Neuroscience and Biobehavioral Reviews 46 (2014) 30–43

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review

Bayesian modeling of flexible cognitive control

Jiefeng Jiang[a,b], Katherine Heller[a,c], Tobias Egner[a,b,*]

[a] Center for Cognitive Neuroscience, Duke University, United States
[b] Department of Psychology & Neuroscience, Duke University, United States
[c] Department of Statistical Science, Duke University, United States

ARTICLE INFO

Article history:
Received 28 June 2013
Received in revised form 31 March 2014
Accepted 3 June 2014
Available online 11 June 2014

Keywords:
Cognitive control
Bayesian models
Conflict
Congruency sequence effect
Conflict adaptation
Proportion congruency effect

ABSTRACT

"Cognitive control" describes endogenous guidance of behavior in situations where routine stimulus-response associations are suboptimal for achieving a desired goal. The computational and neural mechanisms underlying this capacity remain poorly understood. We examine recent advances stemming from the application of a Bayesian learner perspective that provides optimal prediction for control processes. In reviewing the application of Bayesian models to cognitive control, we note that an important limitation in current models is a lack of a plausible mechanism for the flexible adjustment of control over conflict levels changing at varying temporal scales. We then show that flexible cognitive control can be achieved by a Bayesian model with a volatility-driven learning mechanism that modulates dynamically the relative dependence on recent and remote experiences in its prediction of future control demand. We conclude that the emergent Bayesian perspective on computational mechanisms of cognitive control holds considerable promise, especially if future studies can identify neural substrates of the variables encoded by these models, and determine the nature (Bayesian or otherwise) of their neural implementation.

Contents

* Corresponding author at: Center for Cognitive Neuroscience, Duke University, LSRC Box 90999, Durham, NC 27708, United States. Tel.: +1 919 684 1049.
E-mail address: tobias.egner@duke.edu (T. Egner).

http://dx.doi.org/10.1016/j.neubiorev.2014.06.001

1. Cognitive control as statistical prediction

"Cognitive control" describes the ability to guide one's behavior in line with internal goals. A key characteristic of cognitive control is thought to be flexibility: control processes must be capable of dynamically adapting (both qualitatively and quantitatively) to ongoing changes in the environment. How this type of contextual regulation of control occurs (in the absence of an all-knowing *homunculus*) is a key question in current cognitive psychology and neuroscience research. In the present paper, we review recent attempts of modeling the "control of control", with a particular focus on the increasingly popular idea that prediction using Bayesian algorithms, which behave similar to reinforcement learning algorithms with varying learning rates (see below), may furnish a potent means for flexibly adapting control settings to contextual changes in the environment. In Section 1, we review two influential (non-Bayesian) models of cognitive control and highlight some limitations in their ability to adapt control to changing circumstances, specifically with respect to integrating contextual information across different time scales. We then suggest that Bayesian methods can achieve time-varying, self-adaptive integration of control-relevant contextual information. In Section 2, we review recent efforts to use Bayesian models to simulate various aspects of cognitive control. In Section 3, we outline a novel Bayesian model of conflict-control and demonstrate how it can account for various key behavioral phenomena. In Section 4, possible directions for future research regarding the application of Bayesian models to cognitive control are discussed.

1.1. Cognitive control as 'guided' information processing

In interacting with our environment, we transform sensory input into internal representations and select cognitive or motor actions based on these representations and our current goals. Given the fact that there is an enormous amount of sensory information and many possible actions available in contrast to only a few desired responses, appropriate action selection is a difficult task. To simplify this task, stimuli and actions that are frequently paired become mnemonically associated (e.g., via Hebbian learning) into stimulus-response (S-R) ensembles (or pathways) or more complex and extended action schemas (Norman and Shallice, 1986) that facilitate prompt reaction. Because much sensory information is processed in different pathways in parallel but only few actions can (or should) be taken simultaneously, stimulus representations and S-R pathways are believed to compete for being selected to drive behavior (Desimone and Duncan, 1995; Miller and Cohen, 2001; Norman and Shallice, 1986). The results of this competition are largely driven by the strength of associative pathways: stronger (i.e., more frequently activated) pathways are more likely to win the competition than weaker or novel ones. Once selected (and executed), the strength of a particular pathway may be reinforced or reduced depending on the assessment of how well the selected actions have fulfilled the organism's intended goals (Balleine and Dickinson, 1998).

This competition mechanism (or "contention scheduling", see Norman and Shallice, 1986) can generate appropriate behavior in many situations, but strong, stereotyped pathways can also result in suboptimal and even hazardous actions in some situations. For example, a US citizen's habitual driving on the right side of the road may have serious consequences when performed in the UK. In this case, a set of weaker or even novel associations (e.g., driving on the left side of the road) must be biased to win the competition in order to achieve the organism's goals. This "top-down" biasing of information processing to favor goal-directed stimuli and actions is the essence of cognitive control (e.g., Norman and Shallice, 1986; Botvinick et al., 2001; Miller and Cohen, 2001). In present-day neuroanatomical models, cognitive control is closely tied to the prefrontal cortex (PFC), which is proposed to harbor temporary representations of current goals, goal-relevant stimuli and strategies (Badre, 2008; Botvinick et al., 2001; Braver and Barch, 2002; Duncan, 2001; Fuster, 2008; Koechlin et al., 2003; Miller and Cohen, 2001; Norman and Shallice, 1986). To implement control, representations of goals, context and related methods (like rules) are thought to be actively maintained in the PFC, which sends biasing signals to posterior brain regions to guide the information flowing through the desired pathways and reach the selection of appropriate actions (e.g., Miller and Cohen, 2001).

In the laboratory, cognitive control is traditionally tested in interference (or "conflict") tasks such as the Stroop task (MacLeod, 1991), which entail conditions that require subjects to overcome a stronger habitual response in favor of a weaker (but correct) response. Consider, for instance, a variant of the Stroop task we employ in the empirical section of this paper (Section 3). This task requires a subject to respond to the gender of a face image, while ignoring a word label (either "male" or "female") that is overlaid on the image and which can be either congruent (e.g., "male" overlaid on a male face) or incongruent (e.g., "female" overlaid on a male face) with the face image (Egner et al., 2008). In order to arrive at the correct response during an incongruent trial, the subject has to overcome the highly automatic processing of the word-meaning in favor of categorizing the face's gender. Correct response selection on incongruent trials therefore requires the application of cognitive control in the PFC, strengthening the information flowing through the task-relevant processing pathway to win out over the task-irrelevant (though more habitual) one (Botvinick et al., 2001; Braver and Barch, 2002; Cohen et al., 1990). Accordingly, many neuroimaging studies of these types of tasks have documented higher activation in the PFC associated with higher conflict and control levels (Barch et al., 2001; Botvinick et al., 2004; Ridderinkhof et al., 2004), and modulated activity in brain regions related to processing task relevant- and irrelevant stimuli (Egner and Hirsch, 2005; King et al., 2010; Liu et al., 2004; Wittfoth et al., 2006).

One crucial question regarding this account, however, is how cognitive control itself is controlled. For example, when does