Cognition 172 (2018) 89–100

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Learning and transfer of working memory gating policies

Apoorva Bhandari[a,*], David Badre[a,b]

[a] Cognitive, Linguistics, & Psychological Science, Brown University, Providence, USA
[b] Brown Institute for Brain Sciences, Brown University, Providence, USA

ARTICLE INFO

Keywords:
Working memory
Gating
Cognitive control
Generalization
Transfer

ABSTRACT

Abstract knowledge about the tasks we encounter enables us to rapidly and flexibly adapt to novel task contexts. Previous research has focused primarily on abstract rules that leverage shared structure in stimulus-response (S-R) mappings as the basis of such task knowledge. Here we provide evidence that working memory (WM) gating policies – a type of control policy required for internal control of WM during a task – constitute a form of abstract task knowledge that can be transferred across contexts. In two experiments, we report specific evidence for the transfer of selective WM gating policies across changes of task context. We show that this transfer is not tied to shared structure in S-R mappings, but instead in the dynamic structure of the task. Collectively, our results highlight the importance of WM gating policies in particular, and control policies in general, as a key component of the task knowledge that supports flexible behavior and task generalization.

1. Introduction

Humans display remarkable cognitive flexibility in novel task environments (McClelland, 2009). Given only verbal instruction, we rapidly adapt to new tasks, often achieving asymptotic levels of performance within just a few trials (Ackerman, 1988; Bhandari & Duncan, 2014; Ruge & Wolfensteller, 2010; Wolfensteller & Ruge, 2011). Such rapid adaptation relies, in part, on abstract task knowledge transferred from prior experience with other tasks. Abstract task knowledge captures regularities in the space of task environments, and can thus speed up learning in the new environment by reducing the size of the learning problem (Botvinick, Niv, & Barto, 2009; Cole, Etzel, Zacks, Schneider, & Braver, 2011; Collins & Frank, 2013; Gershman & Niv, 2010).

What form does such abstract task knowledge take? The vast majority of prior studies seeking to address this question have focused on rules, or stimulus-response (S-R) mappings as the basis of task knowledge. In these frameworks, abstract rules generalize prior knowledge and thus constrain the (usually) very large space of stimulus-response-outcome contingencies afforded by a novel task environment (Badre, Kayser, & D'Esposito, 2010). Such rules can both be instructed (Cohen-Kdoshay & Meiran, 2007, 2009; Cole, Bagic, Kass, & Schneider, 2010; Meiran, Pereg, Kessler, Cole, & Braver, 2015; Ruge & Wolfensteller, 2010) or transferred from prior experiences (Cole et al., 2011; Collins & Frank, 2013) to rapidly enable successful behavior in novel environments.

The implementation of a task, however, requires more than just the knowledge of stimulus-response contingencies. Even the simplest everyday task environments have dynamical structure, with events unfolding in a specific order, and with specific timing (Radvansky & Zacks, 2014). To achieve task goals in a dynamic task environment, then, one must also learn an *internal control policy* or *task model* aligned to the task's dynamic structure for the moment-by-moment control of internal cognitive processing (Bhandari & Duncan, 2014; Duncan et al., 2008). Such implementational control policies are not typically communicated via instruction and must be discovered and implemented "on the fly", through task experience. In other words, a 'task-set' must incorporate knowledge about implementational control contingencies beyond those specified in stimulus-response mappings (Rogers & Monsell, 1995).

In this paper, we ask whether control policies are themselves a form of abstract task knowledge that, like rules, can be transferred to novel task contexts. Just like different real-world tasks often share stimulus-response-outcome contingencies, they also share other forms dynamic structure (Botvinick, Weinstein, Solway, & Barto, 2015; Schank & Abelson, 1977). Such shared structure affords an opportunity for generalization of internal control policies. Instead of learning new control policies from scratch, humans may build repertoires of internal control policies that are re-used in novel tasks.

We operationalize this question within the domain of working memory (WM) control – i.e. the selective use of working memory. WM control has been extensively analyzed within the *gating framework* (see Fig. 1), in which access to WM is controlled by a set of input and output gates (Chatham & Badre, 2015; O'Reilly and Frank, 2006; Todd, Niv, & Cohen, 2009). The contents of WM can be selectively updated by

* Corresponding author.
E-mail address: apoorva_bhandari@brown.edu (A. Bhandari).

https://doi.org/10.1016/j.cognition.2017.12.001
Received 24 May 2017; Received in revised form 29 November 2017; Accepted 2 December 2017

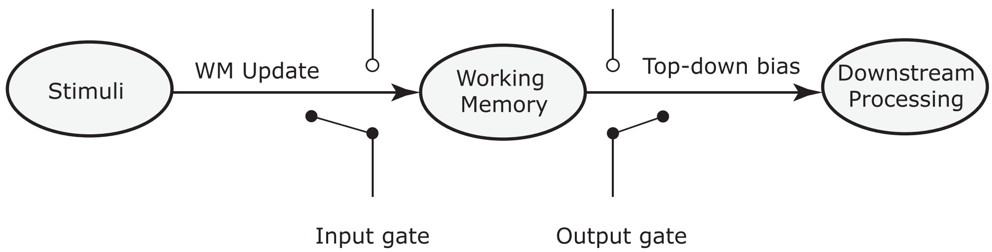

Fig. 1. Simple model of working memory control within the gating framework. Access to WM is controlled via the operation of an input gate that determines whether a stimulus is updated into WM. On the other side, an output gate controls whether or not information within WM can influence on behavior. Two broad classes of policies can be distinguished for selective use of WM – ones that achieve selection via selective input-gating, and ones that rely on selection via output-gating. Adapted from Hazy, Frank, and O'Reilly R (2007).

operating an input gate that determines whether stimulus information can enter WM. Similarly, operating a selective output gate allows WM to selectively influence downstream. Learning to perform a WM task, therefore, involves learning a *gating policy* for operating input and output gates in a moment-by-moment, task-appropriate manner (Frank & Badre, 2012). In the context of WM, a gating policy is an example of a control policy that must be aligned to the dynamic structure of the task. By learning such WM gating policies and transferring them across task contexts, humans may be able to exploit regularities in the dynamic structure of tasks.

To test this possibility, we adopt the 2nd order WM control task employed by Chatham, Frank, and Badre (2014). In their task, participants saw a sequence of three items on every trial, one of which specified a context. The context signaled which of the other two items in the sequence was the target item. Critically, there were two kinds of task structures – 'context first' (CF) trials, on which the first item in the sequence was the context item, and 'context last' (CL) trials, in which the last item in the sequence was the context item. CF and CL trials afford the use of different WM gating policies. On a CL trial, subjects had to employ a 'selective output-gating policy' that allowed the storage of both lower level items in WM (a non-selective input-gating operation), and the retrieval of the target item for guiding response selection (a selective output-gating operation). On a CF trial, while a similar selective output-gating policy could be employed, a more efficient 'selective input-gating policy' was possible. Such a policy would enable proactive coding of the contextual cue in WM, followed by selective input-gating of only the relevant lower-level item contingent on context. This allows a reduction in both, WM load, and interference from the competing non-target during response selection. Indeed, Chatham et al. (2014) presented evidence that CF and CL trials are treated differently and that well-trained subjects employ selective input-gating policies on CF trials to improve performance relative to CL trials on which the selective output-gating policy is required.

In the context of this WM control task, we ask whether selective gating policies learned in one task setting are transferred to a novel task setting. For instance, subjects exposed to an environment with only CL trials would learn a selective output-gating policy. Would this policy transfer to a new block with CF trials? In Experiment 1 we find a pattern of transfer effects that support the hypothesis that a previously learned gating policy influences initial behavior in a novel setting. We replicate these findings in Experiment 2. In addition, we provide evidence that transferred gating policies are dissociable from S-R mappings and have a much larger influence on subsequent behavior. We interpret these findings as evidence that internal control policies comprise an important form of structural task knowledge that supports behavior in novel situations.

2. Experiment 1

2.1. Methods

2.1.1. Participants

85 adult, right-handed participants (34 males, 51 females; age-range: 18–30, M = 21.4, SD = 2.7) from the Providence, RI area were recruited to take part in a computer-based behavioral experiment. We endeavored to collect between 18 and 20 participants in each of four groups based on approximate effect sizes suggested by pilot data. 1 participant was excluded for prior neurological injuries, 3 were excluded as they were on psychoactive medication. 5 participants were excluded because of low performance (< 70% accuracy) on the task. This left 76 participants (30 males, 46 females; age-range: 18–29, M = 21.3, SD = 2.6). Participants were randomly assigned to four groups of 19 each and there were no group differences in age or gender ratio ($p > .250$). We subsequently recruited another 38 participants to serve as additional control groups (17 males, 21 females; age-range: 18–30, M = 22, SD = 3.9). All participants had normal or corrected-to-normal vision, and no reported neurological or psychological disorders. All participants gave informed, written consent as approved by the Human Research Protections Office of Brown University, and they were compensated for their participation.

2.1.2. Apparatus

Experiments were conducted on a computer running the Mac OSX operating system. The stimulus delivery program was written in MATLAB 2013b, using the Psychophysics Toolbox. Responses were collected via a standard keyboard. All analyses were carried out in MATLAB 2013b and SPSS 22.

2.1.3. Task and experiment design

Participants were instructed to perform a 2nd order working memory control task (Fig. 2). On each trial, they saw a sequence of three items on the computer screen: a number (11 or 53), a letter (A or G), and a symbol (π or ⊙). The number served as a higher-level contextual cue, which specified the lower level items (letter or symbol) that would be the target on each trial. The relationships between the contextual cues and the lower level items were specified by the rule trees shown in Fig. 2(a). Participants were asked to memorize these rule trees before the task began, and were given an opportunity to review the trees for as long as they wanted at the beginning of each block.

To illustrate a trial, consider the sequence 11…A…π (Fig. 2(b)). As per the rule trees, 11 indicates that either of the letters, A or G, would be a target on that trial. Therefore, in the above example sequence, A, and not π, is the target. Each trial concluded with a response panel (presented at the same time as the last item in the sequence) that consisted of two pairs of items at the lower left or right of the screen.

Participants had to indicate whether the target item for that trial appeared as part of the left pair (left key) or the right pair (right key). In the example, the left key would be the correct response as it contained the 'A' target. The location of the letters and symbols in the target panel was randomized trial-to-trial such that all items appeared with equal frequency on the left and right across all trials. Further, 50% of the trials were 'incongruent' in that the lower-level items which appeared in the sequence were associated with different response keys in the response panel, and participants could not perform accurately without attending to the contextual cue. Participants were instructed to respond as fast as possible, while being accurate. Response panels were randomized so that on half the trials the left key was the correct response.

Apart from the position of the contextual item in the sequence