Abstract: We support Enlightenment Bayesianism's commitment to grounding Bayesian analysis in empirical details of psychological and neural mechanisms. Recent philosophical accounts of mechanistic science illuminate some of the challenges this approach faces. In particular, mechanistic decomposition of mechanisms into their component parts and operations gives rise to a notion of levels distinct from and more challenging to accommodate than Marr's.

We find attractive *Enlightenment Bayesianism*'s commitment to grounding Bayesian analysis in knowledge of the neural and psychological mechanisms underlying cognition. Our concern is with elucidating what the commitment to mechanism involves. While referring to a number of examples of mechanistic accounts in cognitive science and ways that Bayesians can integrate mechanistic analysis, Jones & Love (J&L) say little about the details of mechanistic explanation. In the last two decades, several philosophers of science have provided accounts of mechanistic explanation and mechanistic research as these have been practiced in biology (Bechtel & Abrahamsen 2005; Bechtel & Richardson 1993/2010; Machamer et al. 2000) and the cognitive sciences (Bechtel 2008; Craver 2007). Drawing on these can help illuminate some of the challenges of integrating mechanistic analysis into Bayesian accounts.

At the core of mechanistic science is the attempt to explain how a mechanism produces a phenomenon by decomposing it into its parts and operations and then recomposing the mechanism to show how parts and operations are organized, such that when the mechanism is situated in an appropriate environment, it generates the phenomenon. One of the best-developed examples in cognitive science is the decomposition of visual processing into a variety of brain regions, each of which is capable of processing different information from visual input. When organized together, they enable individuals to acquire information about the visible world. Decomposition can be performed iteratively by treating the parts of a given mechanism (e.g., V1) as themselves mechanisms and decomposing them into their parts and operations.

A hierarchical ordering in which parts are at a lower level than the mechanism is thus fundamental to a mechanistic perspective. This notion of levels is importantly different from that advanced by Marr (1982), to which J&L appeal, which does not make central the decomposition of a mechanism into its parts and operations. To illustrate the mechanistic conception of levels in terms of mathematical accounts, it is often valuable to provide a mathematical analysis of the phenomenon for which the mechanism is responsible. In such an account (e.g., the Haken-Kelso-Bunz [HKB] model of bimanual cordination described by Kelso 1995), the variables and parameters refer to characteristics of the mechanism as a whole and aspects of the environment with which the mechanism interacts. But to explain how such a mechanism functions one must identify the relevant parts and their operations. The functioning of these parts and operations may also require mathematical modeling (especially when the operations are nonlinear and the organization non-sequential; see Bechtel & Abrahamsen 2010). These models are at a lower level of organization and their parts and operations are characterized in a different vocabulary than that used to describe the phenomenon (as the objective is to show how the phenomenon is produced by the joint action of parts that alone cannot produce it).

We can now pose the question: At what level do Enlightenment Bayesian accounts operate? Do they, like Bayesian Fundamentalist accounts, operate at the level of the whole person, where the hypothesis space reflects people's actual beliefs? Beliefs are most naturally construed as doxastic states of the person that arise from the execution of various operations within the mind/brain. J&L's invocation of Gigerenzer's work on cognitive heuristics (e.g., Gigerenzer & Todd 1999) suggests this is a perspective they might embrace – the heuristics are inference strategies of agents and do not specify the operations that enable agents to execute the heuristics. The resulting Bayesian model may reflect but does not directly embody the results of decomposing the mind into the component operations that enable it to form beliefs.

Another possibility is that the Bayesian hypothesis space might directly incorporate details of the operations performed by components (e.g., brain regions identified in cognitive neuroscience research). Now an additional question arises – with respect to what environment is optimization evaluated? Since we are working a level down from the whole mechanism, one might think that the relevant environment is the internal environment of the local component (comprising other neural components). But this seems not to be the strategy in the research J&L cite (Beck et al. 2008; Wilder et al. 2009). Rather, optimization is still with respect to the task the agent performs. In Beck et al.'s account, a brain region (lateral intraparietal cortex: LIP) is presented as computing a Bayesian probability. This directly links the Bayesian account to parts of the mechanism, but if this approach is to be generalized, it requires that one find brain components that are computing Bayesian probabilities in each instance one applies a Bayesian analysis.

Although we find the prospect of integrating mechanistic and Bayesian approaches attractive, we are unclear how the results of mechanistic decomposition – which often leave the agent-level representations behind to explain how they are realized through a mechanism's parts and operations characterized in a different vocabulary than that which characterizes the agent's beliefs – are to be incorporated into a Bayesian account. We suspect that the most promising strategy is more indirect: Mechanistic research at lower levels of organization helps constrain the account of knowledge possessed by the agent, and Bayesian inference then applies to such agent-level representations.

A further challenge for understanding how mechanism fits into Bayesian analysis stems from the fact that Bayesian analyses are designed to elicit optimal hypotheses. As J&L note, mechanisms, especially when they evolve through descent with modification, are seldom optimal. What then is the point of integrating mechanistic accounts into normative Bayesian models? One possibility is that the normative accounts serve as discovery heuristics – mismatches between the normative model and cognitive agents' actual behavior motivate hypotheses as to features of the mechanism that account for their limitations. While this is plausible, we wonder about its advantages over investigating the nature of the mechanism more directly, by studying its current form or by examining how it evolved through a process of descent with modification. Often, understanding descent reveals how biological mechanisms have been kludged to perform a function satisfactorily but far from optimally.

What the Bayesian framework has contributed to understanding cognition: Causal learning as a case study

doi:10.1017/S0140525X1100032X

Keith J. Holyoak[a] and Hongjing Lu[a,b]

Departments of [a]Psychology and [b]Statistics, University of California, Los Angeles, CA 90095-1563.

holyoak@lifesci.ucla.edu **hongjing@ucla.edu**
http://www.reasoninglaboratory.dreamhosters.com
http://cvl.psych.ucla.edu/

Abstract: The field of causal learning and reasoning (largely overlooked in the target article) provides an illuminating case study of how the modern Bayesian framework has deepened theoretical understanding, resolved long-standing controversies, and guided development of new and more principled algorithmic models. This progress was guided in large part by the systematic formulation and empirical comparison of multiple alternative Bayesian models.

Jones & Love (J&L) raise the specter of Bayesian Fundamentalism sweeping through cognitive science, isolating it from algorithmic models and neuroscience, ushering in a Dark Ages dominated

by an unholy marriage of radical behaviorism with evolutionary "just so" stories. While we agree that a critical assessment of the Bayesian framework for cognition could be salutary, the target article suffers from a serious imbalance: long on speculation grounded in murky metaphors, short on discussion of actual applications of the Bayesian framework to modeling of cognitive processes. Our commentary aims to redress that imbalance.

The target article virtually ignores the topic of causal inference (citing only Griffiths & Tenenbaum 2009). This omission is odd, as causal inference is both a core cognitive process and one of the most prominent research areas in which modern Bayesian models have been applied. To quote a recent article by Holyoak and Cheng in *Annual Review of Psychology*, "The most important methodological advance in the past decade in psychological work on causal learning has been the introduction of Bayesian inference to causal inference. This began with the work of Griffiths & Tenenbaum (2005, 2009; Tenenbaum & Griffiths 2001; see also Waldmann & Martignon 1998)" (Holyoak & Cheng 2011, pp. 142–43). Here we recap how and why the Bayesian framework has had its impact.

Earlier, Pearl's (1988) concept of "causal Bayes nets" had inspired the hypothesis that people learn causal models (Waldmann & Holyoak 1992), and it had been argued that causal induction is fundamentally rational (the power PC [probabilistic contrast] theory of Cheng 1997). However, for about a quarter century, the view that people infer cause-effect relations from non-causal contingency data in a fundamentally rational fashion was pitted against a host of alternatives based either on heuristics and biases (e.g., Schustack & Sternberg 1981) or on associative learning models, most notably Rescorla and Wagner's (1972) learning rule (e.g., Shanks & Dickinson 1987). A decisive resolution of this debate proved to be elusive in part because none of the competing models provided a principled account of how *uncertainty* influences human causal judgments (Cheng & Holyoak 1995).

J&L assert that, "Taken as a psychological theory, the Bayesian framework does not have much to say" (sect. 2.2, para. 3). In fact, the Bayesian framework says that the assessment of causal strength should not be based simply on a point estimate, as had previously been assumed, but on a probability distribution that explicitly quantifies the uncertainty associated with the estimate. It also says that causal judgments should depend jointly on prior knowledge and the likelihoods of the observed data. Griffiths and Tenenbaum (2005) made the critical contribution of showing that different likelihood functions are derived from the different assumptions about cause-effect representations postulated by the power PC theory versus associative learning theory. Both theories can be formulated within a common Bayesian framework, with each being granted exactly the same basis for representing uncertainty about causal strength. Hence, a comparison of these two Bayesian models can help identify the fundamental representations underlying human causal inference.

A persistent complaint that J&L direct at Bayesian modeling is that, "Comparing multiple Bayesian models of the same task is rare" (target article, Abstract); "[i]t is extremely rare to find a comparison among alternative Bayesian models of the same task to determine which is most consistent with empirical data" (sect. 1, para. 6). One of J&L's concluding admonishments is that, "there are generally many Bayesian models of any task. . . . Comparison among alternative models would potentially reveal a great deal" (sect. 7, para. 2). But as the work of Griffiths and Tenenbaum (2005) exemplifies, a basis for comparison of multiple models is exactly what the Bayesian framework provided to the field of causal learning.

Lu et al. (2008b) carried the project a step further, implementing and testing a 2×2 design of Bayesian models of learning causal strength: the two likelihood functions crossed with two priors (uninformative vs. a preference for sparse and strong causes). When compared to human data, model comparisons established that human causal learning is better explained by the assumptions underlying the power PC theory, rather than by those underlying associative models. The sparse-and-strong prior accounted for subtle interactions involving generative and preventive causes that could not be explained by uninformative priors.

J&L acknowledge that, "An important argument in favor of rational over mechanistic modeling is that the proliferation of mechanistic modeling approaches over the past several decades has led to a state of disorganization" (sect. 4.1, para. 2). Perhaps no field better exemplified this state of affairs than causal learning, which had produced roughly 40 algorithmic models by a recent count (Hattori & Oaksford 2007). Almost all of these are non-normative, defined (following Perales & Shanks 2007) as not derived from a well-specified computational analysis of the goals of causal learning. Lu et al. (2008b) compared their Bayesian models to those which Perales and Shanks had tested in a large meta-analysis. The Bayesian extensions of the power PC theory (with zero or one parameter) accounted for up to 92% of the variance, performing at least as well as the most successful non-normative model (with four free parameters), and much better than the Rescorla-Wagner model (see also Griffiths & Tenenbaum 2009).

New Bayesian models of causal learning have thus built upon and significantly extended previous proposals (e.g., the power PC theory), and have in turn been extended to completely new areas. For example, the Bayesian power PC theory has been applied to analogical inferences based on a single example (Holyoak et al. 2010). Rather than blindly applying some single privileged Bayesian theory, alternative models have been systematically formulated and compared to human data. Rather than preempting algorithmic models, the advances in Bayesian modeling have inspired new algorithmic models of sequential causal learning, addressing phenomena related to learning curves and trial order (Daw et al. 2007; Kruschke 2006; Lu et al. 2008a). Efforts are under way to link computation-level theory with algorithmic and neuroscientific models. In short, rather than monolithic Bayesian Fundamentalism, normal science holds sway. Perhaps J&L will happily (if belatedly) acknowledge the past decade of work on causal learning as a shining example of "Bayesian Enlightenment."

Come down from the clouds: Grounding Bayesian insights in developmental and behavioral processes

doi:10.1017/S0140525X11000331

Gavin W. Jenkins, Larissa K. Samuelson, and John P. Spencer

Department of Psychology and Delta Center, University of Iowa, Iowa City, IA 52242-1407.

gavin-jenkins@uiowa.edu larissa-samuelson@uiowa.edu john-spencer@uiowa.edu

http://www.psychology.uiowa.edu/people/gavin_jenkins

http://www.psychology.uiowa.edu/people/larissa_samuelson

http://www.psychology.uiowa.edu/people/john_spencer

Abstract: According to Jones & Love (J&L), Bayesian theories are too often isolated from other theories and behavioral processes. Here, we highlight examples of two types of isolation from the field of word learning. Specifically, Bayesian theories ignore emergence, critical to development theory, and have not probed the behavioral details of several key phenomena, such as the "suspicious coincidence" effect.

A central failing of the "Bayesian Fundamentalist" perspective, as described by Jones & Love (J&L), is its isolation from other theoretical accounts and the rich tradition of empirical work in psychology. Bayesian fundamentalists examine phenomena exclusively at the computational level. This limits contact with other theoretical advances, diminishing the relevance and impact of Bayesian models. This also limits Bayesians' concern