Source Invariance and Probabilistic Transfer: A Testable Theory of Probabilistic Neural Representations

Samuel Lippl (samuel.lippl@columbia.edu)
Center for Theoretical Neuroscience, Columbia University
3227 Broadway, New York, NY 10027, USA

Raphael Gerraty (rgerraty@gmail.com)
Zuckerman Mind Brain Behavior Institute, Columbia University
3227 Broadway, New York, NY 10027, USA

John Morrison (jmorriso@barnard.edu)
Department of Philosophy, Barnard College
3009 Broadway, New York, NY 10027, USA

Nikolaus Kriegeskorte (nk2765@columbia.edu)
Zuckerman Mind Brain Behavior Institute, Columbia University
3227 Broadway, New York, NY 10027, USA

Abstract

As animals interact with their environments, they must infer properties of their surroundings. Some animals, including humans, can represent uncertainty about those properties. But when, if ever, do they use probability distributions to represent their uncertainty? It depends on which definition we choose. In this paper, we argue that existing definitions are inadequate because they are untestable. We then propose our own definition, which defines probabilistic representations in terms of two properties: (1) invariance to the source of uncertainty and (2) consistency in how this uncertainty is taken into account by downstream computations across multiple tasks.

Keywords: probabilistic representation; task transfer; uncertainty representation; Bayesian inference

Introduction

Uncertainty poses a fundamental challenge for perceptual systems. There is strong evidence that humans and some animals address this challenge by representing their uncertainty in order to adjust their behavior and make better decisions (Bach & Dolan, 2012). However, it has remained unclear *how* exactly they represent their uncertainty. Probability theory is a normative language for representing uncertainty and it is therefore tempting to think that humans and animals "use probability distributions" (Ma & Jazayeri, 2014). Indeed probabilistic accounts of perception are widespread in psychology, neuroscience, and philosophy and have been successful at predicting brain activity and behavior (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Vilares & Kording, 2011).

However, theories of probabilistic representation have long been controversial. In particular, a precise definition of probabilistic representation remains lacking and many researchers have argued that probabilistic representations do not constitute a falsifiable theory (Jones & Love, 2011; Bowers & Davis, 2012). Below we describe common definitions of probabilistic representations arguing that they fail to make falsifiable predictions. We then propose our own theory that grounds probabilistic representations in two criteria: source invariance and probabilistic transfer.

Defining probabilistic representation

All definitions of probabilistic representation must answer three questions. Below we describe these three question and survey common answers.

(1) What does it mean to specify a probability distribution? At one extreme, we may call any representation that allows the experimenter to infer a probability distribution probabilistic. However, any neural representation could be used as "evidence" to determine a posterior distribution over its represented variable. Under this view, every representation would therefore be probabilistic. While this can be a useful perspective for describing representations, an empirically testable notion of probabilistic representation has to be more constrained.

Perhaps this constraint can come from the nature of a probability distribution. Thus, at the opposite extreme, we may require that the brain represent the set of mathematical concepts that are used to define probability distributions. This would capture humans explicitly thinking about probability theory, but would exclude humans and animals taking uncertainty into account without representing these mathematical concepts.

The apparent dichotomy between these views is one source of skepticism about probabilistic representation. We suggest an alternative: probability theory defines not just probability distributions but also a set of rules to construct and use these distributions. Perhaps, being a probabilistic representation is a matter of approximating these rules (Harman, 1982). Though rarely stated explicitly, this is the view most experimental investigations seem to take.

This approach also creates a fundamental issue, however. If the generative model guiding probabilistic inference is not explicitly represented, we need to infer it. We call this issue *model indeterminacy* and show below that it renders many commonly used definitions unfalsifiable. This is not to say that such definitions are not useful in describing neural representations. However, they are best understood as a "language" that enables any neural representation to be described from a probabilistic perspective rather than as a testable theory.

(2) What aspects of a distribution must be represented?

MAP estimation and priors cannot be distinguished from optimizing an objective function. The simplest version of representing an aspect of a probability distribution is to represent its most likely value, i.e. to maximize the conditional probability $p(z|x)$ about the variable z given the input x. Indeed probabilistic models in neuroscience often take this form (Ernst & Banks, 2002). However, we can frame the same systems as maximizing an objective function and indeed almost any system that maximizes an objective function could alternately be taken to implement MAP estimation. To see why, suppose you compute $\max_z L(x,z)$ for some objective function L. As long as $C(x) := \int_z L(x,z) < \infty$, we can define a probability distribution proportional to this objective function, $p(x|z) = L(x,z)/C(x)$, which could equivalently be maximized. Note that this argument extends to taking into account priors as well. Thus, MAP estimation and priors are not testable features of probabilistic representations.

Representations of uncertainty are not necessarily probabilistic. In response, we may require that probabilistic representations encode not just a point estimate but a range of uncertainty. However, the same issue arises: virtually any heuristic uncertainty representation amounts to optimal probabilistic inference under some prior. This means that either any uncertainty representation is probabilistic or we need to further constrain our definition. In addition, this perspective introduces a new problem: when a single nuisance variable is used as a source of uncertainty, it becomes ambiguous whether a representation encodes uncertainty or that source of uncertainty.

Indeed, this is a common set-up in studies investigating probabilistic representations and we call this problem *representational indeterminacy*.

(3) What constraints are there on the generative models? One prominent response to the issue of model indeterminacy has been to posit that probabilistic representations are the probability distributions arising from the optimal generative model. This is often justified by the fact that probabilistic computations can be necessary for optimal behavior (Ramsey, 1926). However, this does not imply that optimal behavior is necessary for probabilistic computations. Indeed probability is a language in which we can state our prior assumptions and make consistent inferences whether or not those assumptions are correct. An optimal-observer constraint excludes probabilistic inferences based on imperfect knowledge – arguably the prototypical and most frequent scenario.

A Testable Theory: Source Invariance and Probabilistic Transfer

We now propose our own theory, using two criteria: source invariance (which resolves representational indeterminacy) and probabilistic transfer (which resolves model indeterminacy).

A test for source-invariant uncertainty encoding

To address the issue of representational indeterminacy, uncertainty representations should encode uncertainty invariant to its source (Sahani & Dayan, 2003; Walker et al., 2022). Source invariance not only renders uncertainty representations testable, but also benefits the animal by enabling reuse of the same uncertainty representation across multiple uncertainty sources. Source invariance comes in degrees: a representation may only be invariant to certain sources of uncertainty, or may only be approximately invariant. The more strongly uncertainty is encoded as opposed to other information, the more it becomes a representation of uncertainty.

A test for task-transferable uncertainty decoding

To address the issue of model indeterminacy, we propose to analyze how subjects generalize across tasks. If a subject is using a probabilistic computation, we can use their performance on the first task to characterize their generative model. We can then use those constraints on the generative model to predict their behavior on a second task. If an organism is using a probabilistic representation, probability theory provides a clear justification for these predictions. But if the subject is using a non-probabilistic representation, there is no reason to expect probability theory to make accurate predictions. We call this generalization “probabilistic transfer.” Importantly, probabilistic transfer tests whether subjects (self-)consistently operate over uncertainty representations. This is the defining feature of probability theory.

Notably, only some probabilistic computations allow for such a test. In particular, MAP estimation can never be used to test for a probabilistic computation, as we could always alternately understand the computation as maximizing an objective function (even across tasks). On the other hand, marginalization (e.g. computing an expected value or a marginal probability) as well as change of variables (i.e. transforming a probability distribution of z into a probability distribution of $f(z)$) can serve as discernible probabilistic computations.

Like source invariance, probabilistic transfer benefits the animal. Probabilistic transfer enables the animal to generalize to new levels of uncertainty that it has never experienced on a particular task (Maloney & Mamassian, 2009; Koblinger, Fiser, & Lengyel, 2021). Like source invariance, probabilistic transfer comes in degrees. To be considered probabilistic, a representation of uncertainty does not need to be capable of marginalization and change of variables for all possible downstream computations and it does not need to approximate the laws of probability theory perfectly. Rather, the more functions it can generalize to and the closer its approximation, the more probabilistic a representation.

Evidence

Some studies indicate that probabilistic representations are used for marginalization in perceptual decision making, in particular within the related framework of Bayesian transfer (Trommershäuser, Gepshtein, Maloney, Landy, & Banks, 2005; Whiteley & Sahani, 2008). Bayesian transfer uses an analogous task-transfer criterion to test for evidence that a neural computation relies on Bayesian decision theory. As a result, there is substantial overlap with our own criterion, but there are instances of Bayesian transfer that are not instances of probabilistic transfer (e.g. computations that can be explained as MAP estimation (Sato & Kording, 2014)) and there are instances of probabilistic transfer that are not Bayesian transfer (e.g. change of variables that need not fall under the paradigm of Bayesian decision theory).

Finally, our theory can inform different theories of neural probabilistic codes. In particular, marginalization and change of variables are linear operations for two prominent theories, distributed distributional codes (Zemel, Dayan, & Pouget, 1998) and sampling-based inference (Moreno-Bote, Knill, & Pouget, 2011). This provides a straightforward basis for probabilistic transfer in these formats. In contrast, while probabilistic population codes (Ma, Beck, Latham, & Pouget, 2006) can learn to marginalize (Beck, Latham, & Pouget, 2011), it is not clear how they would generalize such a marginalization operation to new levels of uncertainty.

Conclusion

Researchers have long debated whether humans and animals have probabilistic representations. A central issue in this debate has been a lack of agreement on what it means to have a probabilistic representation. Here we have argued that this question should be given careful consideration and have highlighted issues with prevalent frameworks. We have then proposed our own theory which we argue overcomes these issues.

Acknowledgement

This work was supported in part by a subaward from Duke University of Templeton World Charity Foundation Grant “Summer Seminars in Neuroscience and Philosophy” to Walter Sinnott-Armstrong.

References

Bach, D. R., & Dolan, R. J. (2012, August). Knowing how much you don’t know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*(8), 572–586. Retrieved from https://doi.org/10.1038/nrn3289 doi: 10.1038/nrn3289

Beck, J. M., Latham, P. E., & Pouget, A. (2011, October). Marginalization in Neural Circuits with Divisive Normalization. *The Journal of Neuroscience*, *31*(43), 15310. Retrieved from http://www.jneurosci.org/content/31/43/15310.abstract doi: 10.1523/JNEUROSCI.1706-11.2011

Bowers, J. S., & Davis, C. J. (2012, May). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, *138*(3), 389–414. (Place: United States) doi: 10.1037/a0026450

Ernst, M. O., & Banks, M. S. (2002, January). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. Retrieved from https://doi.org/10.1038/415429a doi: 10.1038/415429a

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010, August). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. Retrieved 2022-12-23, from https://www.sciencedirect.com/science/article/pii/S1364661310001129 doi: 10.1016/j.tics.2010.05.004

Harman, G. (1982). Conceptual Role Semantics. *Notre Dame Journal of Formal Logic*, *28*(April), 242–56. (Publisher: Academic Press) doi: 10.1305/ndjfl/1093883628

Jones, M., & Love, B. C. (2011, August). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *The Behavioral and brain sciences*, *34*(4), 169–188; disuccsion 188–231. (Place: England) doi: 10.1017/S0140525X10003134

Koblinger, , Fiser, J., & Lengyel, M. (2021, April). Representations of uncertainty: where art thou? *Current Opinion in Behavioral Sciences*, *38*, 150–162. Retrieved 2022-12-27, from https://www.sciencedirect.com/science/article/pii/S2352154621000577 doi: 10.1016/j.cobeha.2021.03.009

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006, November). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. Retrieved from https://doi.org/10.1038/nn1790 doi: 10.1038/nn1790

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual review of neuroscience*, *37*(1), 205–220.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual neuroscience*, *26*(1), 147–155. (Publisher: Cambridge University Press)

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011, July). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, *108*(30), 12491–12496. Retrieved 2024-04-09, from https://www.pnas.org/doi/abs/10.1073/pnas.1101430108 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.1101430108

Ramsey, F. P. (1926). Truth and Probability. *Histoy of Economic Thought Chapters*, 156–198. (Publisher: McMaster University Archive for the History of Economic Thought)

Sahani, M., & Dayan, P. (2003, October). Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural computation*, *15*(10), 2255–2279. (Place: United States) doi: 10.1162/089976603322362356

Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 13–13. (Publisher: The Association for Research in Vision and Ophthalmology)

Trommershäuser, J., Gepshtein, S., Maloney, L. T., Landy, M. S., & Banks, M. S. (2005, August). Optimal compensation for changes in task-relevant movement variability. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *25*(31), 7169–7178. doi: 10.1523/JNEUROSCI.1906-05.2005

Vilares, I., & Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. (Publisher: Wiley Online Library)

Walker, E. Y., Pohl, S., Denison, R. N., Barack, D. L., Lee, J., Block, N., ... Meyniel, F. (2022, December). *Studying the neural representations of uncertainty.* arXiv. Retrieved 2023-06-01, from http://arxiv.org/abs/2202.04324 (arXiv:2202.04324 [q-bio]) doi: 10.48550/arXiv.2202.04324

Whiteley, L., & Sahani, M. (2008, March). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision*, *8*(3), 2–2. Retrieved 2022-08-30, from https://doi.org/10.1167/8.3.2 doi: 10.1167/8.3.2

Zemel, R. S., Dayan, P., & Pouget, A. (1998, February). Probabilistic Interpretation of Population Codes. *Neural Computation*, *10*(2), 403–430. Retrieved 2022-01-15, from https://doi.org/10.1162/089976698300017818 doi: 10.1162/089976698300017818