, the research program risks becoming an exercise in post hoc modeling. Additionally, we pointed out that the word “optimality” is used with many different meetings (see below).

The best rejoinder in Goodman et al (this issue) is their correct assertion that in a particular case — their “rational speech act” (RSA) theory of communication— there has been more consistency than we acknowledged. Throughout their work in this domain, the authors have, to their considerable credit, *consistently* used a particular choice rule that we suggested was arbitrary (viz. the “soft max” choice).

But our larger point stands: looking beyond RSA, Bayesian models in other domains use various choice rules, and no general rule for their selection has been proposed. (Battaglia, Hamrick, & Tenenbaum, 2013), (Kemp & Tenenbaum, 2008) and (Cain, Vul, & Mitroff, 2012) use hard max; (Gweon, Tenenbaum, & Schulz, 2010) use probability matching; (Griffiths & Tenenbaum, 2006 and 2011) use the median. (Smith & Vul, 2013) use a hard max rule for a task involving predicting a bouncing ball; for a very similar task, (Smith, Dechter, Tenenbaum, & Vul, 2013) use two separate soft maxes, with four parameters tuned to fit the data. Even in the rejoinder, we see no *principle* for deciding which rule applies in any given situation.

Moreover, though the choice rule in RSA is fixed, other aspects of the model remain fluid or arbitrary. Frank and Goodman (2012) assume without justification, for example, that the hearer knows what word choices are available to the speaker; this hardly seems plausible, yet the issue persists in (Kao, Wu, Bergen, & Goodman, 2014).

These same problems beset other papers that we did not include. For example, Gweon, Tenenbaum, and Schulz’s (2010)“squeaky toy” experiment purports to show that infants compute a posterior probability on hypotheses. Their model posits that the babies choose between four different hypotheses. But there was no principled justification for that particular model. In a more recent analysis, we found (Davis & Marcus, 2014) that there are forty-three different hypotheses that the babies might consider, all about equally plausible *a priori,* and over 7500 different Bayesian models that a theorist might use, all equally well motivated.

Rips, Asmuth, and Bloomfield (2013) point out the same flaw in the Bayesian theory of number learning proposed in (Piantadosi, Tenenbaum, & Goodman, 2012). The model relies on having a limited vocabulary of primitive concepts under consideration, and it is not explained how the child learner would select the appropriate vocabulary.

## Optimality

In our original critique we noted that strong, unwarranted, claims for the optimality of performance are often made, and that the notion of optimality varies between papers with no systematic criterion. We stand by these assertions.

**Strong claims:**  The word “optimal” or “optimize” is used in the titles of (Griffiths and Tenenbaum, 2006; Kording, Tenenbaum, & Shadmehr, 2007; Cain, Vul, & Mitroff, 2012; Piantadosi, Tiley, & Gibson, 2011) and claims of optimality or near optimality are made in (Teglas et al., 2011; Kao, Wu, Bergen, and Goodman, 2014) (Kao, Wu, Bergen, & Goodman, 2014) and many more. Oaksford and Chater (2009) argue that, “Behavioral predictions [should be] derived from the assumption that the cognitive system is solving this problem, optimally (or, more plausibly, approximately), under ... constraints.” Griffiths and Tenenbaum (2006) state that “everyday cognitive judgments follow . . . optimal statistical principles” and that there is “close correspondence between people’s implicit probabilistic models and the statistics of the world”. Sanborn, Masinghka, and Griffiths (2013) propose that “people’s judgments [about physical events] are based on optimal statistical inference over a Newtonian physical model that incorporates sensory noise and intrinsic uncertainty about the physical properties of the objects being viewed.” Frank’s (2013) more moderate view is the exception rather than the rule.

**Arbitrary criteria.** In our original paper, we demonstrated that the optimality claims in two of the experiments reported in (Griffiths & Tenenbaum, 2006) depended on arbitrary assumptions about what information in the problems the subjects are considering and which they are ignoring. Similarly, the justifications in (Oakford & Chater, 2009) and in (Tenenbaum & Griffiths, 2001) for viewing subjects’ non-normative answers to questions as in fact optimal depends on arbitrary assumptions about how the subjects’ interpretations differ from the experimenters. In the title of (Piantadosi, Tiley, & Gibson, Word lengths are optimized for efficient communication, 2011), “Word length is optimized for efficient communication,” the word “optimized” means little more than “pretty good”. The varying choice rules discussed above are each “optimal” in a different sense.

If each paper means something different by optimal, the overall claim becomes nearly meaningless. The cleverly-worded reply in Goodman et al., “*an* optimal analysis is not *the* optimal analysis” merely sidesteps the problem.

## Literature

What’s left? Goodman et al note that we didn’t cite, well, everything, including a variety of papers that hasn’t come out before our critique went to press. True enough, but we were hardly lax. We cited 12 papers by the authors of Goodman et al (compared to just one by ourselves), exceeding the maximum allowable number of references by four in order to squeeze as many as we did .More importantly, the additional papers that Goodman et al. mention hardly refute our argument. (Goodman et al. also chide us for focusing on work that was not “mature”; but in fact we focused primarily on articles in prestigious outlets like *Science* (Frank & Goodman, 2012), *PNAS* (Battaglia, Hamrick, & Tenenbaum, 2013)and *Psychological Science* (Griffiths & Tenenbaum, 2006). Since both Griffiths and Tenenbaum list that highly-cited paper as among their key publications, it hardly seems unfair to focus attention on it.)

## Conclusion

What’s most telling, however, is what’s absent. In our original piece we concluded that in many of the studies that use Bayesian models to characterize high-level cognition, there are many possible Bayesian models and many possible standards of optimality, and Bayesian theory offers no principled way to choose between them. We see nothing in the response of Goodman et al. that alleviates those concerns.

# References

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences* *, 110* (45), 18327-18332.

Cain, M., Vul, E. C., & Mitroff, S. (2012). A Bayesian Optimal Foraging Model of Human Visual Search. *Psychological Science* *, 23* (9), 1047-1054.

Davis, E., & Marcus, G. (2014 йил 22-April). *The Hypothesis Space in Gweon, Tenenbaum, and Schulz (2010).* From http://www.cs.nyu.edu/faculty/davise/papers/GTS.html

Endress, A. (2013). Bayesian learning and the psychology of rule induction. *Cognition* *, 127*, 159-176.

Frank, M. (2013). Throwing out the Bayesian baby with the optimal bathwater: Response to Endress. *Cognition* *, 128*, 417-423.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science* *, 336* (6084), 998.

Frank, M., Goodman, N., Lai, P., & Tenenbaum, J. .. (2009). Informative communication in word production and word learning. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, (pp. 206-211).

Griffiths, T., & Tenenbaum, J. (2006). Optimal prediction in everyday cognition. *Psychological Science* *, 17* (9), 767-773.

Gweon, H., Tenenbaum, J., & Schulz, L. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences* *, 107* (20), 9066-9071.

Kao, J., Wu, J., Bergen, N., & Goodman, N. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* .

Kemp, C., & Tenenbaum, J. (2008). The discovery of structural form. *Proceedings of the national academy of sciences* *, 105* (31), 10687-10692.

Kording, K., Tenenbaum, J., & Shadmehr, R. (2007). The dynamics of memory are a consequence of optimal adaptation to a changing body. *Nature Neuroscience* *, 10* (6), 779-786.

Oakford, M., & Chater, N. (2009). Precis of Bayesian rationality; The probabilistic approach to human reasoning. *Behavioral and Brain Sciences* *, 32*, 69-120.

Piantadosi, S., Tenenbaum, J., & Goodman, N. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* *, 123*, 197-217.

Piantadosi, S., Tiley, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* .

Rips, L., Asmuth, J., & Bloomfield, A. (2013). Can statistical learning bootstrap the integers? *Cognition* *, 128*, 320-330.

Smith, K., & Vul, E. (2013). Sources of Uncertainty in Intuitive Physics. *Topics in Cognitive Science* *, 5*, 185-199.

Smith, K., Dechter, E., Tenenbaum, J., & Vul, E. (2013). Physical predictions over time. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.*

Teglas, E., Vul, E. G., Gonzalez, M., Tenenbaum, J., & Bonatti, L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science* *, 332*, 1054-1059.

Tenenbaum, J., & Griffiths, T. (2001). The rational basis of representativeness. *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, (pp. 1036-1041).

Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review* *, 114* (2), 245-272.