Human Development 2016;59:26–33
DOI: 10.1159/000447339

Human Development

Piaget, Probability, Causality, and Contradiction

Commentary on Tourmen

Amy Perfors

School of Psychology, North Terrace Campus, University of Adelaide, Adelaide, S.A., Australia

Key Words
Bayesian nets · Cognitive development · Probabilistic models

The goal of the Tourmen [this issue] paper is to provide a systematic review of the Piagetian literature, with the objective of more fully understanding the links between Piaget and modern probabilistic models of learning. This is a worthy aim, and the paper is quite interesting. Piaget was one of the foundational thinkers in the field, upon whose ideas the edifice of modern developmental psychology is built. Probabilistic models of learning are a recent but highly successful approach to understanding human cognition. Although probabilistic models did not emerge directly from Piaget's thinking, it can be useful and enlightening to trace the historical commonalities and divergences between the two. More profoundly, it is especially useful for current theorists to occasionally look back from whence we came, to ensure that the valuable insights and ideas of yesteryear are not lost in the hubbub and excitement of engaging with the new approaches and tools of the present.

My goal in this commentary is to synthesize and build on some of the key points of the Tourmen paper. The overall aim is to contextualize and elaborate on what probabilistic models are and what they offer, as well as to highlight some Piagetian insights that our theories nowadays would do well to grapple more with. To that end, the first part of this paper focuses on making some important distinctions between *core principles* of the probabilistic approach, *theoretical implications* that (sometimes, but not always) follow from those principles, and *empirical findings* that are relevant to researchers who adopt the approach, but that themselves incorporate no special reliance on it. The second part turns to Piaget's insights, focusing on those that are

KARGER

E-Mail karger@karger.com
www.karger.com/hde

0018–716X/16/0591–0026$39.50/0

Amy Perfors
School of Psychology, North Terrace Campus
University of Adelaide
Adelaide, SA 5005 (Australia)
E-Mail amy.perfors@adelaide.edu.au

not directly paralleled in current thought. To what extent can we gain further insight from transporting Piaget's ideas to a modern context and testing them with modern tools?

The Probabilistic Approach: Some Key Distinctions

What do we mean by the probabilistic approach to cognitive science? Identifying the key components of the approach is a matter of some contention within the literature [see, e.g., Bowers & Davis, 2012; Goodman, Frank, Griffiths, Tenenbaum, Battaglia, & Hamrick, 2015; Griffiths, Chater, Norris, & Pouget, 2012; Marcus & Davis, 2015; Tauber, Navarro, Perfors, & Steyvers et al., under review]. Nevertheless, doing so is a requirement for being able to make sensible claims about what we can learn from it. To that end, I think there are three important distinctions to make. First, what are the core principles of the probabilistic approach? What features *define* and *shape* it to the extent that it would be unrecognizable without them? To what extent are these principles shared by other theories and approaches within cognitive science? Second, what theoretical implications can follow from the core principles? To what extent are they integral, as opposed to peripheral? Third, what empirical findings has research within the probabilistic framework helped to elucidate or understand? To what extent do those findings *depend* on the approach for their meaning or importance?

The first issue centers around defining the core principles of the probabilistic approach, as elucidated by Oaksford & Chater [2007], Perfors, Tenenbaum, Griffiths, & Xu [2011], and Tenenbaum, Kemp, Griffiths, & Goodman [2011]. Also known as the Bayesian approach, it conceives of learning as involving the evaluation of hypotheses h according to how well they capture the observed data d in the environment. It is presumed that the learner is sensitive to two kinds of information about hypotheses and data their prior probability of the hypothesis, $p(h)$, and the likelihood of the data given the hypothesis, $p(d|h)$. These two elements are combined using Bayes' rule, which is derived from the rules of probability theory and thus defines a normative standard for reasoning about and weighting both factors appropriately.

What constitutes a hypothesis? Here is where both the power and the flexibility of probabilistic models really come into play (although some might argue *too* much power and *too* much flexibility). Hypotheses can be anything that one can define a sensible prior and likelihood over. This includes representations as simple as points in a metric space, to areas or shapes, to grammars, to equations, and to causal networks. The hypotheses are defined in whatever representational space is appropriate for the problem, as conceptualized by the researcher. They are a key part of the *theoretical claim* being made by the researcher for the particular problem under consideration, but the nature of any one set of hypotheses is not a core part of the probabilistic approach *per se*. Within the Bayesian framework, many different kinds of hypotheses can be represented and evaluated. The approach provides a language for performing that evaluation, but the nature of those hypotheses is not itself a central part of the approach.

This point is important because the Tourmen paper suggests that the probabilistic approach originated from the hypothesis that "Knowledge is structured around causality," [this issue] and goes on to discuss it in ways that appear to presume that

DOI: 10.1159/000447339

the probabilistic approach is identical to modeling with Bayesian nets. While it is true that the work of Pearl [2000] was foundational to the development of Bayesian models in cognitive science, this does not mean that a core principle of such models is that they be causal, or that Bayesian nets are the total of probabilistic models of cognition. There are many probabilistic models that are not causal, primarily because they address different questions entirely [e.g., Griffiths, Steyver, & Tenenbaum, 2007; Kemp, 2012; Kemp & Tenenbaum, 2008; Perfors, Tenenbaum, & Regier, 2011; Sanborn, Navarro, & Griffiths, 2010; Tenenbaum & Griffiths, 2001]. The true breadth of probabilistic models is relevant not just for properly conceptualizing their nature and implications, but also for the discussion later in the paper of how such models map onto causality in Piaget's theories. Just as Piaget's ideas did not entirely revolve around causality – for him, as Tourmen points out, logico-mathematical operations were also central – probabilistic models of cognitive science do not entirely revolve around causal models. It is of course true that *some* Bayesian models are explicitly causal, and that it is their causal nature that contributes to their explanatory power in those domains [e.g., Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Gopnik & Wellman, 2012; Griffiths & Tenenbaum, 2009]. This means that the consideration of causality is important; it is just not a core, definitional consideration or feature of probabilistic models.

If causality is not a central theoretical claim of probabilistic models, what is it? Here I would point to two things: probability and hypothesis testing. Tourmen also highlights both of these aspects of the Bayesian framework. Hypothesis testing involves identifying a set of hypotheses, evaluating how well each accounts for the data, modifying the hypotheses (or re-weighting them) based on that initial calculation, and iterating through this procedure again. Key to this sequence of events is that we learn from experience in a "backwards and forwards movement between prior knowledge (beliefs), hypotheses, and new data" [this issue]. The probabilistic nature of the framework is also an essential part of it: Bayes' rule requires that the elements be specifiable as probabilities from zero to one, and their normative character derives from the fact that they must behave in accordance with probability theory. The approach predicts, therefore, that people must be sensitive to probability or frequency information in the world, and should be able to reason sensibly with it – perhaps not in all contexts or in all ways, but if their behaviors were not well matched to the predictions of probabilistic models in at least some important cases, it would not be a very interesting or appropriate theory. The Tourmen paper does a nice job of considering the role of both hypothesis testing and probability in Piagetian thought, setting up an interesting contrast to how they are viewed within Bayesian models.

That said, it is also important to contrast probabilistic models with other current theoretical approaches. Within cognitive science, there are many modeling frameworks other than Bayesian ones. Some of the most widely used include signal detection theory [Macmillan & Creelman, 2004], diffusion modeling [Ratcliff & Smith, 2004], associationism [Rumelhart & McClelland, 1986], or approaches that are more specific to certain problems like the generalized context model [Nosofsky, 1986] or latent semantic analysis [Landauer, Foltz, & Laham, 1998]. All of these can be captured by figure 1 in the Tourmen paper – that is, they can be described as taking input of some sort, producing an output or an action based on some kind of internal transformation of the input, receiving feedback based on the results of that action or output, and iterating accordingly.

When one realizes that this qualitative description could apply to *any* of these modern theories of learning, the correspondence between it and the Bayesian approach becomes rather less striking. This is not to say that the correspondence is uninteresting – it reveals both the great intellectual debt that all modern theories bear to Piaget, as well as the fact that some aspects of learning are universal. Computational theories differ from each other about the precise quantitative claims about *how* one's output or action is updated in light of some input and how feedback is incorporated. Since Piaget was not in the business of making precise quantitative claims in the same way, his thinking is consistent with all or most modern computational theories, at least at this level of description.

The central constructs of the probabilistic approach thus contain only the notions of hypothesis spaces and hypothesis testing, the representation of theories and learning with probabilities, and the idea of priors and likelihoods that can be combined using Bayes' rule. However, there are some key theoretical implications that often follow from these central constructs, even if they are not central themselves. The idea of causality as represented by Bayesian nets, as well as the effects thereof that are traced and discussed within the Tourmen paper, is one such implication. Another is the idea that new concepts and learning biases are driven by rational inferential learning rather than progression through stages that begin with sensory-motor primitives [e.g., Carey, 2009; Gopnik et al., 2004; Xu & Griffiths, 2011]. Although this idea is certainly consistent with the Bayesian approach – indeed, one would be hard-pressed to find a Bayesian researcher who disagrees with it – it is not definitionally a part of probabilistic modeling. It is at least possible, within the Bayesian framework, to model a learner who begins with sensory-motor primitives or whose learning biases are innately set rather than driven by rational inferential learning.

Finally, another important distinction to be drawn is between the tenets of the probabilistic modeling approach – both central and implicational – and the empirical findings implemented by or even discussed by Bayesian researchers. There is a variety of interesting empirical work suggesting, among other things, that children are capable of sophisticated inferences from and about teachers [e.g., Bonawitz, Shafto, Gweon, Goodman, Spelke, & Schulz, 2011; Buschbaum, Bridgers, Weisberg, & Gopnik, 2012; Shafto, Eaves, Navarro, & Perfors, 2012], that children and babies are already sensitive to numbers, probability, and probabilistic reasoning [e.g., Denison, Reed, & Xu, 2013; Feigenson, Dehaene, & Spelke, 2004; Gweon, Tenenbaum, & Schulz, 2010], and that some of children's basic concepts and realizations emerge far earlier than Piaget suggests [e.g., Baillargeon & Graber, 1988]. This empirical work is fascinating and exciting, but is not itself a part of the probabilistic approach – though at times the Tourmen paper discusses it as though it is [this issue].

Interestingly, much of this evidence suggests that infants and children have a vast store of nonverbal knowledge that is only accessible if the conditions are maximally supportive (which explains why in many cases Piaget was unaware of it; it required sophisticated modern experimental techniques to pick up). This, in turn, suggests that much of what we call "learning" may really be about expanding the contexts and situations in which that knowledge becomes accessible, rather than hypothesis testing in a probabilistic sense. As such, this evidence poses a challenge to the probabilistic approach, at least insofar as it is reliant on the notion that learning proceeds through hypothesis testing. In any case, it is clear that this kind of empirical evidence, arrived

DOI: 10.1159/000447339

at through experimental methods, is a separate kind of thing than Bayesian probabilistic modeling.

These considerations do lead to an open question: what aspects of Piaget's insights *do not* have parallels within the Bayesian approach? This is where some of the most valuable insights of the Tourmen paper lie, because it is in those that we may gain new ideas for future directions and unexplored pathways within the probabilistic framework. It is to that question that I now turn.

The Piagetian Framework: Some Key Insights

Some of the most fascinating parts of the Tourmen paper occur where it highlights the *differences* between Piaget's thought and that of the probabilistic approach. Of those differences, I found three in particular especially interesting, primarily because they contain within them some tantalizing suggestions for where the field might go in the future. None of them have been abandoned because they were proven to be wrong; rather, they are not an intrinsic part of the probabilistic approach simply because not all approaches can be all things, and because some things are easier to incorporate. I highlight these three points because it may be useful to consider, now, how we might begin including their insights into our modern theories.

The first divergence I would like to point out is Piaget's emphasis on the *role of action* in thought. The Tourmen paper, quite rightly, makes much of this, since it is a key part of Piaget's theories and it permeated many of his insights. As the paper points out, for Piaget, knowledge is actually *constructed* through action, and it is the back-and-forth movement between thought and action that leads to theory change. Development, then, is highly constrained and regulated by interaction with the environment, as the human learner acts to maintain equilibrium and resolve contradictions based on feedback from reality in response to one's actions.

Tourmen is certainly correct to point out that probabilistic models are at least *consistent* with a view of the human as actor. In particular, modern theories of causality that incorporate Bayesian nets allow for a crucial role that intervention – that is, acting on the world – plays in driving inferences. With that as the exception, most implementations of probabilistic models do not incorporate action as a key element. The models capture how a learner would update their beliefs after having observed some data in the world, but the default approach is to presume that the learner is a passive observer. At most, Bayesian models might conceive of a learner as an *active* learner who makes different inferences knowing that their data have been provided by a helpful teacher [Shafto, Goodman, & Frank, 2012]. Yet it would not be too difficult to expand the approach further, to incorporate a wider notion of action. For instance, there is some research focusing on how people select their own data and information to choose from [Hendrickson, Navarro, & Perfors, 2016; Hendrickson, Perfors, & Navarro, 2014; Markant & Gureckis, 2013]. One could build on this to determine how doing so changes the information communicated by those data (via different sampling assumptions).

Another divergence worth considering is the Piagetian emphasis on what a learner does when their hypothesis does not explain an element in the world. For Piaget, this kind of contradiction or "disturbing event" led to emotional perturbation and disequilibrium, which the learner dealt with by a process of denial, distortion, or modification of the "undesirable fact," followed by an attempted integration and adjustment. This process seems rather more strenuous and emotional than is captured by Bayesian models, which simply presume that a learner updates the probabilities associated with their existing hypotheses, and chooses a new one if the probabilities so indicate. Something like the Piagetian process also appears more correct as a depiction of what people do, at least if they are faced with radically divergent data that require a massive theory change. People don't just blithely recalculate probabilities and proceed; rather, there is a lot of flailing and emotional discord, and only sometimes a successful resolution.

At least, this is what it looks like. Appearances can be deceiving, but these considerations do suggest that learning may not always just be simple hypothesis testing and updating. One way to incorporate these insights into the existing Bayesian framework may be to realize that in addition to the process of hypothesis testing there is also a process of hypothesis *generation*, which is almost universally recognized as more difficult and fraught for the learner. Although there is some research into how people generate new hypotheses and how they decide when they need to do so [e.g., Dougherty, Thomas, & Lange, 2010] it is relatively little, and even less from within a Bayesian framework. Since development in particular may be a time of massive theory change, this avenue of research seems ripe for the picking.

The final Piagetian insight that I wish to highlight simply concerns how he viewed the child – including all of his theories and ideas throughout all of his life – as one interconnected system that builds on itself. The child might go through stages, but each successive stage can be understood as intertwined with the previous and subsequent one. It is true that almost no modern theorist would seriously argue against the idea that human cognition is all one system, interconnected across time as well as across subsystems. Nevertheless, the modern academic world effectively silos research into questions that can fit into single papers and topics small enough for one person to become expert on. Even the most interdisciplinary work usually means having people with expertise in multiple methods working on one small topic or system: a typical interdisciplinary project might involve a computational modeler, an expert in development, and a linguist work together to understand, say, the vocabulary spurt in language acquisition. It is generally not multiple people working on multiple systems trying to build a unified picture of the organism as a whole.

There are good reasons for the balkanization of the field into many small parts – ranging from the realities of research funding to the sheer fact that no one person or group of people can become expert on everything that is known about psychology, at least not in the same way one might have been able to in Piaget's day. So I am not seriously suggesting that we should all turn our minds to developing one theory of everything. That said, the main thing I am struck by in thinking about Piaget's work now, from the vantage point of a modern researcher, is how the sheer scope of his thinking allowed him to make connections and unify phenomena that look superficially very different. There was value in that, and my hope is that we remember from time to time to think in these terms too.

References

Baillargeon, R., & Graber, M. (1988). Evidence of a location memory in 8-month-old infants in a non-search AB task. *Developmental Psychology*, *24,* 502–511. doi:10.1037/0012–1649.24.4.502

Bonawitz, E.B., Shafto, P., Gweon, H., Goodman, N., Spelke, E., & Schulz, L.E. (2011). The double-edged sword of pedagogy: Teaching limits children's spontaneous exploration and discovery. *Cognition*, *120,* 322–330. doi:10.1016/j.cognition.2010.10.001

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

Buschbaum, D., Bridgers, S., Weisberg, D., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367,* 2202–2212. doi:10.1098/rstb.2012.0122

Carey, S. (2009). *The origin of concepts.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195367638.001.0001

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49,* 243–249. doi:10.1037/a0028278

Dougherty, M.R., Thomas, R.P., & Lange, N. (2010). Toward an integrative theory of hypothesis generation, probability judgment, and hypothesis testing. *Psychology of Learning and Motivation*, *52,* 299–342. doi:10.1016/S0079-7421(10)52008-5

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science*, *8,* 307–314. doi:10.1016/j.tics.2004.05.002

Goodman, N.D., Frank, M.C., Griffiths, T.L., Tenenbaum, J.B., Battaglia, P.W., & Hamrick, J.B. (2015). Relevant and robust: A response to Marcus and Davis (2013). *Psychological Science, 26,* 539–541. doi:10.1177/0956797614559544

Gopnik, A., Glymour, C., Sobel, D.M., Schulz, L.E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 3–32. doi:10.1037/0033-295X.111.1.3

Gopnik, A., & Wellman, H. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory. *Psychological Bulletin, 138,* 1085-1108. doi:10.1037/a0028044

Griffiths, T.L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin, 138,* 415–422. doi:10.1037/a0026884

Griffiths, T.L., Steyvers, M., & Tenenbaum, J.B. (2007). Topics in semantic representation. *Psychological Review*, *114,* 211–244. doi:10.1037/0033-295X.114.2.211

Griffiths, T., & Tenenbaum, J.B. (2009). Theory-based causal induction. *Psychological Review, 116,* 661–716. doi:10.1037/a0017201

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107,* 9066–9071. doi:10.1073/pnas.1003095107

Hendrickson, A., Navarro, D.J., & Perfors, A. (2016). Sensitivity to hypothesis size during information search. *Decision*, *3,* 62–80. doi:10.1037/dec0000039

Hendrickson, A., Perfors, A., & Navarro, D.J. (2014). Adaptive information source selection during hypothesis testing. In P. Bellow, M. Guarani, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 607–612). Québec, QC: Cognitive Science Society.

Kemp, C. (2012). Exploring the conceptual universe. *Psychological Review, 119,* 685–722. doi:10.1037/a0029347

Kemp, C., & Tenenbaum, J.B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105,* 10687–10692. doi:10.1073/pnas.0802631105

Landauer, T., Foltz, P.W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25,* 259–284. doi:10.1080/01638539809545028

Macmillan, N.A., & Creelman, C.D. (2004). *Detection theory: A user's guide*. Hove, UK: Psychology Press.

Marcus, G.F., & Davis, E. (2015). Still searching for principles: A response to Goodman et al. (2015). *Psychological Science*, *26,* 542–544. doi:10.1177/0956797614568433

Markant, D.B., & Gureckis, T.M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143,* 94–122. doi:10.1037/a0032108

Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115,* 39–57. doi:10.1037/0096-3445.115.1.39

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198524496.001.0001

Pearl, J. (2000). *Models, reasoning, and inference.* Cambridge, MA: Cambridge University Press.

Perfors, A., Tenenbaum, J.B., Griffiths, T.L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120,* 302–321. doi:10.1016/j.cognition.2010.11.015

Perfors, A., Tenenbaum, J.B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118,* 306–338. doi:10.1016/j.cognition.2010.11.001

Ratcliff, R., & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333–367. doi:10.1037/0033-295X.111.2.333

Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1.* Cambridge, MA: MIT Press.

Sanborn, A., Navarro, D.J., & Griffiths, T.L. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117,* 1144–1167. doi:10.1037/a0020511

Shafto, P., Eaves, B., Navarro, D.J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15,* 436–447. doi:10.1111/j.1467-7687.2012.01135.x

Shafto, P., Goodman, N.D., & Frank, M.C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7,* 341–351. doi:10.1177/1745691612448481

Tauber, S., Navarro, D.J., Perfors, A., & Steyvers, M. (under review). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory.

Tenenbaum, J.B., & Griffiths, T.L. (2001). Generalization, similarity, and Bayesian inference. *Behavioural and Brain Sciences, 24,* 629–640. doi:10.1017/S0140525X01000061

Tenenbaum, J.B., Kemp, C., Griffiths, T.L., & Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331,* 1279–1285. doi:10.1126/science.1192788

Xu, F., & Griffiths, T. L. (2011). Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition*, *120,* 299–301. doi:10.1016/j.cognition.2011.06.008