Cognitive Psychology 79 (2015) 102–133

Contents lists available at ScienceDirect

Cognitive Psychology

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

# Strategies to intervene on causal systems are adaptively selected

Anna Coenen *, Bob Rehder, Todd M. Gureckis

*Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, United States*

A R T I C L E   I N F O

*Article history:*
Accepted 18 February 2015
Available online 15 May 2015

*Keywords:*
Causal learning
Interventions
Hypothesis testing
Information gain
Self-directed learning

A B S T R A C T

How do people choose interventions to learn about causal systems? Here, we considered two possibilities. First, we test an information sampling model, *information gain*, which values interventions that can discriminate between a learner's hypotheses (i.e. possible causal structures). We compare this discriminatory model to a *positive testing strategy* that instead aims to confirm individual hypotheses. Experiment 1 shows that individual behavior is described best by a mixture of these two alternatives. In Experiment 2 we find that people are able to adaptively alter their behavior and adopt the discriminatory model more often after experiencing that the confirmatory strategy leads to a subjective performance decrement. In Experiment 3, time pressure leads to the opposite effect of inducing a change towards the simpler positive testing strategy. These findings suggest that there is no single strategy that describes how intervention decisions are made. Instead, people select strategies in an adaptive fashion that trades off their expected performance and cognitive effort.



## 1. Introduction

Causal knowledge underlies our intuitive grasp of physics ("Heat causes water to turn to steam."), technology ("This button causes it to go."), and helps us understand our fellow human beings

* Corresponding author.
*E-mail addresses:* anna.coenen@nyu.edu (A. Coenen), bob.rehder@nyu.edu (B. Rehder), todd.gureckis@nyu.edu (T.M. Gureckis).

http://dx.doi.org/10.1016/j.cogpsych.2015.02.004
0010-0285/

("Hunger causes her to be grumpy."). Often, the only way to find out about the causal structure of the world is by manipulating individual variables, and observing the effects of this manipulation. For example, banning sugary drinks can help decide whether they are a cause of diabetes. These decisions to manipulate a system are known as *interventions* (Pearl, 2000) and psychological research has recently explored how people use these interventions to learn (Bonawitz et al., 2010; Lagnado & Sloman, 2004; Schulz, Gopnik, & Glymour, 2007; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005).

Most research into how people make causal intervention decisions has implicitly sought to identify the *single* strategy that characterizes people's choices best across one or more experiments. For example, one proposal is that people search for information that can *discriminate* between possible hypotheses about causal structure, for instance by using an *information gain* (IG) strategy (Bramley, Lagnado, & Speekenbrink, 2014; Nelson, 2005; Shafto, Goodman, & Griffiths, 2014; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). Alternatively, in the broader hypothesis testing literature many studies argue that people seek information that yields positive evidence to *confirm* a single hypothesis, disregarding alternatives (e.g., Klayman & Ha, 1987; Nickerson, 1998; Wason, 1960). This mode of search is often referred to as the *positive test strategy* or PTS because it favors queries that are expected to yield a positive response ("yes", rather than "no") given a single hypothesis. A survey of the literature on information gathering during learning reveals forceful arguments for each of these alternatives (Gureckis & Markant, 2012; Nickerson, 1998) even though the division between these perspectives is not always precise (Navarro & Perfors, 2011; Oaksford & Chater, 1994).

The present paper begins from a slightly different perspective from this past work. In particular, we first ask if any single strategy model provides a plausible account of intervention-based causal learning. To that end, we describe a new hierarchical Bayesian method of identifying decision strategies during causal intervention learning. Using the model, we present evidence that individual participants adopt a *mixture* of strategies when learning through causal interventions (Experiment 1). Next, we ask if such mixtures are stable biases in the way people approach such tasks or if they change in response to environmental factors. Our second and third experiments show that strategy choice can change adaptively depending on the current task environment. Such adaptive adjustment of intervention-based strategies is unanticipated by single strategy models and suggests simple manipulations which might improve the quality of human reasoning.

### *1.1. Two perspectives on information gathering*

Efficient learning from causal interventions is ultimately a problem of information search. The learner must decide which intervention to perform in order to gain information about a system's causal structure. The following section describes two theories of how people make such decisions and how they relate to the task of causal intervention learning.

#### *1.1.1. Discriminatory: information gain*

The first strategy considered here is based on a rational analysis of the structure learning task (Anderson, 1990; Chater & Oaksford, 2008; Marr, 1982). According to this perspective, people should choose interventions that will be maximally useful for distinguishing alternative hypotheses.

To illustrate, consider playing the children's game "Guess Who?". In this game, one player adopts a secret identity (e.g., a fictional character or a celebrity). The job of the other players is to reveal this identity as quickly as possible by asking questions that can be answered with a "yes" or "no". The space of possible hypotheses (identities) is large in the beginning, but can be reduced by asking revealing questions. For example "Is the character male?" is a useful question because (assuming the learner expects a roughly even split of males and females) either answer will cut the number of identities in half. In contrast, very specific questions like "Does the character have pointy ears?" is a lot less informative, because the likelihood of "yes" is very low, and a "no" does not reduce the hypothesis space by much (most people do not have pointy ears). Similarly, a too general query like "Does this character have eyes?" will do little to narrow down the number of plausible hypotheses, because it is true of most.