Cognitive Psychology 103 (2018) 42–84

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Beyond Markov: Accounting for independence violations in causal reasoning[☆]

Bob Rehder[*]

Department of Psychology, New York University, United States

ABSTRACT

Although many theories of causal cognition are based on causal graphical models, a key property of such models—the independence relations stipulated by the Markov condition—is routinely violated by human reasoners. This article presents three new accounts of those independence violations, accounts that share the assumption that people's understanding of the correlational structure of data generated from a causal graph differs from that stipulated by causal graphical model framework. To distinguish these models, experiments assessed how people reason with causal graphs that are larger than those tested in previous studies. A traditional *common cause network* ($Y_1 \leftarrow X \rightarrow Y_2$) was extended so that the effects themselves had effects ($Z_1 \leftarrow Y_1 \leftarrow X \rightarrow Y_2 \rightarrow Z_2$). A traditional *common effect network* ($Y_1 \rightarrow X \leftarrow Y_2$) was extended so that the causes themselves had causes ($Z_1 \rightarrow Y_1 \rightarrow X \leftarrow Y_2 \leftarrow Z_2$). Subjects' inferences were most consistent with *the beta-Q model* in which consistent states of the world—those in which variables are either mostly all present or mostly all absent—are viewed as more probable than stipulated by the causal graphical model framework. Substantial variability in subjects' inferences was also observed, with the result that substantial minorities of subjects were best fit by one of the other models (the *dual prototype or* a *leaky gate* models). The discrepancy between normative and human causal cognition stipulated by these models is foundational in the sense that they locate the error not in people's causal *reasoning* but rather in their causal *representations*. As a result, they are applicable to any cognitive theory grounded in causal graphical models, including theories of analogy, learning, explanation, categorization, decision-making, and counterfactual reasoning. Preliminary evidence that independence violations indeed generalize to other judgment types is presented.

1. Introduction

The last 25 years has witnessed a dramatic increase in research investigating how causal knowledge influences cognition. This work has established that virtually every type of judgment is profoundly influenced by people's causal understanding of a situation. Moreover, theory in this area has been promoted by use of a formalism borrowed from artificial intelligence. This formalism—*causal graphical models*—is generally considered to be normative and thus serves as a standard against which people's causal judgments can be evaluated (Glymour, 1998; Jordan, 1999; Koller & Friedman, 2009; Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 2000). Causal graphical models are used as psychological models of causal reasoning (Ali, Chater, & Oaksford, 2011; Fernbach & Erb, 2013; Hayes, Hawkins, Newell, Pasqualino, & Rehder, 2014; Krynski & Tenenbaum, 2007; Meder, Mayrhofer, & Waldmann, 2014;

☆ I thank Zachary J. Davis and three anonymous reviewers for their comments on previous versions of this article.
* Address: Department of Psychology, 6 Washington Place, New York, NY 10003, United States.
E-mail address: bob.rehder@nyu.edu.

https://doi.org/10.1016/j.cogpsych.2018.01.003
Accepted 23 January 2018
0010-0285/

Oppenheimer, 2004; Rehder, 2014a), including how people predict the effects of intervening on a system (Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005) and draw analogies between content domains (Holyoak, Lee, & Lu, 2010; Lee & Holyoak, 2008). They are the dominant framework for theories of causal learning, regardless of whether the question is how learners induce causal strength (Cheng, 1997; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008; Waldmann, 2000; Waldmann & Holyoak, 1992), causal structure (Gopnik et al., 2004; Griffiths & Tenenbaum, 2005, 2009; Sobel, Tenenbaum, Gopnik, 2004), or learn by manipulating the system in question (Bramley, Lagnado, & Speekenbrink, 2014; Coenen, Rehder, & Gureckis, 2015; Gopnik et al., 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). In categorization research, causal graphical models serve as accounts of how people learn categories (Kemp, Shafto, & Tenenbaum, 2012; Lien & Cheng, 2000; Waldmann, Holyoak, Fratianne, 1995; Waldmann & Hagmayer, 2006), classify objects (Hayes & Rehder, 2012; Oppenheimer, Tenenbaum, & Krynski, 2013; Rehder, 2003a, 2003b, 2014b; Rehder & Kim, 2009, 2010), and generalize properties to other categories and category members (Hayes & Thompson, 2007; Kemp & Jern, 2014; Kemp & Tenenbaum, 2009; Kemp et al., 2012; Lassaline, 1996; Rehder, 2006, 2009, 2014b; Rehder & Burnett, 2005; Shafto, Kemp, Bonawitz, Coley, & Tenenbaum, 2008). There are accounts of how people assess counterfactual possibilities (Lucas & Kemp, 2015; Rips, 2010; Rips & Edwards, 2013), make decisions (Hagmayer & Meder, 2013; Hagmayer & Sloman, 2009), and generate explanations (Lombrozo, 2010) based on the causal graphical model framework.

This research has shown that people are sophisticated causal reasoners and that causal graphical models provide a good first order approximation of those abilities. Nevertheless, there is now considerable evidence that a defining feature of these models—the Markov condition—is routinely violated by human reasoners. The Markov condition defines situations under which variables in a causal network are conditionally or unconditionally independent—it allows those independence relations to be "read off" a causal graph. That reasoners fail to honor these independence constraints is important because it places upper limits on the fidelity with which any cognitive theory based on causal graphical models can reproduce human behavior. Said differently, *all* the theories of reasoning, learning, categorization, explanation, and decision-making cited above must be considered incomplete because all inherit the descriptive flaw of causal graphical models—the fact that people violate the Markov condition.

A number of explanations of independence violations in causal reasoning have been offered, most of which appeal to the possibility that subjects were reasoning with knowledge in addition to, or different than, that assumed by the experimenters. They thus rationalize the so-called violations by noting that they are no longer errors if subjects' full causal model of the situation is taken into account. However, whereas it is plausible that subjects in some studies indeed made use of extra-experimental knowledge, recent work using more advanced methodological techniques (e.g., full counterbalancing of materials, etc.) that control for this possibility has shown that independence violations persist nonetheless (e.g., Rehder, 2014a; Rottman & Hastie, 2016). Accordingly, the remainder of this article assumes that the preponderance of evidence now indicates that humans do not generally honor the Markov condition when drawing causal inferences. The General Discussion will return to this issue when it reviews the conditions under which prior knowledge does—and doesn't—serve as a satisfactory account of causal reasoning errors.

The field of causal cognition's use of the causal graphical model framework thus needs to evolve to accommodate the fact people violate the Markov condition. On one hand, because the evidence for those violations comes primarily from studies of causal reasoning, it is tempting to stipulate the presence of non-normative reasoning rules and strategies—*heuristics*—that are idiosyncratic to how people draw causal inferences. In contrast, this article asserts that the discrepancy between normative and human causal inferences arises for reasons that are more foundational. In particular, it asserts that the fault lies not in people's causal *reasoning* but rather their causal *representations*, namely, their beliefs about what a causal model of the world has to say about the data that one can expect to observe.

The following sections will first describe the constraints that the current causal graphical model formalism imposes on the pattern of inter-variable correlations implied by a causal graph and then propose new three models that embody alternative constraints. The *leaky gate model* is based on the intuition that information flows across nodes in a causal network even in situations where that flow is normatively blocked. The *dual prototype model* is based on the intuition that causal inferences reflect an expectation that the variables in a network are either all present or all absent. The *beta-Q model* favors prototypical states as well but is also biased against states involving a mixture of present and absent variables. I refer to the alternative causal representations specified by these models as foundational because they have implications not only for the conditional probability judgments that are the focus of this article but each of the causal-based tasks mentioned above. The General Discussion will touch upon preliminary evidence that Markov violations indeed manifest themselves in other causal-based tasks such as categorization and hypothesis testing.

This article has the following structure. The first section describes some prominent examples of violations of the Markov condition that have been documented in the literature. I then present the three new models. Because all three account for known violations of the Markov condition, I introduce new causal graphs for which their predictions differ. New experiments assessing how people reason with those graphs are then reported. Having established which of the three new models provides the best descriptive account of people's causal inferences, the article will close with a discussion of the underlying cognitive processes potentially responsible for the observed causal inferences. For example, it will show that those inferences can be characterized as arising from the sort of limited resource *sampling* methods that have become popular in the cognitive sciences.

2. Independence violations in causal reasoning

The causal networks in Figs. 1 and 2 provide two examples of how human reasoners violate the independence relations stipulated by causal graphic models. Although superficially similar, the different direction of the arrows in the two networks entails different patterns of statistical independence. Start with the *common cause network* in Fig. 1 in which X is a common cause of Y_1 and Y_2. On one hand, Y_1 and Y_2 are unconditionally dependent (e.g., one can reason from the state of Y_1 to the likely state of X and then to the likely