

What are probabilistic models of cognition? Please begin by situating them with respect to alternatives (i.e., what are non-probabilistic models of cognition?). Then describe what you see as the two or three key strengths of this approach, illustrating each with examples.

Introduction

On the surface, probabilistic models of cognition are an application of Bayesian inference to modeling how people reason about the world. In particular, rather than viewing Bayes' rule in the traditional sense as $P(A|B) \propto P(B|A)P(A)$, probabilistic models of cognition assign a particular meaning to the variables as hypotheses (H) and data (D). With this interpretation, Bayes' rule becomes $P(H|D) \propto P(D|H)P(H)$. The significance of writing Bayes' rule in this way is that it gives us a framework for describing how people reason about the world by combining prior knowledge about the world (or *inductive bias*) with new observations (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010).

However, probabilistic models of cognition are much more than a simple application of Bayes' rule. I will argue that the strengths of probabilistic models lie in their ability to express complex, *structured representations* and to formalize explicit hypotheses about how people think the world works in terms of their *generative knowledge*. Probabilistic models of cognition are further strengthened by the strategy of *rational analysis* that is commonly used to construct them.

Here, I will begin by discussing the distinction between probabilistic models of cognition and non-probabilistic models of cognition. Second, I will explain how structured representations allow probabilistic models to capture complex forms of knowledge. Third, I will discuss how that knowledge may capture not just the structure of data and hypotheses, but of the generative process by which the data itself is created. Fourth, I will consider how the strategy of rational analysis provides a principled way of constructing probabilistic models.

What are non-probabilistic models of cognition?

Before jumping in to describing what are *non* probabilistic models of cognition, I will briefly summarize one way in which models can be distinguished: that is, what level of analysis they are posed at (Marr, 1971). Marr’s levels of analysis consist of the *computational* level (at which analyses are concerned with what the goal of a system is, and what the solution might look like), the *algorithmic* level (at which analyses are concerned with specific processes by which a solution to the goal can be computed), and the *implementation* level (at which analyses are concerned with the physical realization of the solution).

Non-probabilistic models of cognition usually fall at the *algorithmic* level of analysis: they are frequently committed to a particular type of representation. For example, connectionist models are committed to using distributed representations and a process of error-driven learning (gradient descent via backpropagation). In another example, the model theory (Johnson-Laird, 2012) is committed to a representation in terms of “mental models” and to processes which manipulate those mental models. Qualitative reasoning models are committed to “qualitative” representations of knowledge such as the knowledge that $X > Y$, but not specific values of X or Y (Kuipers, 1986; Forbus, 2011). Models of heuristics and biases (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974) typically assume a particular type of process or representation that might cause such systematic errors.

All of the examples listed above either make a strong commitment to a particular *algorithm* (e.g., backpropagation), to a particular *representation* (e.g., distributed representation), or to both. In contrast, probabilistic models of cognition are almost always formulated at the *computational* level of analysis and typically do not make assumptions about processes, though they may make abstract assumptions about representation. Probabilistic models of cognition, while formulated in terms of Bayesian models, do not make assumptions about the specific process by which those inferences are computed: for example, if a process-level model of response time and accuracy can be described as performing Bayesian inference (Bitzer, Park, Blankenburg, & Kiebel, 2014), then it is not inconsistent with a probabilistic

model formulated at the computational level.

It is important to differentiate the types of representations assumed by computational-level probabilistic models from the representations assumed by algorithmic-level models: the representations used by probabilistic models are concerned with the *structure* of information rather than a particular *encoding*. For example, a probabilistic model might assume a representation involving a causal Bayes' net (Griffiths & Tenenbaum, 2009), but it is agnostic as to whether that representation is actually instantiated as a set of distributed activations in a neural network, or as a logical set of nodes and edges, or as a matrix, etc.¹

Structured representations

One of the core strengths of probabilistic models are their ability to express abstract, structured representations. By interpreting Bayes' rule in terms of hypotheses and data, probabilistic models have the option of formulating the structure of hypotheses and data in whatever way is most relevant to the problem at hand. Here, I will give examples from several papers that formulate probabilistic models in terms of powerful representations.

In their paper on theory-based causal induction, Griffiths and Tenenbaum, 2009 outline a framework for how mental theories (that is, people's own theories about the world, not the modeler's theory about people) might be constructed. Griffiths and Tenenbaum, 2009 take the idea that Bayes' nets can be used for inductive reasoning and show how a probabilistic model operating over the structure of the Bayes' nets themselves can explain inductive inferences about causal structure in the world.

In another example, Kemp and Tenenbaum, 2008 also make use of a graph representation—specifically, a *graph grammar*—to show how people might extract structured representations from data. In this paper, the grammar defines various types of graphs (e.g., chains, trees,

¹There are some probabilistic models that do seem to make stronger commitments about the type of representation. For example, Kemp, Tenenbaum, Niyogi, and Griffiths, 2010 assume knowledge to be represented in a binary matrix or tensor. I would interpret this assumption to not be a core piece of the model, however, as long as whatever representation was used somehow expressed relationships between pairs of entities. In order to actually compute the model predictions, a specific representation must sometimes be chosen.

hierarchies, etc.) and the probabilistic model over the grammar expresses how a particular graph representation might be chosen for a specific set of data.

Using a different type of grammar, Ullman, Goodman, and Tenenbaum, 2012 express theories in terms of propositions and perform inference over theories with the help of a probabilistic grammar specifying how propositions combine into Horn Clauses to form laws within the theory.

Generative knowledge

Closely related to the idea of structured representations is the idea of *generative knowledge*. The term “generative” is not always used in the same way, so I will attempt to provide a definition of what I mean by “generative” in this context. First, in machine learning, a generative model is defined as modeling the joint distribution $P(H, D)$ from which the posterior $P(H|D)$ can be computed. This is in contrast to a discriminative model, which models the posterior distribution directly (Ng & Jordan, 2002). In this sense, most probabilistic models are generative models as they almost always define the full joint distribution. In the context of cognitive science, however, there is a more nuanced interpretation of “generative”.

Battaglia, Kersten, and Schrater, 2012 make the distinction between the generative *process* and generative *knowledge*. They define the generative process to be the procedure by which the world gives rise to our perceptions, and generative knowledge to be a person’s hypothesis about how generative process works. The generative process should be distinguished from structured representations, which apply to the form of the variables in a probabilistic model (e.g., H and D). In contrast, the generative process reflects assumptions about the *distributions* and processes by which H is generated or by which D is generated from H .

The generative knowledge in a probabilistic model may be relatively simple: for example, knowledge about the statistics of the world. Griffiths and Tenenbaum, 2006 and Lewandowsky, Griffiths, and Kalish, 2009 demonstrate that people have detailed knowledge about the empirical distributions of every-day phenomena such as cake baking times, length

of poems, or lifespans.

Probabilistic models can also express egocentric generative knowledge, such as the reliability of various sensory cues. This type of generative knowledge has been widely explored in probabilistic models of perception (Battaglia et al., 2012), and has been used to explain phenomena including visual illusions (Weiss, Simoncelli, & Adelson, 2002) and sensory cue integration when the reliability of a cue is manipulated (Ernst & Banks, 2002). It has further been suggested that such generative knowledge may underlie other perceptual processes such as visual segmentation (Yuille & Kersten, 2006) or speech perception (Halle & Stevens, 1959, 1962; Bever & Poeppel, 2010).

Finally, probabilistic models can encode generative knowledge about complex, nonlinear processes such as how physical objects behave (Teglas et al., 2011; Sanborn, Mansinghka, & Griffiths, 2013; Smith & Vul, 2013).

Rational analysis

Given assumptions about structured representations and generative knowledge, a probabilistic model of cognition encodes a highly specific and falsifiable hypothesis about how people reason: hypotheses are structured in *this particular* form, and generative knowledge specifies that hypotheses and data are generated in *this particular* way. However, in general there always exists some probabilistic model that can capture any phenomena; how, then, should one choose *which* probabilistic to act as a hypothesis? The strategy typically taken is that of *rational analysis* (Anderson, 1990). The original version of rational analysis proposed by Anderson, 1990 argued for constructing models which are optimal solutions to problems posed at the computational level of analysis (Marr, 1971) while taking into account environmental constraints. This approach has been remarkably successful in generating hypotheses (instantiated as probabilistic models of cognition) for topics ranging from notions of similarity (Tenenbaum & Griffiths, 2001), physical intuitions (Teglas et al., 2011; Sanborn et al., 2013), perception (Weiss et al., 2002; Ernst & Banks, 2002; Körding & Wolpert, 2004; Yuille

& Kersten, 2006), and theory learning (Kemp, Perfors, & Tenenbaum, 2007; Griffiths & Tenenbaum, 2009; Kemp et al., 2010; Ullman et al., 2012), just to name a few domains.

Recently, it has been proposed that rational analysis can go even further than just guiding hypothesis construction at the computational level. Griffiths, Lieder, and Goodman, 2015 suggest a new strategy of *resource-rational analysis*, in which a hypothesis is first generated at the computational level. Second, a class of approximation algorithms are posited for actually computing the solution proposed by the hypothesis. Third, these approximation algorithms are analyzed in order to determine the optimal way to use computational resources under the assumptions of the algorithm. For example, particle filters are an approximation algorithm for Bayesian inference, and they can be analyzed to see *how many* particles are optimal for a bounded agent to use. This approach has already seen success at bridging between the computational and algorithmic levels of analysis, with rational process models suggesting explanations for process-level heuristics and biases such as anchoring (Lieder, Griffiths, & Goodman, 2012), availability (Lieder, Hsu, & Griffiths, 2014), primacy and recency (Abbott & Griffiths, 2011), probability matching (Vul, Goodman, Griffiths, & Tenenbaum, 2014), and theory change (Ullman et al., 2012).

Conclusion

To summarize, probabilistic models are distinct from non-probabilistic models in that they are posed at the computational level of analysis. Probabilistic models involve two key components: the ability to express structured representations, and the ability to express people's generative knowledge of how the world works. Probabilistic models are further supplemented by the strategy of rational analysis. Although I would argue that rational analysis is not a *component* of probabilistic models of cognition, it is a crucial *strategy* by which hypotheses are selected and refined. In the resource-rational formulation of rational analysis, it additionally strengthens the probabilistic modeling enterprise by helping to connect probabilistic models posed at the computational level to the algorithmic level.

References

- Abbott, J. T. & Griffiths, T. L. (2011). Exploring the influence of particle filter parameters on order effects in causal learning. In *Proceedings of the 33rd annual conference of the cognitive science society*. Retrieved from https://cocosci.berkeley.edu/josh/papers/cogsci2011_pfOrderEffects.pdf
- Anderson, J. R. (1990). Introduction. In *The adaptive character of thought* (Chap. 1, pp. 1–40). Lawrence Erlbaum Associates.
- Battaglia, P. W., Kersten, D., & Schrater, P. R. (2012). The Role of Generative Knowledge in Object Perception. In J. Trommershauser, K. P. Körding, & M. S. Landy (Eds.), *Sensory cue integration*. Oxford University Press.
- Bever, T. G. & Poeppel, D. (2010). Analysis by Synthesis: A (Re-)Emerging Program of Research for Language and Vision. *Biolinguistics*, *43*(2), 174–200. Retrieved from http://www.psych.nyu.edu/clash/dp_papers/bever.poeppel.pdf
- Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, *8*(102). doi:[10.3389/fnhum.2014.00102](https://doi.org/10.3389/fnhum.2014.00102)
- Ernst, M. O. & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:[10.1038/415429a](https://doi.org/10.1038/415429a)
- Forbus, K. D. (2011). Qualitative modeling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 374–391. doi:[10.1002/wcs.115](https://doi.org/10.1002/wcs.115)
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. doi:[10.1016/j.tics.2010.05.004](https://doi.org/10.1016/j.tics.2010.05.004)
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. doi:[10.1111/tops.12142](https://doi.org/10.1111/tops.12142)
- Griffiths, T. L. & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773. doi:[10.1111/j.1467-9280.2006.01780.x](https://doi.org/10.1111/j.1467-9280.2006.01780.x)
- Griffiths, T. L. & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661–716. doi:[10.1037/a0017201](https://doi.org/10.1037/a0017201)
- Halle, M. & Stevens, K. M. (1959). Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (Eds.), *Proceedings of the seminar of speech compression and processing* (Vol. 2, D7).
- Halle, M. & Stevens, K. M. (1962). Speech recognition: a model and a program for research. *IRE Transactions on Information Theory*, *8*(2), 155–159. doi:[10.1109/TIT.1962.1057686](https://doi.org/10.1109/TIT.1962.1057686)
- Johnson-Laird, P. N. (2012). Inference with Mental Models. In *The oxford handbook of thinking and reasoning* (pp. 134–145). doi:[10.1093/oxfordhb/9780199734689.001.0001](https://doi.org/10.1093/oxfordhb/9780199734689.001.0001)
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. doi:[10.1037/h0034747](https://doi.org/10.1037/h0034747)
- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. doi:[10.1111/j.1467-7687.2007.00585.x](https://doi.org/10.1111/j.1467-7687.2007.00585.x)

- Kemp, C. & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10687–92. doi:[10.1073/pnas.0802631105](https://doi.org/10.1073/pnas.0802631105)
- Kemp, C., Tenenbaum, J. B., Niyogi, S., & Griffiths, T. L. (2010). A probabilistic model of theory formation. *Cognition*, *114*(2), 165–196. doi:[10.1016/j.cognition.2009.09.003](https://doi.org/10.1016/j.cognition.2009.09.003)
- Körding, K. P. & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. doi:[10.1038/nature02169](https://doi.org/10.1038/nature02169)
- Kuipers, B. (1986). Qualitative Simulation. *Artificial Intelligence*, *29*(3), 289–338. doi:[10.1016/0004-3702\(86\)90073-1](https://doi.org/10.1016/0004-3702(86)90073-1)
- Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The Wisdom of Individuals: Exploring People’s Knowledge About Everyday Events Using Iterated Learning. *Cognitive science*, *33*, 969–998. doi:[10.1111/j.1551-6709.2009.01045.x](https://doi.org/10.1111/j.1551-6709.2009.01045.x)
- Lieder, F., Griffiths, T. L., & Goodman, N. D. (2012). Burn-in, bias, and the rationality of anchoring. *Advances in Neural Information Processing Systems*, *25*. Retrieved from <http://papers.nips.cc/paper/4719-burn-in-bias-and-the-rationality-of-anchoring>
- Lieder, F., Hsu, M., & Griffiths, T. L. (2014). The high availability of extreme events serves resource-rational decision-making. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, (2). Retrieved from <https://mindmodeling.org/cogsci2014/papers/445/paper445.pdf>
- Marr, D. (1971). The Philosophy and the Approach. In *Vision* (Chap. 1). Retrieved from http://web.stanford.edu/class/psych209a/ReadingsByDate/01_07/Marr82Philosophy.pdf
- Ng, A. Y. & Jordan, M. I. (2002). On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes. *Advances in Neural Information Processing Systems*, *14*.
- Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 1–77. doi:[10.1037/a0031912](https://doi.org/10.1037/a0031912)
- Smith, K. A. & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*, *5*(1), 185–199. doi:[10.1111/tops.12009](https://doi.org/10.1111/tops.12009)
- Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–9. doi:[10.1126/science.1196404](https://doi.org/10.1126/science.1196404)
- Tenenbaum, J. B. & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *The Behavioral and Brain Sciences*, *24*, 629–640, discussion 652–791. doi:[10.1017/S0140525X01000061](https://doi.org/10.1017/S0140525X01000061)
- Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi:[10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124)
- Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4). doi:[10.1016/j.cogdev.2012.07.005](https://doi.org/10.1016/j.cogdev.2012.07.005)
- Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*(4), 599–637. doi:[10.1111/cogs.12101](https://doi.org/10.1111/cogs.12101)

- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604. doi:[10.1038/nn858](https://doi.org/10.1038/nn858)
- Yuille, A. L. & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308. doi:[10.1016/j.tics.2006.05.002](https://doi.org/10.1016/j.tics.2006.05.002)