Cognitive Psychology 76 (2015) 1–29

Contents lists available at ScienceDirect

Cognitive Psychology

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

Identifying expectations about the strength of causal relationships

Saiwing Yeung [a,*], Thomas L. Griffiths [b]

[a] Institute of Education, Beijing Institute of Technology, China
[b] Department of Psychology, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Accepted 3 November 2014
Available online 15 December 2014

Keywords:
Causal reasoning
Bayesian models
Computational modeling
Iterated learning

ABSTRACT

When we try to identify causal relationships, how strong do we expect that relationship to be? Bayesian models of causal induction rely on assumptions regarding people's a priori beliefs about causal systems, with recent research focusing on people's expectations about the strength of causes. These expectations are expressed in terms of prior probability distributions. While proposals about the form of such prior distributions have been made previously, many different distributions are possible, making it difficult to test such proposals exhaustively. In Experiment 1 we used iterated learning—a method in which participants make inferences about data generated based on their own responses in previous trials—to estimate participants' prior beliefs about the strengths of causes. This method produced estimated prior distributions that were quite different from those previously proposed in the literature. Experiment 2 collected a large set of human judgments on the strength of causal relationships to be used as a benchmark for evaluating different models, using stimuli that cover a wider and more systematic set of contingencies than previous research. Using these judgments, we evaluated the predictions of various Bayesian models. The Bayesian model with priors estimated via iterated learning compared favorably against the others. Experiment 3 estimated participants' prior beliefs concerning different causal systems, revealing key similarities in their expectations across diverse scenarios.

* Corresponding author at: Institute of Education, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China.
E-mail address: saiwing.yeung@gmail.com (S. Yeung).

http://dx.doi.org/10.1016/j.cogpsych.2014.11.001

1. Introduction

Inferring the relationship between causes and effects is an important skill that people rely on every day in order to understand the structure of their environment. Psychological models of human causal induction have often focused on the role of associative learning—how people's judgments might be related to the number of instances of an effect occurring in the presence and absence of a cause (e.g., Cheng, 1997; Shanks, 1995; Ward & Jenkins, 1965). Recent work has explored how ideas from Bayesian statistics might help to explain people's intuitions, using causal graphical models to precisely define the problem of causal induction (Griffiths & Tenenbaum, 2005; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008) and to formalize the influence of prior knowledge (Griffiths & Tenenbaum, 2009).

A key part of Bayesian models of causal induction is the assumptions they make about the expectations people have about causal relationships. These assumptions are expressed in the form of *prior distributions* (often shortened to just *priors*), and reflect the expectations of the learners about causal systems prior to seeing any data. Bayesian models of human causal induction focus especially on the learners' prior beliefs about the *strength* of causal relationships, and two different specifications have been proposed thus far—using either a non-informative prior that makes no assumptions about the prior beliefs (Griffiths & Tenenbaum, 2005), or a prior based on theoretical assumptions about people's inductive biases (Lu et al., 2008). In this paper we present a new approach to identifying the expectations people have about causal systems, using the technique of *iterated learning* (Griffiths & Kalish, 2007; Griffiths, Christian, & Kalish, 2008). This technique allows us to estimate the values of unobservable cognitive constructs, such as people's priors, and produced predictions about human judgments that are more accurate than previous methods.

The plan for the rest of the paper is as follows. In the next section we summarize relevant previous work on Bayesian models of human causal induction. We then introduce the basic ideas behind iterated learning, and apply it in Experiment 1 to empirically estimate participants' prior beliefs about causal systems. Next we discussed a potential issue in previous work on human causal induction—models of causal induction are often evaluated based on small sets of stimuli that have certain specific characteristics. To address this issue, in Experiment 2 we collected a large benchmark data set against which the performance of different models can be better compared. We followed up with an investigation on prior beliefs about several different causal systems in Experiment 3. Finally we conclude with a discussion of the implications of these results for understanding causal induction as well as some possible future directions.

2. Bayesian models of human causal induction

Historically, there have been two major approaches to psychological theories of causal induction (Newsome, 2003). The mechanism-based approach focuses on understanding how knowledge about the causal mechanisms influences reasoning. Here learners attempt to discover a process in which causal power possessed by entities is transmitted to generate the event (Ahn, Kalish, Medin, & Gelman, 1995). In contrast, the covariation-based approach focuses on how people use the contingency about cause and effect to identify causal relationships (Cheng, 1997; Shanks, 1995). In particular, much psychological research on causal induction has focused on the problem of *elemental causal induction*: given a number of observations of two binary variables with a plausible causal relationship, how do people assess the relationship between them? For example, we can imagine a scientist studying the effect of a certain chemical on clovers. While most clovers have three leaflets, some have four, and the scientist wants to know whether the presence of the chemical has the power to increase the proportion of four-leaf clovers. By planting a number of clover plants and applies this chemical on some of them, he can use the resulted contingency data—number of clover plants exposed to or not exposed to the chemical, number of four-leaf clovers resulted in each case—to evaluate this potential causal relationship.

More recently, researchers have proposed various models based on Bayesian statistics. While these models rely on covariation data, they explicitly represent the learners' subjective beliefs about the