

Is adaptive control in language production mediated by learning?

Michael Freund¹ and Nazbanou Nozari^{1,2}

¹Department of Neurology, Johns Hopkins University. ²Department of Cognitive Science, Johns Hopkins University.

Corresponding author: Nazbanou Nozari, MD, PhD

1629 Thames Street, Suite 350, Baltimore, MD 21231, USA

Phone: 443-287-1712

Fax: 410-955-0188

Email: nozari@jhu.edu

Abstract

Recent work using the Picture Word Interference (PWI) paradigm has revealed that language production, similar to non-verbal tasks, shows a robust Congruency Sequence Effect (CSE), defined as a decreased congruency effect following incongruent trials. Although CSE is considered an index of adaptive control, its mechanism is debated. In two experiments, we tested the predictions of a learning model of adaptive control in production, using a task-switching paradigm fully balanced to evaluate CSE on a PWI trial as a function of the congruency of a 2-back PWI trial (within-task CSE), as well as a 1-back trial belonging to a different task (cross-task CSE). The second task was a visuospatial task with congruent and incongruent trials in Experiment 1, and a self-paced reading task with ambiguous and unambiguous sentences in Experiment 2 that imposed a gap between the two PWI trials twice as long of that in Experiment 1. A learning model posits that CSE is the result of changes to the connection weights between task-specific representations and a control center, which leads to two predictions in our paradigm: (a) a robust within-task CSE unaffected by the intervening trial and the gap duration, and (b) an absent or reversed cross-task CSE. These predictions were contrasted with two versions of an activation model of CSE. In accord with the predictions of the learning model, we found robust within-task CSE in PWI in both Experiments with a comparable effect size. Similarly, evidence of within-task CSE was also found in the visuospatial and sentence reading tasks. On the other hand, examination of cross-task CSE from PWI to the other tasks and vice versa revealed either absent or reversed CSE. Collectively, these results support a learning model of adaptive control in language production.

Keywords: word production; language monitoring; cognitive control; domain generality; congruency sequence effect (CSE); conflict adaptation, learning

Highlights:

- We tested predictions of learning versus an activation model of adaptive control.
- In two experiments, PWI trials were interleaved with trials from two other tasks.
- A persistent within-task adaptation was observed in PWI in both experiments.
- Adaptation between PWI and other tasks was either absent or reversed.
- Results support a learning model of adaptive control in language production.

Is adaptive control in language production mediated by learning?

1. Introduction

Cognitive control can be defined as operations required to resolve competition in favor of the most goal-appropriate response. The importance of cognitive control in language production has been implicitly acknowledged since early psycholinguistics research, in which it was shown that simultaneous activation of competing lexical representations was a natural product of spreading activation in a highly interconnected system (e.g., Dell, 1986). But compared to other cognitive areas, study of cognitive control in language production did not gain much attention until recently, perhaps partly due to the proposal of non-competitive accounts of lexical selection (e.g., Mahon et al., 2007). However, in recent years, a series of studies have demonstrated the true susceptibility of the language production system to interference during both semantic–lexical (e.g., Belke, Meyer, & Damian, 2005; Costa, Alario & Caramazza, 2005; Schnur et al., 2009; Schnur, 2014) and lexical–phonological mapping (Breining, Nozari, & Rapp, 2016; Nozari et al., 2016; O'Seaghda, & Marin, 2000; Sadat et al., 2014), inciting new interest in mechanisms that resolve such interference. Two sets of such studies have investigated such mechanisms: one set comprises studies that have reported a correlation between production performance (e.g., picture naming latencies or production errors) and performance on inhibitory control tasks (Shao et al., 2015; Shao, Meyer, & Roelofs, 2013; Shao, Roelofs, & Meyer, 2012; Trude & Nozari, 2017). The second set comprises lesion studies that have linked a cortical region such as the lateral prefrontal cortex, usually considered important for competition resolution, to performance on a production task that requires resolution of lexical competition (e.g., de Zubicaray et al., 2006; Piai, Riès, & Swick, 2016; Piai et al., 2013; Riès et al., 2015; see Nozari & Thompson-Schill, 2015, for a review).

While these studies have been critical in demonstrating the importance of cognitive control in language production, they mostly rely on indirect demonstrations of the need for, or the implementation of, control in production. It is thus difficult to establish a causal role between control and production abilities, or to understand how fluctuations in control demands lead to regulations of control in order to optimize performance. This regulatory process, called “adaptive control”, can be studied by investigating how performance on a demanding trial changes future performance on a trial with similar demands. Gratton, Coles, and Donchin (1992) were the first to report such a change, by showing a decrease in the size of the Flanker effect (defined as the difference in accuracy or response times between incongruent and congruent trials) after an incongruent compared to a congruent trial. The “Gratton effect” was later replicated in other tasks such as the Simon task (e.g., Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002), and the button-press Stroop task (e.g., Kerns et al., 2004), and received the more general label of “congruency sequence effect” (CSE) that we opt to use throughout this paper. CSE paradigms are currently the gold standard for examining adaptive control.

Recently, we and three other research groups have successfully replicated the canonical CSE pattern in word production using the Picture-Word Interference (PWI) paradigm (Duthoo et al., 2014; Freund, Gordon & Nozari, 2015; Shitova et al., 2017; Van Maanen & Van Rijn, 2010). In this paradigm, participants must name a picture with a word superimposed on it. On congruent trials, the word is the name of the picture. On incongruent trials, the word is a different name (often a semantic competitor of the picture; Schriefers, Meyer, & Levelt, 1990). The congruency effect manifests as lower accuracy and/or longer RTs on incongruent compared to congruent trials, and CSE is defined as a reduction of the congruency effect after incongruent trials. This finding is exciting because it demonstrates that word production in the presence of competitors

can be regulated online, just like non-linguistic tasks such as arrow Flanker and Simon tasks. More importantly, it opens a promising avenue for studying the nature of cognitive control processes that operate in language production. Our interest is in the mechanism by which CSE is generated in language production. Specifically, we test whether a learning account of CSE is suitable for implementing adaptive control in word production.

1.1. Accounts of CSE

Three classes of accounts—*associative*, *control-based*, and *hybrid*—have been proposed to explain CSE. Associative accounts view CSE as a consequence of forming specific associations between response choices and physical stimulus properties. The most prominent account is “feature integration” (Hommel, Proctor, & Vu, 2004), which proposes the binding of co-occurring features (e.g., blue hue + word “red” + evoking the left response button) as an “event file” in episodic memory. When any of the features are repeated on a subsequent trial, the entire memory is retrieved. Processing is facilitated if the new event has complete overlap with the previous one (i.e., if it is the exact same trial). Partial overlap between the previous and the new event, on the other hand, hinders performance, as the binding needs to be undone in order for the memory to be updated (Mayr, Awh, & Laurey, 2003). Associative accounts successfully explain CSE in tasks with a small stimulus or response sets, in which the probability of feature overlap from one trial to the next is high.

However, CSEs have also been observed in the absence of feature overlap and other low-level confounds. Memory confounds are usually controlled for by increasing the stimulus set size (often from two to four) to increase the number of unique combinations. This solution, however, creates a new problem in which the probability of each congruent stimulus is higher than each unique incongruent stimulus. For example, in order to maintain a 1:1 ratio of congruent and

incongruent trials in a four-choice Stroop task, the trial of “red” + red (i.e., “word” + hue) must happen four times more frequently than “red” + blue, “red” + green, and “red” + yellow. The emergence of CSEs under these circumstances may reflect the learning of these contingencies (Mordkoff, 2012; Schmidt & De Houwer, 2011). Importantly, though, when controlling for both memory confounds and contingency learning, CSEs have still been observed (Freitas & Clark, 2015; Hengstler, Holland, Steenbergen, & Knippenberg, 2014; Kunde & Wühr, 2006; Weissman et al., 2015; see Egner, 2014 for a review). Thus, associative accounts alone are insufficient to explain the CSE.

In contrast to associative accounts, control-based accounts propose the effect is driven by abstract control processes that operate independently from stimulus features. The common feature of control-based theories is their assumption that the CSE is driven by dynamic adjustments in top-down control, regardless of the specific nature of the representations involved in the task. The original control-based account of CSE is grounded in the modulation of expectations (Gratton et al., 1992), and assumes that individuals typically expect events to repeat in time. This expectancy-based account thus posits that encountering an incongruent trial generates an expectation (and ensuing preparation) for a subsequent incongruent trial, and it is this preparation that generates the CSE. A second influential control-based account, the conflict-based account, proposes that the adjustment of top-down control is mediated by monitoring the level of conflict on each trial (Botvinick et al., 2001). When conflict is high (i.e., in an incongruent trial), a signal is sent to recruit more control. This control, in turn, benefits performance on a subsequent high-conflict (incongruent) trial. The conflict-based account has had success in explaining a wide range of CSE, but has been criticized by Lamers and Roelofs (2011), who found larger CSE for post-congruent trials compared to both post-incongruent and

post-neutral trials. These authors argued that a conflict-based account would have predicted differences between the post-incongruent and post-neutral trials.

While there is abundant evidence for a CSE in the absence of low-level feature overlap or contingency learning (see above references), there is little doubt that feature overlap enhances the CSE (Hommel, 1998; Nieuwenhuis et al., 2006; Notebaert et al., 2006). This finding has prompted hybrid accounts of the CSE, which posit an interaction between top-down control mechanisms and bottom-up stimulus features. The most prominent hybrid account is the “adaptation-by-binding” model (Verguts & Notebaert, 2008, 2009), which proposes that conflict is used as a signal to increase the weights between stimulus features and attentional units that maintain task goals, through Hebbian learning. How abstract these stimulus features should be is less clear. While the adaptation-by-binding model is implemented on low-level features, Egner (2014) argues that such learning must include more abstract features. Critically, though, both Egner (2014) and Verguts and Notebaert (2008, 2009) view adaptive control as a learning mechanism.

1.2. Testing a learning account of CSE in word production

We use the general framework of Verguts and Notebaert’s (2008) adaptation-by-binding model to discuss the key predictions of a learning model of adaptive control. Many aspects of this model are not critical to our purpose, so instead of presenting the model in full detail, we will focus on the main mechanism and adapt it to our current purpose. The left panel of Figure 1 shows a schematic of this model. At a general level, the model includes two sets of representations: task-specific representations (e.g., color representations, orthographical representations), and task demand units, which implement top-down control over task-specific

representations in order to maintain task goals (e.g., “name the color/do not read the word” in a Stroop task). We use the label “Control center” instead of “task demand units” to cover a broader range of top-down control operations, such as resolving competition during lexical selection in PWI. Similarly, task-specific representations can be extended to include task-specific operations, such as mapping of semantic features to lexical items.

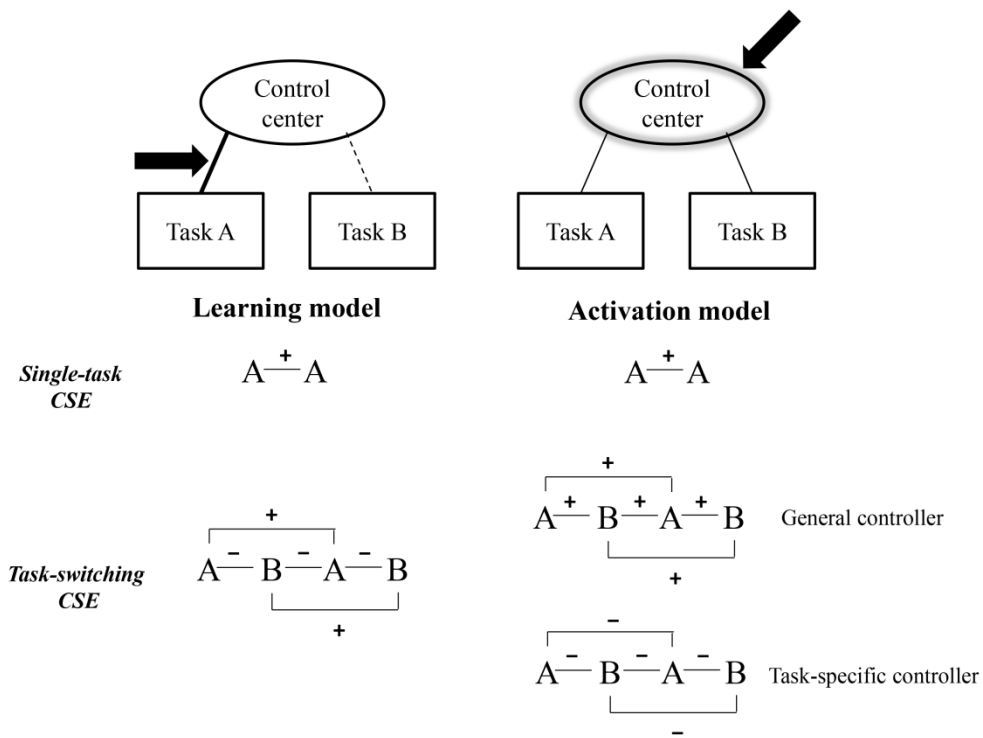


Figure 1. Schematic of the learning and activation models of adaptive control. The important difference between the two models is in the locus of the effect (black arrows), which is changes to the weights in the learning model and increased activation of the Control center in the activation model. Both models predict CSE between task A trials without an intervening task

(*single-task CSE*), but make different predictions about CSE during task switching (see text for predictions). Canonical CSE is marked by +. Absence of the canonical CSE is marked by –.

The black arrow indicates the locus of the CSE in the model: after each trial, the model “learns” the mapping between the Control center and the task-specific units involved on that trial by strengthening the connections between the two, in proportion to the amount of control required on that trial. For example, if the trial was associated with high levels of conflict, the connections will be strengthened by a larger factor compared to when the trial was associated with low levels of conflict. This will help facilitate performance on the following high-conflict trial within the same task—a *within-task CSE*. Because the change to the connection weights is stable and long-term, this model would predict within-task CSE either when the two trials from the same task are presented back-to-back (Figure 1, single-task CSE: CSE transfer from A to A), or when they are separated by the insertion of an intervening trial from another task (Figure 1, task-switching CSE: CSE transfer from A to A, as well as from B to B).

Importantly, only the connections between the Control center and the task-specific representations *involved in the current task* grow stronger. Connections between the Control center and task-specific representations of any other task either remain unchanged, or actually undergo a weakening referred to as the “dark side” of learning (Breining et al., 2016; under review; Oppenheim et al., 2010). This means that once top-down control has been diverted towards one task, implementing control in another task would either be unaffected (in the case where connections to the task-demand units for other tasks are unchanged), or would in fact suffer (in the case where such connections are weakened), manifesting as a *null or reversed*

cross-task CSE (Figure 1, Task-switching CSE: null or reversed CSE transfer from A to B and vice versa).

An alternative to the learning account is an activation account, a schematic of which is presented in the right panel of Figure 1. Similar to the learning account, the activation account assumes that upon registering a demand for control, neuronal populations in the Control center are activated and implement control over task-specific representations. But, the activation account differs critically from the learning account in that it assumes this process of top-down control does *not* involve any long-term changes to the connections between the Control center and task-specific representations. When trials from the same task are presented back-to-back, the increase in control demand on the current trial activates the Control center. This increased activation (or lowered threshold for reactivation) of the Control center makes for faster and more efficient implementation of control on a subsequent trial of the same task through priming (Figure 1, Single-task CSE: CSE transfer from A to A). Thus, in the case of single-task CSE, the activation account makes a comparable prediction to the learning account.

When participants alternate between two tasks, however, predictions of the activation account diverge from the learning account. This divergence is dependent on whether the Control center is viewed as a domain-general controller shared between tasks from different domains (e.g., Fedorenko, Duncan, & Kanwisher, 2013; Hsu & Novick, 2016; Kan et al., 2013; Novick et al., 2005, 2009; Nozari, Trueswell, & Thompson-Schill, 2016; Nozari & Thompson-Schill, 2015; Thompson-Schill et al., 1997), or as a constellation of predominantly task-specific sub-units or networks (e.g., Fedorenko, Behr, & Kanwisher, 2011; Jiang & Egner, 2014). While we remain agnostic about the generality of the controller, we discuss the prediction of each view in contrast with that of a learning model. The general controller view holds that the same neuronal

population in the Control center implements control in tasks A and B. Thus, once these neurons are activated by one task, they are primed to implement control in the other task. In other words, the same priming mechanism that leads to single-task CSE between two subsequent trials from task A also leads to cross-task CSE between subsequent trials from tasks A and B because the nature of the task does not affect the population of neurons engaged in implementing control (Figure 1, task-switching CSE: CSE transfer from A to B and vice versa in the general controller). Note that the constant engagement of the Control center by either task could create the illusion of a persistent within-task CSE, e.g., from the 2-back A to the current A trial, but in fact it is the intervening trial B that mediates this persistence. In contrast to the general-controller version, the task-specific controller view posits little to no overlap in the neuronal populations that implement control in tasks A and B. Thus, there should be no priming from task A to B or vice versa. Further, because activation is short-lived, priming from a 2-back A trial to the next A trial would be strongly diminished or absent if the two are separated by an intervening task B trial (Figure 1, task-switching CSE: absent CSE transfer from A to B, vice versa, and from A to A, or B to B in the task-specific controller).

To summarize, both the learning and the activation account predict within-task CSE in single-task experiments. Thus, the only way to distinguish between the two accounts is to employ a design where trials from two tasks A and B are interleaved. Under these circumstances, the learning account, which entails long-term changes to connections, predicts (a) a robust within-task CSE regardless of the duration of the intervening trial, and (b) a null or reversed cross-task CSE. Of the activation accounts, the general controller version would predict (a) a robust within-task CSE, and (b) a robust cross-task CSE. Finally, the task-specific controller version of the activation account would predict (a) a null within-task CSE or a CSE the

magnitude of which is very sensitive to the gap, and (b) a null cross-task CSE. In the next section, we review the empirical evidence for or against these predictions and propose a design to unambiguously distinguish between these theoretical alternatives.

1.2.1. Testing within-task CSE. The first way to distinguish between learning and activation accounts is to examine within-task CSE and its resilience against the passage of time. Studies of higher-order (i.e., more than 1-back) CSE are less common, and most have focused on the sensitivity of performance to the cumulative effect of demand on previous trials. For example, Clayson and Larson (2011) showed that both response latencies and ERP measures were sensitive to modulations of cognitive control across several trials with similar control demands (e.g., four consecutive incongruent trials; see also Thomas, Gonsalvez, & Johnstone; 2009). Similarly, Durston et al. (2003) showed that activation in several regions linked to conflict monitoring and control was decreased as a function of the number of preceding incongruent trials in the Flanker task (see also Clayson & Larson, 2011). Such cumulative effects are compatible with both learning and activation models, however, because both learning and priming mechanisms predict stronger effects with more repetitions. Two studies have specifically looked at the persistence of CSE against passage of time and have yielded different results. Egner, Ely, and Grinband (2010) used a face-word Stroop task in which participants had to categorize faces as male or female by pressing a button while ignoring the words “male” or “female” superimposed on the pictures. The gaps between trials were manipulated from 500 milliseconds (ms) to 7 seconds. CSE decreased quickly in magnitude and was no longer reliable past the 4–5 second interval. Wühr and Ansorge (2005) also found a decline in the size of CSE as a function of time, but unlike Egner et al. (2010), they showed robust CSE at 6 seconds in their Simon task.

No studies, to our knowledge, have examined the resilience of CSE to the passage of time in the language production system. Given the discrepancy between the results of the studies in the non-production literature, a test of the persistence of CSE in the production system would not only help shed light on the mechanism of control implementation in the language production system, but is also an important piece of evidence for or against learning models of CSE in general. To test this, we examined CSE in a paradigm where PWI trials were interleaved in an ABAB format with trials from a different task that also contained high and low control demand trials (Figure 2). We chose PWI for several reasons: (1) it involves all cognitive processes critical to word production, namely, conceptualization, lexical activation, lexical selection, phonological encoding, and articulation. (2) It is one of the simplest and most reliable paradigms to elicit competition in the production system, and is thus ideal for manipulation of control demands. (3) Although the superposition of a written word on a picture is not what speakers experience during everyday speech, interference in PWI has been localized to the level of lexical selection and not stimulus encoding (Shitova et al., 2017). Thus, for the purpose of studying control adjustments, PWI taps into the same processing level as natural word production.

In Experiment 1, task B was a non-linguistic spatial task, called the “prime-probe” task (Weissman et al., 2015). On each trial, participants were presented with a large prime arrow followed shortly by a smaller probe arrow. On congruent trials, the directions of the two arrows were the same. On incongruent trials, the arrows faced opposite directions. The large prime arrow elicits a response matching the direction of that arrow, but this response representation must be suppressed on trials where the probe arrow faces the opposite direction. This suppression requires cognitive control (Weissman et al., 2015). We chose the prime-probe task over other visuospatial tasks such as Flanker or the Necker Cube (Kan et al., 2013) for two

reasons: (1) the prime-probe task avoids many of the low-level confounds known to influence CSE (Weissman et al., 2015), and (2) because the competing stimuli do not appear simultaneously, competition is less likely to arise at the level of stimulus encoding and more likely to be engendered by two responses evoked separately by the two stimuli. This makes the locus of competition more comparable between PWI and prime-probe, as opposed to other visuospatial tasks, and gives cross-task CSE its best shot (see below).

Importantly, the duration of the prime-probe trial imposed a 3666 ms gap between the two PWI trials during which participants performed a task with representations from a domain different from that in the PWI task. Using this paradigm, we can assess within-task CSE by examining changes in performance on a current PWI trial as a function of control demands on a 2-back PWI trial. A learning account would predict a robust within-task CSE in PWI, the size of which should be no smaller than that obtained in a previous single-task PWI study with similar materials (13 ms in Freund et al., 2015; see also 24 ms in Duthoo et al. 2014 and 23 ms in Shitova et al., 2017, albeit with different materials and in a different language).

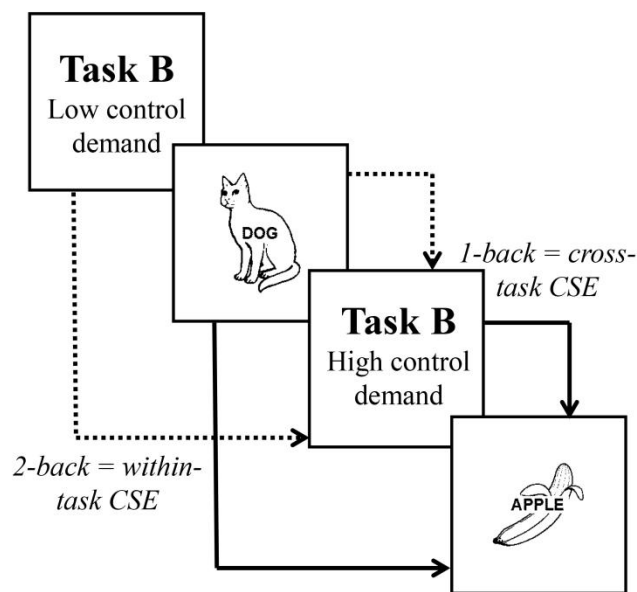


Figure 2. Schematic of the general design for Experiments 1 and 2. PWI trials are interleaved with trials from a different task—prime-probe in Experiment 1 and sentence reading in Experiment 2—in a predictable ABAB fashion. Prime-probe trials induced a gap of 3666 ms between the two PWI trials, and this gap was more than doubled in the sentence reading trials. Changes to performance on the current trial as a function of the previous trial in the same task (i.e., a 2-back trial) constitute within-task CSE (longer arrows on the left side of the figure). A learning model predicts robust within-task CSE despite the intervening trial (prediction 1). Changes to performance on the current trial as a function of the previous trial belonging to a different task (i.e., 1-back trial) constitute cross-task CSE. A learning model predicts no cross-task CSE or a reversed cross-task CSE.

Experiment 2 pushed the manipulation further by doubling the temporal gap between the two PWI trials and by using an intervening task with linguistic materials more likely to cause interference with the linguistic PWI materials than those of a spatial task (e.g., Shah, & Miyake, 1996). Participants performed self-paced reading of sentences that either did or did not contain local ambiguity (high and low control demand, respectively). Several studies have previously established a link between ambiguity resolution and cognitive control, especially in “garden path” sentences, in which the initial incorrect interpretation is the more common one (see Novick et al., 2005 for a review). For example, upon hearing “The primary suspect established the alibi had been a total lie” (Garnsey et al., 1997), most listeners initially parse “the alibi” as the direct object of the verb, only realizing the need for re-parsing when the rest of the sentence is heard. This re-parsing requires suppression of the initial, more common alternative, and has been shown

to require cognitive control (Hsu & Novick, 2016; Novick et al., 2005; 2009). The main reason for choosing the comprehension of ambiguous sentences in Experiment 2 was to use a task that relies upon representations with greater overlap with those involved in language production than the visuospatial task used in Experiment 1. This overlap may be important for the general-controller version of the activation accounts, if one assumes that the more similar representations are more likely to activate overlapping neuronal populations in the Control center. In other words, if domain-generality of the controller is determined by representational similarity between tasks, then Experiment 2 which uses linguistic representations in both tasks should give the general-controller version of the activation account its best shot. Prior studies (e.g., Kan et al., 2013; Hsu & Novick, 2016) have used syntactically ambiguous sentences. To be consistent, we also used syntactic ambiguity in half of the sentences. However, the need for control in PWI arises during lexical selection, a process that is distinct from syntactic operations (e.g., Ferreira & Slevc, 2007). Thus, we also constructed half of the materials in the sentence comprehension task to contain lexical (but not syntactic) ambiguity, in order to engender conflict more similar to that experienced during the PWI task. Again, this choice should give the general controller version of the activation model its best chance.

We examined the statistical robustness of the CSE in PWI as a function of the 2-back PWI trial after a gap of around 8 seconds, during which participants engaged in reading comprehension of sentences with low or high control demands. Moreover, we compared the size of CSE for the PWI trials in Experiments 1 and 2, which, according to a learning account of adaptive control, should be unchanged. The version of the activation model with a task-specific controller would, on the other hand, predict a sizeable reduction in CSE, and perhaps its total elimination at gaps as long as 8 seconds.

1.2.2. Testing cross-task CSE. The second prediction concerned the task-specificity of CSE. Empirical results in this regard are mixed. In keeping with the prediction of the learning model, CSE in task-switching paradigms—like the ones used in the current paper—has been shown to be limited to tasks that probe the same dimension. For example, Kiesel, Kunde, and Hoffmann (2006) found CSE when all trials required numerical parity judgement, but no cross-task CSE when trials alternated between parity and magnitude judgments. Similarly, Notebaert and Verguts (2008) showed the canonical CSE in a task-switching paradigm when the two tasks probed the same dimension (stimulus orientation), but found a reversed CSE when they probed different dimensions (color vs. orientation). Moreover, factorial combinations of two tasks, in which two types of conflict are merged into a single task (e.g., administering a manual Stroop task via lateralized stimuli and response buttons so that Simon conflict is orthogonally generated; Egner, 2008) have yielded CSEs specific to the source of conflict (Akçay & Hazeltine, 2011; Boy, Husain, & Sumner, 2010; Egner, Delano & Hirshc, 2007; Kim, Chung, & Kim, 2012; Kunde, Augst, & Kleinsorge, 2012; Kunde & Stöcker, 2002; Kunde & Wühr, 2006; Schlaghecken, Refaat, & Maylor, 2011; Wendt, Kluwe, & Peters, 2006).

In contrast to these results stand those which have found a reliable cross-task adaptation between tasks with different representations and task goals (Freitas, Bahar, Yang, & Banai, 2007; Hsu & Novick, 2016; Kan et al., 2013; Kleiman, Hassin, & Trope, 2014). For example, Kan et al. (2013) reported a reduced congruency effect in button-press Stroop after reading locally ambiguous sentences (Experiment 1), and after passively viewing a bi-stable Necker Cube (Experiment 2). In the same vein, Hsu and Novick (2016) reported more efficient processing of sentences with ambiguity following incongruent button-press Stroop trials.

To test our second prediction, we chose our tasks—a word production task, a visuospatial task, and a sentence comprehension task—to be very similar to those used in a previous study in which a reliable cross-task CSE was observed (Kan et al., 2013), in order to give cross-task CSE its best shot (see above for the reasoning behind our choice of each specific paradigm). Cross-task CSE was assessed by examining changes to the performance on a current trial, as a function of the control demand on the previous trial of the other task. A learning account would predict absent or reversed cross-task CSE. The version of the activation account with domain-general controller, on the other hand, would predict a robust canonical cross-task CSE.

2. Experiment 1

2.1. Methods

2.1.1. Participants. Thirty-two native English speakers recruited from the Johns Hopkins University community (24 women; mean age = 24.8 years) participated for payment. All participants gave informed consent under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine.

2.1.2. Materials. For the PWI task, a list of 128 target-distractor word pairs was compiled to form the incongruent PWI stimuli (see Appendix A). Each target and distractor were semantically related (e.g., “bus” and “car”), were matched in length and frequency, and had minimal phonological overlap. Next, 128 300-by-300 pixel black-and-white line drawings corresponding to each target word were selected from the IPNP corpus (Szekely et al., 2004) and from online images marked for reuse. The word (the distractor for incongruent and the target for congruent stimuli) was superimposed in the center of each image in black uppercase 36-point Helvetica, creating 128 congruent and 128 incongruent stimuli. Each PWI stimulus was assigned

once to each of our four cross-task CSE conditions (*cC*, *iC*, *iI*, *cI*; where lowercase “*c*” and “*i*” denote the congruent and incongruent status of the previous trial, respectively, and capital “*C*” and “*I*” the status of the current trial).

Prime-probe materials were adopted from Weissman et al. (2015; Experiment 1, "sequential" condition). Stimuli consisted of black prime arrows and black probe arrows (75% smaller than the prime arrows) pointing in one of the four cardinal directions, presented on a white background. Primes and probes pointed in the same direction on congruent trials and in opposite directions on incongruent trials. Similar to PWI, each prime-probe stimulus was assigned to each cross-task CSE condition an equal number of times.

Sequences of experimental trials were formed by interleaving PWI and prime-probe trials in an ABAB task-switching pattern (Figure 2). This predictable pattern minimized the behavioral cost of the task-switch while maximizing the number of switch trials. To minimize order effects, the materials were organized in four lists with different orders, assigned randomly to participants, so each participant in the study was presented with one of the four lists. A list contained four blocks of 256 experimental trials (128 PWI stimuli, 128 prime-probe stimuli), and two “filler” trials (one of each task) to begin the block and thus start the congruency sequence (These filler trials were not analyzed.) The order of blocks within each list was shuffled across participants according to a balanced Latin square. Additionally, the trials within each list obeyed several constraints: (a) Each block began with a PWI trial. (b) The four cross-task CSE conditions in both tasks occurred with equal frequency within each block. (c) All target responses in both tasks appeared with equal frequency in each cross-task CSE condition. (d) Each PWI target was presented once before each cross-task CSE condition of prime-probe. Similarly, each condition of prime-probe was presented once before each PWI target. This

constraint ensured that any item-specific effects of PWI on subsequent prime-probe (or vice versa) would be balanced between cross-task CSE conditions. (e) The higher-order congruency sequences in prime-probe were balanced. That is, each of the eight possible sequences of congruency across three trials (e.g., *ccC*) occurred with equal frequency. This constraint allowed us to analyze 2-back sequences for evidence of within-task CSE. (f) Semantically related PWI targets (as defined by category membership) were spaced by at least 12 unrelated PWI targets. This constraint minimized cumulative semantic interference (Schnur, 2014). (g) A maximum of six consecutive trials (three from each task) of the same congruency were allowed to occur.

2.1.3. Procedures. The experiment was run in E-Prime 2.0 software (Psychology Software Tools, Pittsburg, PA). Stimuli were displayed at the center of a 15-by-12 inch Dell monitor approximately 25 inches in front of the participants. RTs for PWI were registered using an Audio-Technica microphone connected to E-Prime's SR-BOX. Responses were also recorded digitally and transcribed offline for the identification of errors. RTs for the prime-probe were registered using a Dell keyboard.

First, participants silently reviewed a slideshow containing labeled images of all PWI targets in the experiment. Next, they received task instructions and completed three practice blocks. The first was a 10-trial PWI block, the second was a 48-trial prime-probe block, and the third was a 20-trial task-switching block. For PWI, participants were instructed to “name the picture as fast and accurately as possible.” For prime-probe, participants were instructed to respond to the direction of the probe arrow by pressing, with index and middle fingers of left and right hands, one of the four arrows on the keyboard corresponding to the correct direction. After practice, participants completed 4 experimental blocks of 258 trials.

PWI stimuli were presented for 3000 ms or until a response was registered. A blank screen then appeared for 1000 ms before a prime-probe trial. Primes and probes were each presented for 133 ms, separated by a blank screen presented for 33 ms. After probe presentation, a blank screen, the offset of which marked the response deadline, appeared for 1367 ms. The next PWI trial started after an inter-trial interval of 1000 ms. In total, this process induced a gap of 3666 ms between the two PWI trials.

2.1.4. Analyses. While we are particularly interested in CSE in PWI, the predictions of a general learning model of adaptive control should hold for any task. We thus analyze RTs and error rates in both PWI and the prime probe task. The data for all subsequent analyses are publicly available in Freund and Nozari (2018). The first set of analyses focuses on within-task (i.e., 2-back) CSE to test the first prediction of the learning model. The second set of analyses focuses on cross-task (i.e., 1-back) CSE to test the second prediction of the learning model.

Trials with inaccurate, incomplete or late (past the deadline) responses were excluded from RT analyses, as were the two filler trials initiating each block. All RTs were log-transformed prior to analysis in order to better approximate a Gaussian distribution. Data were analyzed using linear and generalized linear mixed-effect models with the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) in R v3.4.0. In models of RTs, p-values were estimated via Satterthwaite approximation. For analyzing errors, logistic versions of the models were used and p-values were calculated using lme4's default Wald z-test. Unless otherwise specified, all RT models were structured with maximal random effects, in line with recently proposed guidelines of model-building for psycholinguistic hypothesis testing (Barr, Levy, Scheepers, & Tily, 2013). Because of the relatively small number of errors, models of errors often did not tolerate full random effects, so these models are reported with random intercepts for participant and item.

Fixed effects in the models testing within-task CSE included current-trial congruency, 2-back trial congruency and, critically, the interaction between the two. Fixed effects in the models testing cross-task CSE included current-trial congruency, 1-back congruency, and the interaction between the two. In models that tolerated larger random effect structures, the slopes of all fixed effects on subjects and items were included in addition to subject and item random intercepts. Full results of all RT analyses (where the majority of the critical effects are), including the random effects, are reported in Appendix B. To save space, we only report the critical findings in the manuscript.

2.2. Results

2.2.1. Within-task CSE

Figure 3 shows the within-task CSE in RTs and error rates for PWI (upper panels) and prime-probe (lower panels) tasks. Recall that this is adaptation in each task as a function of a 2-back trial from the same task, ignoring the intervening trial from a different task. This effect is reflected in the interaction between current-trial congruency and 2-back congruency.

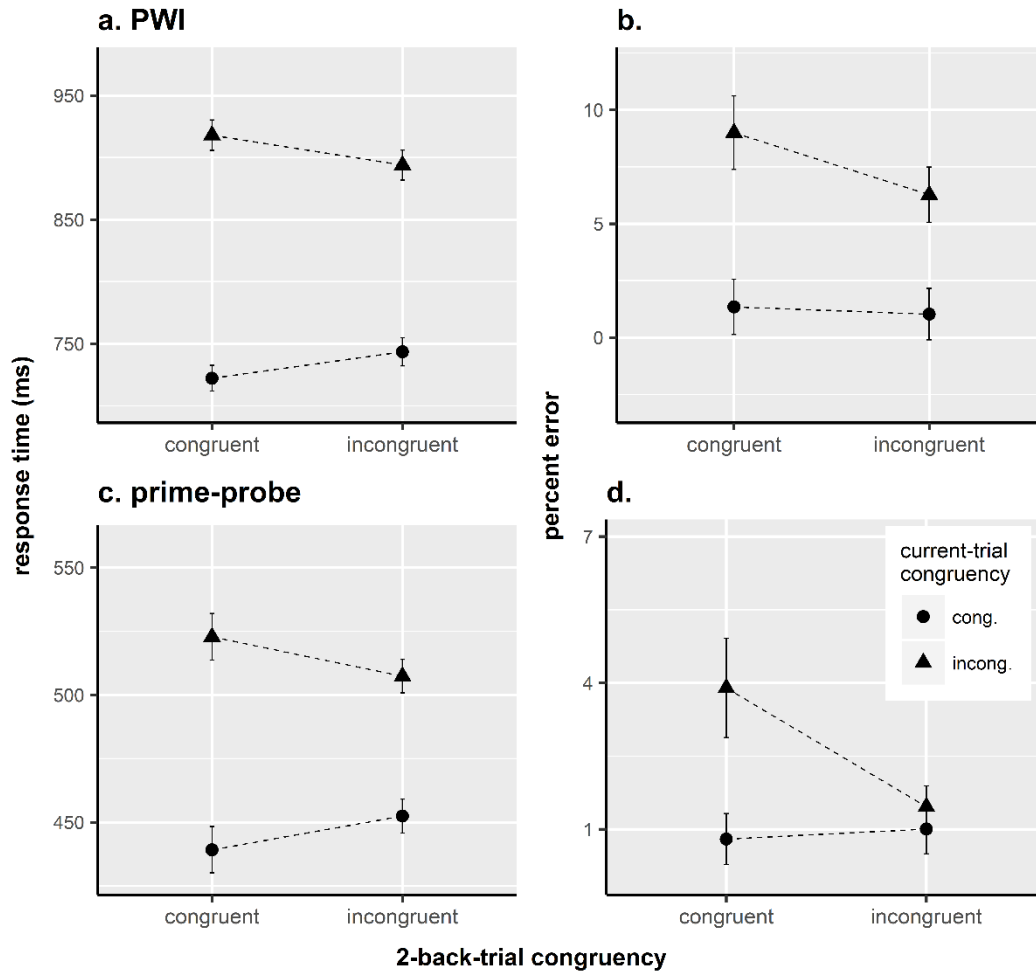


Figure 3. Within-task CSEs in Experiment 1. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 2-back PWI trial. Mean RT (c) and percent error (d) for congruent and incongruent prime-probe trials as a function of the congruency of a 2-back prime-probe trial. Error bars indicate 95% confidence intervals of within-participant variability (Morey, 2008).

2.2.1.1. PWI. Errors and trials in which the response was not accurately detected (e.g., early or late microphone triggers) constituted 4.7% and 2.3% of the data, respectively, and were removed for analysis. To control for potential error-related sequence effects (see Danielmeier & Ullsperger, 2011 for a review)—which can be confounded with CSE effects, as errors tend to occur more often on incongruent trials—we first assessed their occurrence in our data. RTs were

slower following previous prime-probe errors ($\beta = 0.04, t = 2.65, p = 0.01$) and following 2-back PWI errors ($\beta = 0.04, t = 2.94, p < 0.01$), and PWI errors were more frequent following 2-back PWI errors ($\beta = 0.62, z = 4.37, p < 0.01$). We therefore also excluded these types of trials from RT (4.7%) and error (1.8%) analyses, respectively (e.g., Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014). Finally, outliers, defined via visual inspection of QQ-plots (e.g., Schmidt & Weissman, 2015) of both log-transformed and raw RTs (after the aforementioned exclusions), were also removed from all analyses ($< 1\%$). Total exclusion percentages were 12.5% for RT and 6.8% for error analyses.

In PWI, RTs were significantly slower on incongruent relative to congruent trials ($\beta = 0.24, t = 21.26, p < 0.001$), and following an incongruent relative to a congruent 2-back (PWI) trial ($\beta = 0.04, t = 6.03, p < 0.001$). Critically, the interaction between current and 2-back congruency was also significant ($\beta = -0.06, t = -5.97, p < 0.001$; see Table B1 in Appendix B for the full results of this analysis), showing a marked reduction in the within-task congruency effect, with an average size of 45 ms (SE = 7 ms) across participants (see Table B1 for the model's estimated effect size represented by the coefficient). Post-hoc tests revealed a reliable 2-back effect on both current congruent ($\beta = 0.04, t = 5.94, p < 0.001$) and incongruent ($\beta = -0.02, t = -2.54, p = 0.01$) trials.

Error rates were significantly higher on incongruent relative to congruent trials ($\beta = 2.13, z = 13.03, p < 0.001$). But neither the main effect of 2-back congruency ($\beta = -0.20, z = -0.91, p = 0.36$), nor its interaction with current-trial congruency ($\beta = -0.10, z = -0.39, p = 0.70$) was statistically significant.

2.2.1.2. Prime-probe. Response errors and outliers accounted for 1.8% and $< 1\%$ of the data, respectively, and were excluded from the RT analyses. Preliminary analyses of error-related

sequence effects indicated that prime-probe RTs were slower following previous PWI errors ($\beta = 0.06, t = 3.8, p < 0.01$), previous PWI microphone trigger malfunctions ($\beta = 0.03, t = 4.5, p < 0.01$), and 2-back prime-probe errors ($\beta = 0.1, t = 4.9, p < 0.01$). These types of trials (8.7% of the data) were therefore excluded from RT analyses as well (e.g., Weissman et al., 2015). Correspondingly, prime-probe errors were more frequent following previous PWI errors ($\beta = 0.6, z = 2.85, p < 0.01$). This trial type (4.7%) was also excluded from error analyses. Total exclusion percentages were 10.2% and 4.7% of RT and error data, respectively.

In prime-probe, RTs were significantly slower on incongruent relative to congruent trials ($\beta = 0.18, t = 11.13, p < 0.001$), and following an incongruent relative to a congruent 2-back (prime-probe) trial ($\beta = 0.03, t = 4.27, p < 0.01$). Critically, the interaction between current and 2-back congruency was also significant ($\beta = -0.06, t = -5.69, p < 0.001$; see Table B2 in Appendix B for the full results of this analysis), showing the canonical CSE in spite of the intervening task, similar to PWI. Post-hoc tests revealed a significant 2-back effect on both current congruent ($\beta = 0.031, t = 4.94, p < 0.01$) and incongruent ($\beta = -0.031, t = -3.49, p = 0.03$) trials.

Error rates were significantly higher on incongruent relative to congruent trials ($\beta = 1.67, z = 8.50, p < 0.001$). Although there was no main effect of 2-back congruency on error rates ($\beta = 0.24, z = 1.01, p = 0.31$), the interaction between current-trial congruency and 2-back congruency was significant ($\beta = -1.28, z = -4.50, p < 0.001$), also suggesting a persistent canonical CSE despite the intervening PWI trial. Post-hoc tests showed that this effect was significant for incongruent ($\beta = -1.03, z = -6.66, p < 0.001$), but not congruent ($\beta = 0.27, z = 1.19, p = 0.24$) trials.

2.2.2. Cross-Task CSE

Figure 4 shows the cross-task CSE in RTs and error rates for PWI (upper panels) and prime-probe (lower panels) tasks. Recall that this is adaptation in each task as a function of the control demand on the previous trial, which belongs to the other task—that is, adaptation in PWI as a function of demand change in prime-probe and vice versa. This is reflected in the interaction between current-trial congruency and previous-trial congruency.

2.2.2.1. PWI. The same trials excluded in the previous analyses were also excluded here. The RT analysis returned a significant main effect of current-trial congruency ($\beta = 0.21, t = 18.78, p < 0.001$), but neither the main effect of 1-back congruency ($\beta = 0.001, t = 0.03, p = 0.98$), nor its interaction with the current-trial congruency ($\beta = 0.01, t = 0.75, p = 0.46$) was significant (see Table B3 in Appendix B for the full results of this analysis).

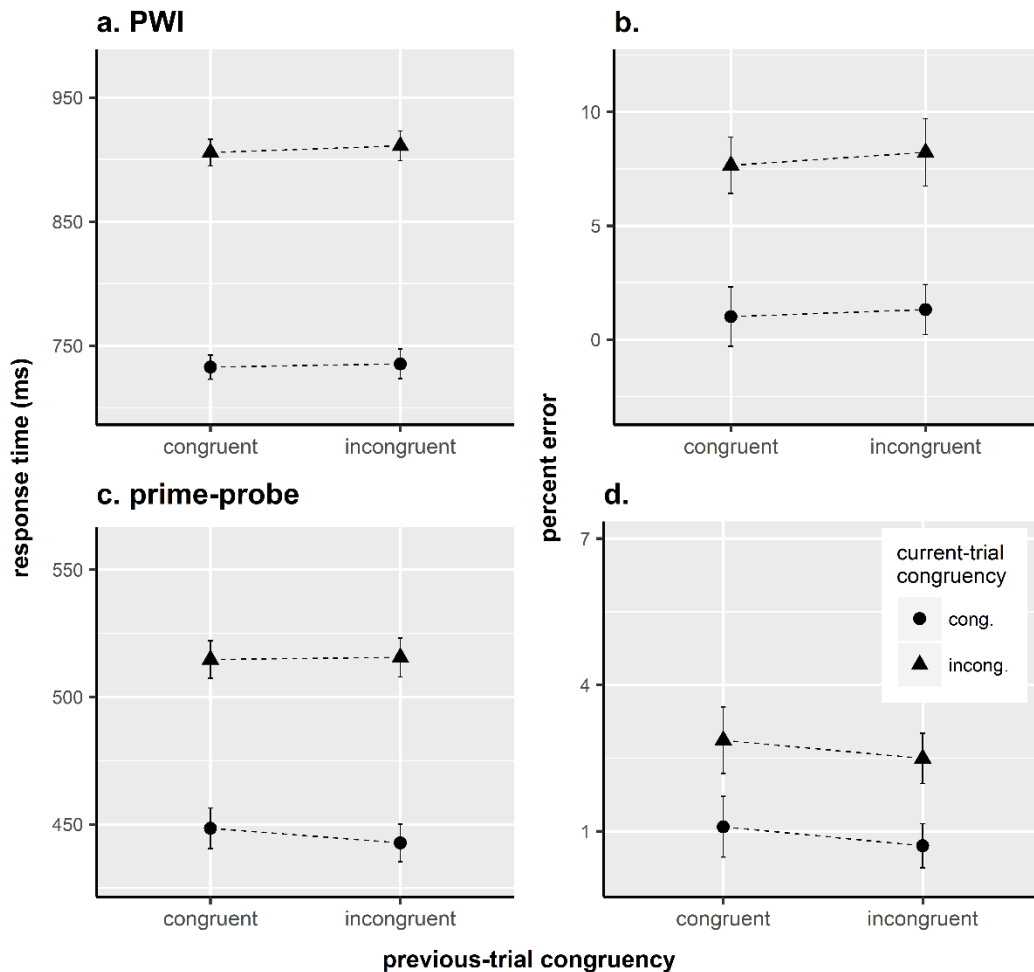


Figure 4. Cross-task CSEs in Experiment 1. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 1-back prime-probe trial. Mean RT (c) and percent error (d) for congruent and incongruent prime-probe trials as a function of the congruency of a 1-back PWI trial. Error bars indicate 95% confidence intervals of within-participant variability.

Analysis of errors also revealed a significantly larger number of errors on incongruent relative to congruent trials ($\beta = 2.22$, $z = 12.85$, $p < 0.001$), while neither the main effect of 1-back congruency ($\beta = 0.25$, $z = 1.18$, $p = 0.24$) nor its interaction with the current-trial congruency was significant ($\beta = -0.16$, $z = -0.69$, $p = 0.49$).

2.2.2.2. Prime-probe. RTs were significantly slower on incongruent relative to congruent trials ($\beta = 0.14$, $t = 10.40$, $p < 0.001$) and faster on trials following incongruent relative to congruent PWI ($\beta = -0.01$, $t = -2.87$, $p = 0.03$). The interaction of these effects was also marginally significant ($\beta = 0.01$, $t = 2.33$, $p = 0.05$), but note that the direction of the effect is the opposite of canonical CSE, i.e., the congruency effect grew *larger* after an incongruent trial, thus showing a reversed adaptation effect.

Prime-probe errors were more common on incongruent trials relative to congruent ($\beta = 1.01$, $z = 5.72$, $p < 0.001$). Following incongruent relative to congruent PWI, there was a marginally significant reduction in prime-probe error rate ($\beta = -0.46$, $z = -1.88$, $p = 0.06$). The interaction between 1-back and current-trial congruency, however, was not significant ($\beta = 0.31$, $z = 1.09$, $p = 0.28$).

2.3. Discussion

To test the first prediction, namely the resilience of CSE, we investigated within-task (i.e., 2-back) CSE in PWI and the prime-probe tasks. Results showed robust CSE in PWI RTs as

a function of control demands in a previous PWI trial, despite an intervening prime-probe trial. Importantly, the average size of CSE was 45 ms, which was no smaller than previous PWI studies in which there was no intervening task between the two PWI trials (13 ms in Freund et al., 2015 with similar materials and 23, and 24 ms in Shitova et al., 2017; and Duthoo et al., 2014 with different materials). A robust CSE was also found for the prime-probe trials, in which both RTs and errors showed the canonical CSE as a function of the congruency of a 2-back prime-probe trial without interference from the intervening PWI trial. Contrary to the version of the activation account with a task-specific controller, and in keeping with the first prediction of the learning model, these findings suggest that within-task CSE shows some degree of resilience against passage of time, at least when that time is spent on performing a task with representations from a different domain.

To test the second prediction, we examined cross-task CSE. In PWI, neither the RTs nor the error rates showed any evidence of adaptation as a function of control demand fluctuations on a previous prime-probe trial (see also Kiesel, Kunde, & Hoffmann, 2006). Note that the statistically significant 2-back CSE helps rule out alternative explanations for the null cross-task CSE observed in PWI: if statistical power had been an issue, e.g., 2-back CSE should also have been affected, and perhaps even more so given the potential noise induced by an intervening trial. Similarly, if switching between two tasks had caused interference (Braem, Abrahamse, Duthoo, & Notebaert, 2014) or had imposed a cognitive load that eliminated cross-task CSE, finding a 2-back CSE which involves two switches should have been even more difficult. Finally, RTs in the prime-probe task showed a marginally significant *reversed* adaptation effect (see also Notebaert & Verguts, 2008), further ruling out the possibility that we may have missed a canonical cross-task CSE that had in fact been present in these data. We thus conclude that the

results of Experiment 1 are well aligned with the predictions of the learning account, and incompatible with those of the task-specific-controller version of the activation account (in its prediction regarding within-task CSE) and those of the general-controller version of the activation account (in its prediction regarding cross-task CSE).

Experiment 2 set out to test the same predictions under more extreme conditions, that is, when the temporal gap was longer, and when the two tasks tapped into representations from the same (linguistic) domain, but with different goals. Kan et al. (2013) found a more robust CSE between Stroop and the sentence comprehension task, as opposed to Stroop and the Necker cube task. One possible explanation is that in the former case both tasks tapped into materials from the same domain. In Experiment 2, we use a sentence reading task similar to Kan et al. (2013) to increase the chance of uncovering a cross-task CSE.

3. Experiment 2

3.1. Methods

3.1.1. Participants. Thirty-two native English speakers, none of whom participated in Experiment 1, were recruited from the Johns Hopkins University community (23 women; mean age = 21.23 years). All gave informed consent under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine.

3.1.2. Materials. Comprehension materials consisted of 256 sentences: 64 semantically ambiguous, 64 syntactically ambiguous, and 128 unambiguous controls (see Appendix C). Sentences with semantic ambiguity (e.g., a1) were either created (from biased homographs in Nelson, McEvoy, Walling, & Wheeler, 1980) or adapted (from biased sentences in Duffy, Morris, & Rayner, 1988). Sentences with syntactic ambiguity (e.g., b1) were sampled from eight different sentence structures, each of which engendered a certain type of temporary syntactic

ambiguity (see Appendix C for complete list). Control (unambiguous) sentences were created by replacing the biased homograph with a monosemic word (as in a2), or by applying minimal revisions to the structure to remove temporal ambiguity (as in b2).

Semantic ambiguity: (a1) Reggie enjoyed his first *date* so much that he ate seven more. where “date” is used in its less frequent sense (a dried fruit) which only becomes apparent when the reader encounters “ate”.

Control for semantic ambiguity: (a2) Reggie enjoyed his first *raisin* so much that he ate seven more.

Syntactic ambiguity: (b1) The alley mice run rampant in is damp and dimly lit.

(Grodner, Gibson, & Tunstall, 2002)

where “alley” is likely to be parsed as a modifier for “mice” until the reader encounters the verb “is” which is incongruent with the parsed syntactic structures, prompting a revision.

Control for syntactic ambiguity: (b2) The alley *which* mice run rampant in is damp and dimly lit. (Grodner, Gibson, & Tunstall, 2002)

Additionally, for each pair of sentences, we defined (or imported, from the sourced experiment) the region of disambiguation. A spillover window was also marked as the region immediately following the region of disambiguation, as reading time effects in sentence comprehension tasks, especially those elicited within moving-window procedures, may emerge late (e.g., Ferreira & Henderson, 1990).

Due to the constraints imposed by time and number of sentences in the comprehension task in Experiment 2, a subset of PWI items (N = 32) were randomly selected from items of Experiment 1 that had picture-name agreement values above 0.9 as indexed by norms from IPNP or Amazon’s Mechanical Turk (see Appendix A). List construction and trial ordering followed

the same rules as Experiment 1. Importantly, each trial order was structured with balanced numbers of each higher-order congruency sequence for each task (e.g., *ccC*). Furthermore, each PWI target appeared with equal frequency in each participant's cross-task CSE conditions, minimizing any differences between conditions due to item properties. Each list contained four blocks of 64 experimental trials (interleaved PWI and sentence reading in an ABAB pattern). Also included within each block were five filler sentences followed by comprehension questions to ensure that participants paid attention to the meaning of the sentences they read. Each filler question was also followed by a filler PWI trial, thus restarting the congruency sequence for the following trials. In contrast to Experiment 1, we did not use additional “filler” stimuli to begin each block, but instead ensured that the trials that did begin each block—to be discarded prior to analyses—came from each experimental condition, e.g., *ccC*, equally.

3.1.3. Procedures. Procedures were kept as similar as possible to Experiment 1. After the PWI practice sessions, a block of 10 practice sentences was administered, followed by a task-switching practice block of 24 trials. Sentences were read via the self-paced moving-window procedure (Just, Carpenter, & Woolley, 1982), in which each character of a sentence is initially masked by a hyphen. Participants then pressed the spacebar to make each word appear sequentially, and the previous word was again masked. Participants were warned about the comprehension questions, in order to encourage them to try to comprehend what they were reading. All four blocks of the experiment were then completed with short (< 5 minute) intervening breaks.

PWI trials were presented with timing parameters equivalent to Experiment 1. Each sentence-reading trial began with a 1000 ms blank screen, followed by a 500 ms central fixation, then by the masked sentence. Text was presented in vertically centered, left-justified, 14 point

Consolas. After the sentence was read, a blank screen appeared for 1000 ms, followed by a 500 ms fixation cross before the PWI trial was presented. The average number of words in all experimental and control sentences (excluding fillers) was 13 (SD = 2.5) words. Assuming, on average, 400 ms reading time for each word, we estimated average reading time of 5200 ms per sentence, which, combined with the blank and fixation screens, was expected to impose an average gap of ~ 8200 ms between the two PWI trials—more than twice as long of a gap as in Experiment 1. For trials that were followed by comprehension questions, participants' responses were registered, and the next trial began after a 1500 ms inter-trial interval.

3.2. Results

In addition to RT and error data from PWI, we collected word reading durations from the comprehension task. Before testing for CSE, we must first demonstrate the effectiveness of the ambiguity manipulation in changing control demands on the sentence reading task. If the ambiguity manipulation has indeed increased control demands, this must be reflected in increased reading times in the critical region(s) in the sentence. To analyze these data, we used a two-step regression technique (adapted from Hofmeister, 2011), in which an initial model containing nuisance variables (e.g., word length) was fit to all reading time data—that is, experimental and filler items (to improve estimates of these variables)—and the residuals of this initial model were then analyzed for effects of interest. These residual reading times can be understood as the variance in reading times unaccounted for by the nuisance predictors. We opted for this two-step technique as opposed to adding the nuisance variables to a model of the raw reading time data to avoid non-convergence due to fitting over-specified models. For our CSE analyses, trial exclusion and model building followed the same criteria as in Experiment 1.

3.2.1. Ambiguity effects

Reading times longer than 2500 ms were removed. None of the participants performed below chance on the comprehension questions (mean accuracy = 77%) and none had a mean reading time further than 2.5 standard deviations from the group mean, thus no participant was excluded. A multilevel model was fit to these data. This model contained the following fixed effects, as recommended by Hofmeister (2011): (a) word length (in number of characters), (b) log-transformed position of the trial within the experimental block, (c) position of the block within the experiment, and (d) “subtype” of ambiguity (e.g., prepositional phrase attachment ambiguity, reduced relative clause ambiguity, etc.; see Appendix C), and a random effect for participants. Factor (d) was included to capture variability in reading times unique to the sentence structures that engendered each specific type of ambiguity (cf. the “construction type” factor in Hofmeister, 2011; each level of this factor, as might be expected, included both ambiguous and unambiguous sentences). Unlike Hofmeister (2011), however, we did not include word position within the sentence, as our sentences had varying lengths.

We then extracted the residuals of this model to estimate the effects of ambiguity. First, we removed data from filler sentences. Then, for sentences with disambiguating and spillover regions that were multiple words long, we took the mean residual log reading time—yielding, for each subject, two “composite” residual log reading times for every sentence they saw—one value for the disambiguating region and one for spillover. For single-word regions, the residual values were unchanged. As a final step, we removed values 3 standard deviations away from the grand mean (2% of the data). These “composite” residuals were then analyzed for evidence of ambiguity effects. The final ambiguity models contained a fixed effect for ambiguity (ambiguous or unambiguous) and random effects for participant and item (with random intercepts, and

random slopes for ambiguity). Figure 5 shows the results of this analysis in the disambiguating (left panel) and spillover (right panel) regions for each ambiguity type.

Relative to unambiguous sentences, residual log reading times in the disambiguating region were significantly slower in syntactically ambiguous sentences ($\beta = 0.05$, $t = 3.5$, $p < 0.01$), but not in semantically ambiguous sentences ($\beta = 0.02$, $t = 1.41$, $p = 0.17$). However, significant ambiguity effects emerged in the spillover regions of both types of sentences (syntactic: $\beta = 0.04$, $t = 3.31$, $p < 0.01$; semantic: $\beta = 0.04$, $t = 3.16$, $p < 0.01$). Further, these ambiguity effects were independent of the reading times of previous words in the sentence, as each (significant) effect remained significant regardless of whether the residual reading times of the previous one, two, or three windows were included in the regression as fixed effects. We were therefore assured that our ambiguity manipulations were effective. The regions in which a significant increase in control demand was observed (i.e., the disambiguation region for syntactic, and the spillover region for both sentence types) were used to analyze CSE in the sentence reading task.

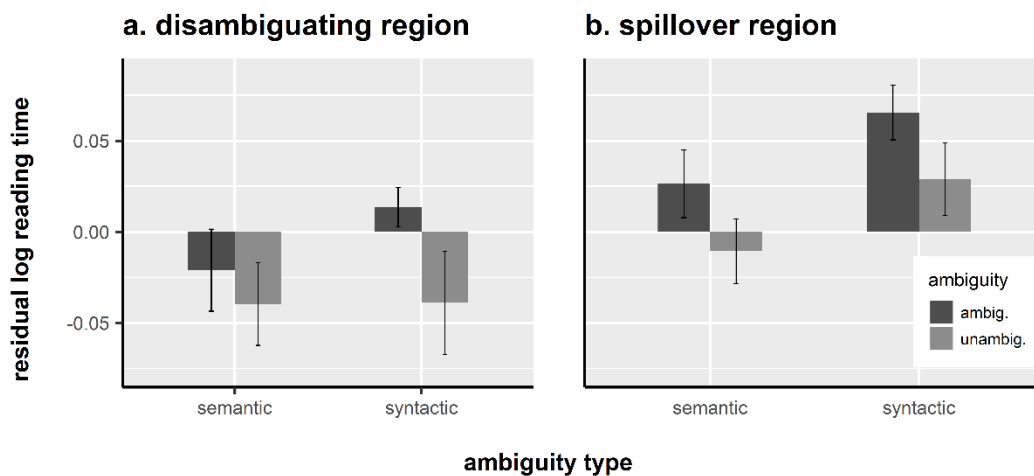


Figure 5. Ambiguity effects in Experiment 2. Mean residual log reading time for the disambiguating (a) and “spillover” (b) regions of semantically and syntactically ambiguous sentences. Error bars indicate 95% confidence intervals of within-participant variability.

3.2.2. Within-task CSE

Figures 6 and 7 show the within-task CSE for PWI and reading comprehension, respectively. Data for—or following—semantically and syntactically ambiguous sentences and their controls have been graphed separately to show the consistency of the pattern across the two ambiguity types. For the analyses, they are combined to increase statistical power. The average reading time for the experimental and control sentences was 5484 (SD = 2132) ms. Combined with the 3000 ms of blank and fixation screens, the total gap between the two PWI sentences was on average 8484 ms, more than twice as long as the gap between the two PWI trials in Experiment 1, as planned.

3.2.2.1. PWI. Response errors and outliers accounted for 3.6% and < 1% of the data and were excluded. Preliminary analyses of error-related sequence effects indicated that PWI errors were more frequent following 2-back PWI errors ($\beta = 1.50$, $z = 2.82$, $p < 0.003$). This type of trial (2.1%) was therefore excluded from error analyses (e.g., Weissman et al., 2015). Total exclusion percentages were 3.6% and 3.7% of RT and error data, respectively.

RTs were reliably slower on incongruent relative to congruent PWI trials ($\beta = 0.18$, $t = 9.74$, $p < 0.001$), and following an incongruent relative to a congruent 2-back PWI trial ($\beta = 0.05$, $t = 4.20$, $p < 0.001$). Similar to Experiment 1, the interaction between current and 2-back congruency was also significant ($\beta = -0.06$, $t = -5.05$, $p < 0.001$; see Table B5 in Appendix B for the full results of this analysis). This interaction supported a reliable canonical CSE with an average size of 54 ms (SE = 6 ms) across participants (see Table B5 for the model’s estimated

effect size represented by the coefficient). Post-hoc tests revealed a reliable effect of 2-back congruency on both congruent ($\beta = 0.05$, $t = 4.45$, $p < 0.01$) and incongruent ($\beta = -0.02$, $t = -2.30$, $p = 0.02$) trials.

To formally test whether the magnitude of the CSE obtained in Experiment 2 (54 ms) differed from that of Experiment 1 (45 ms), the data of the two experiments were combined and analyzed in a model with current-trial congruency, two-back congruency, Experiment, and the 2- and 3-way interactions of these effects, as well as a full random effect structure. Results revealed significant effects of current-trial congruency ($\beta = 0.24$, $t = 19.66$, $p < 0.001$), two-back congruency ($\beta = 0.30$, $t = 5.67$, $p < 0.01$), and the interaction between the two (i.e., canonical CSE; $\beta = -0.05$, $t = -6.57$, $p < 0.01$). Importantly, however, this interaction did not interact with Experiment (i.e., the 3-way interaction of current-trial congruency, two-back congruency, and Experiment was not significant; $\beta = -0.01$, $t = -1.03$, $p = 0.30$).

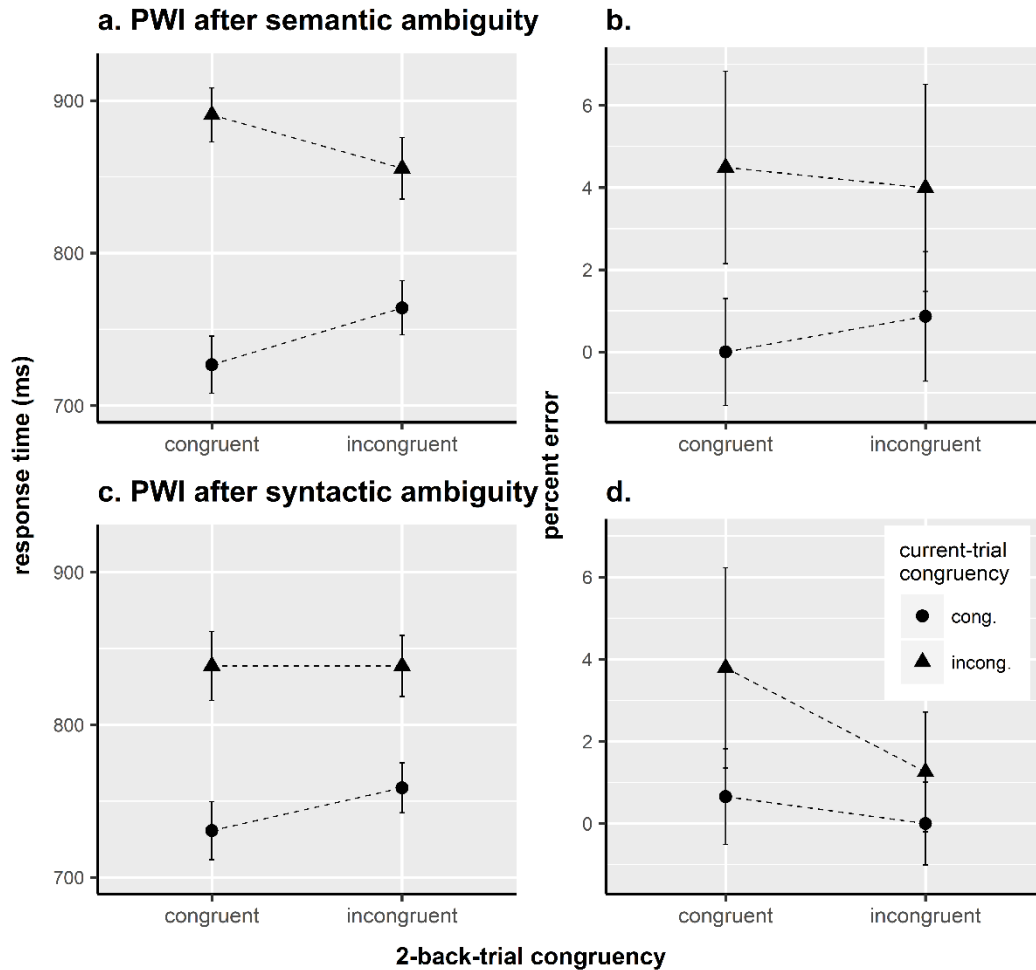


Figure 6. Within-task CSEs in PWI in Experiment 2. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 2-back PWI trial with an intervening semantically ambiguous sentence or its control. Mean RT (c) and percent error (d) for congruent and incongruent PWI trials as trials as a function of the congruency of a 2-back PWI trial with an intervening syntactically ambiguous sentence or its control. Error bars indicate 95% confidence intervals of within-participant variability. Note the consistency of the pattern of results across the four graphs in the direction predicted by the canonical CSE.

Error rates were significantly higher on incongruent relative to congruent trials ($\beta = 2.88$, $z = 4.67$, $p < 0.01$). But neither the main effect of 2-back congruency ($\beta = 0.49$, $z = 0.63$, $p =$

0.53), nor its interaction with current-trial congruency ($\beta = -0.94$, $z = -1.14$, $p = 0.26$) reached significance.

3.2.2.2. Sentence reading. The residual log reading times used in our ambiguity analysis served as the dependent variable for the CSE analyses. Preliminary analyses of error-related sequence effects indicated that residual log reading times were larger following PWI errors ($\beta = 0.1$, $t = 2.52$, $p = 0.02$), therefore, we excluded this type of trial (1.9%) from further analyses (e.g., Weissman et al., 2015).

As seen in Figure 7, both semantic and syntactic sentences show the canonical CSE pattern as a function of ambiguity on a 2-back trial. A model of residual log reading times did not converge with a full random effect structure, so the random effects were reduced to fit the model (see Table B6 in Appendix B). Results showed significantly slower reading times in ambiguous compared to unambiguous sentences ($\beta = 0.06$, $t = 6.04$, $p < 0.001$). Reading times were also significantly slower after a 2-back ambiguous trial ($\beta = 0.02$, $t = 2.13$, $p = 0.03$). There was also a significant interaction between the current-trial ambiguity and 2-back ambiguity ($\beta = -0.03$, $t = -2.06$, $p = 0.04$). Post-hoc tests revealed a reliable effect on the unambiguous ($\beta = 0.02$, $t = 2.14$, $p = 0.04$) but not on the ambiguous ($\beta = -0.01$, $t = -0.74$, $p = 0.47$) trials.

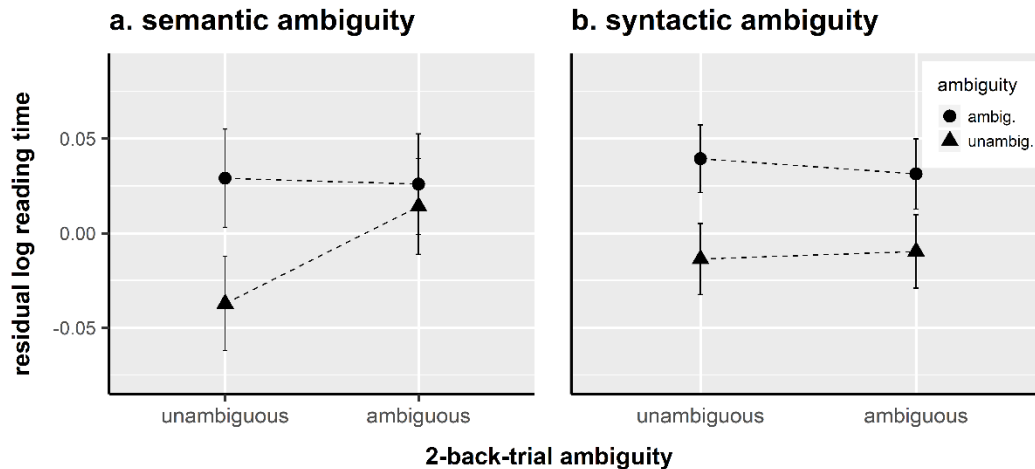


Figure 7. Within-task CSEs in sentence reading in Experiment 2. Mean residual log reading times for semantically ambiguous sentences and their controls as a function of the ambiguity of a 2-back sentence with an intervening PWI trial (a). Mean residual log reading times for syntactically ambiguous sentences and their controls as a function of the ambiguity of a 2-back sentence with an intervening PWI trial (b). Error bars indicate 95% confidence intervals of within-participant variability. Note the consistency of the pattern of results across the two sentence types in the direction predicted by the canonical CSE.

3.2.3. Cross-task CSE

Figures 8 and 9 show the cross-task CSE for PWI and reading tasks, respectively. Data for—or following—semantically and syntactically ambiguous sentences and their controls have been graphed separately to show the consistency of the pattern across these the two ambiguity types. For analyses, they are combined to increase statistical power.

3.2.3.1. PWI. The same trials excluded in the previous analyses were also excluded here. The RT analysis showed that current incongruent trials were significantly slower than the congruent ones ($\beta = 0.15$, $t = 8.78$, $p < 0.001$), and that there was general slowing after an

ambiguous sentence ($\beta = 0.04$, $t = 3.19$, $p = 0.003$). We found no evidence, however, that ambiguity led to cross-task CSE: PWI congruency did not interact with previous-trial ambiguity ($\beta = -0.01$, $t = -1.06$, $p = 0.30$; see Table B7 in Appendix B for the full results of this analysis).

Error analysis showed that participants made more errors on incongruent trials ($\beta = 2.42$, $z = 3.88$, $p < 0.001$), but PWI error rates were not significantly impacted by previous-trial ambiguity ($\beta = 0.27$, $z = 0.35$, $p = 0.73$). As with PWI RTs, previous-trial ambiguity did not significantly interact with current-trial congruency in the aggregate model ($\beta = 0.32$, $z = 0.04$, $p = 0.97$).

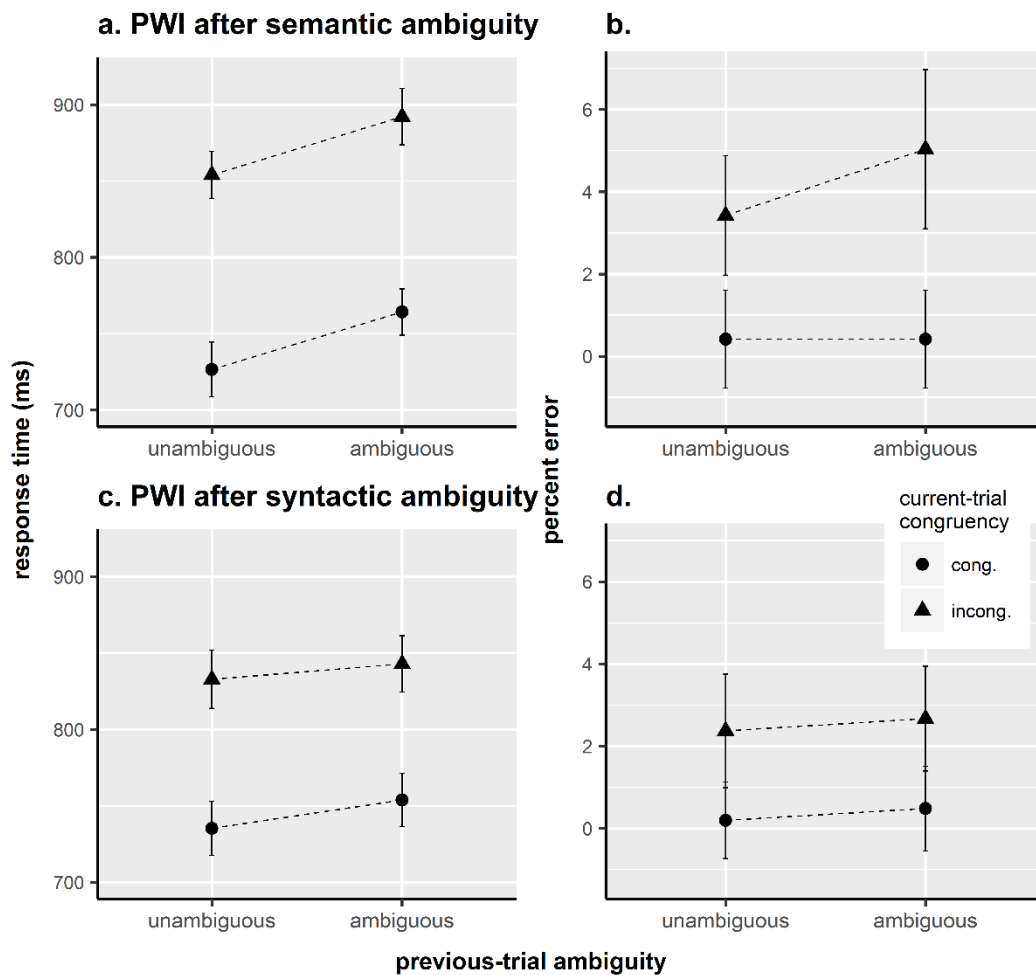


Figure 8. Cross-task CSEs in PWI in Experiment 2. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the ambiguity of a 1-back semantically ambiguous sentence or its control. Mean RT (c) and percent error (d) for congruent and incongruent PWI trials as a function of the ambiguity of a 1-back syntactically ambiguous sentence or its control. Error bars indicate 95% confidence intervals of within-participant variability. Note that there is no evidence of the canonical CSE in any of the four plots.

3.2.3.2. Sentence comprehension. As seen in Figure 9, both sentence types show a pattern compatible with a *larger* congruency (ambiguity) effects following an incongruent PWI trial. Results of a model of residual log reading times with the same random effect structure used to fit the within-CSE model is reported in Table B8 in Appendix B. Results revealed significantly slower reading times in ambiguous compared to unambiguous sentences ($\beta = 0.03$, $t = 3.63$, $p < 0.001$), but were not sensitive to the congruency of the previous PWI trial ($\beta = 0.0002$, $t = -0.02$, $p = 0.99$). Similarly, the interaction between current-trial ambiguity and previous-trial PWI congruency was not significant ($\beta = 0.01$, $t = 1.22$, $p = 0.22$).

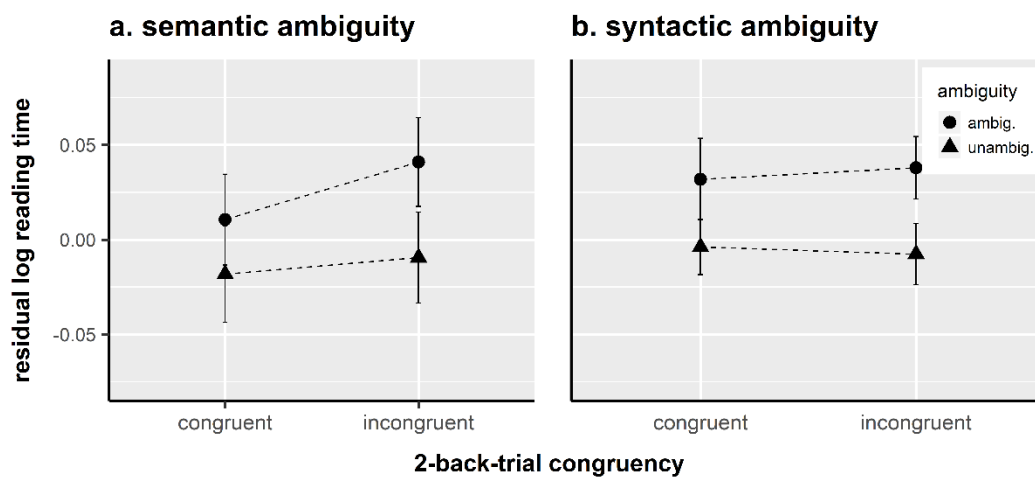


Figure 9. Cross-task CSEs in sentence reading in Experiment 2. Mean residual log reading times for semantically ambiguous sentences and their controls as a function of the ambiguity of a preceding PWI trial (a). Mean residual log reading times for syntactically ambiguous sentences and their controls as a function of the ambiguity of a preceding PWI trial (b). Error bars indicate 95% confidence intervals of within-participant variability. Note the consistent pattern across the two sentence types pointing to a reversed CSE.

3.3. Discussion

Experiment 2 followed the same logic and tested the same predictions as Experiment 1, with the difference that the temporal gap between the two PWI trials was more than doubled, and the intervening task tapped into representations from the same cognitive domain, namely language, as PWI. Results converged with those obtained in Experiment 1. PWI trials showed robust adaptation as a function of a 2-back PWI congruency, with a size that was no smaller than that obtained in Experiment 1 (54 ms in Experiment 2 vs. 45 ms in Experiment 1), and not significantly different when tested in a model of the combined data from both experiments. In addition, residual reading times also showed a statistically significant CSE as a function of the ambiguity on the previous reading trial despite the intervening PWI.

Results of the two experiments also converged in tests of the cross-task CSE. Canonical cross-task CSE was not found in either PWI or the sentence reading tasks. If anything, the pattern of the results (although not statistically significant) was in the direction of a reversed CSE effect, similar to that observed in the prime-probe task in Experiment 1. Together, these results are well aligned with the predictions of the learning account, and are incompatible with those of the task-specific-controller version of the activation account (in its prediction regarding

within-task CSE) and those of the general-controller version of the activation account (in its prediction regarding cross-task CSE).

4. General Discussion

In two experiments, we tested the viability of a learning model of adaptive control in language production against two versions of an activation model. Results of both experiments were consistent with one another and with predictions of the learning model. A robust CSE was found in PWI, even when the previous PWI trial was over 8 seconds earlier, and despite the fact that participants performed a secondary task in between the two PWI trials. In addition, the nature of the secondary task did not matter: the involvement of linguistic representations instead of visuospatial representations in the sentence reading task did not decrease CSE in PWI, ruling out an effect that depends heavily on domain-specific working memory processes (e.g., Shah & Miyake, 1996). To check the consistency of these findings, we also tested CSE in the other tasks in Experiments 1 and 2, both of which showed within-task CSE despite the intervening PWI trial. Together, these results showed a persistent within-task CSE in all our tasks, ruling out a version of the activation model with a task-specific controller in which short-lived priming of the controller should vanish, or strongly decrease, as a function of the gap between two trials of the same task.

As predicted by the learning model, the canonical CSE was only found within the same task, while cross-task CSE was either absent or was in the form of a reversed CSE. This ruled out the version of the activation account with a general controller, which would predict a reliable CSE between all trials regardless of the task. One might object that we have failed to find cross-task CSE because PWI and the two other tasks differ in various ways (e.g., complexity, number

of responses, etc.) and might thus elicit different kinds of control. But, keep in mind that we specifically chose tasks that have been shown in past studies to elicit cross-task CSE with button-press responses (Hsu & Novick, 2016; Kan et al., 2013). In fact, our sentence comprehension–PWI paradigm closely mirrors the sentence comprehension–Stroop paradigm used in Hsu and Novick (2016) and Kan et al. (2013), except in response modality—which, in our study, engages the production system. Another objection might be that we failed to find cross-task CSE because of low statistical power. Note, however, that in the same tasks where no canonical 1-back CSE was found, we found solid evidence for 2-back CSE in all tasks. In addition, we found evidence of a statistically-significant reversed 1-back CSE in both the prime-probe and the sentence reading tasks as a function of the previous PWI trial. Together, these findings make it highly unlikely that the absence of a canonical cross-task CSE in the language production task was due to either poor task choice, or low statistical power.

Finally, one might argue that, in both experiments, we have reported post-error slowing effects (which led to the exclusion of a subset of trials). Perhaps such post-error slowing regardless of the task is evidence for cross-task CSE. A large literature has investigated post-error effects, and different accounts have been proposed for why such effects arise. While a full discussion is out of the scope of the current paper, we present the gist of the findings in order to explain that post-error slowing differs from the CSE of interest in the current studies.

In an attempt to formally model post-error slowing, Dutilh et al. (2012) fitted a drift diffusion model to a large dataset obtained from the lexical decision task. First, these authors reported a response repetition effect, a tendency to repeat the same response—regardless of its accuracy—made on the previous trial. This finding indeed sounds like learning, and could point to similar processes underlying post-error and the CSE reported here. However, the rest of the

findings suggest otherwise: the main model parameter that was sensitive to errors was the boundary separation. That is, after an error, participants widened their response boundaries (i.e., became more cautious) in order to accumulate more evidence before committing to a response and thus to avoid further errors. This widening manifested as a speed-accuracy trade-off whereby longer RTs led to more accurate responses. This differs from within-task CSE, in which responses to the same type of trial (e.g., high-conflict trials) become *faster* and no less accurate (sometimes even *more* accurate) after experiencing another high-conflict trial. There was also an effect of the response bias parameter in the model, but the effect was only visible for nonwords (and not words), which, as acknowledged by the authors, makes the interpretation of findings difficult.

The speed-accuracy account of post-error slowing has, however, not gone undisputed. The main criticism is that empirical data do not always suggest such a tradeoff (see Ullsperger et al., 2014 for a review). Most recently, Purcell and Kiani (2016) showed that both humans and monkeys slowed down after errors without any reliable increase in performance accuracy. The drift diffusion model can handle this finding through a combination of an increased decision bound and a reduced sensitivity of the accumulator to perceptual information. While the former causes an increase in RTs, the latter prevents a rise in accuracy as a function of longer accumulation periods. These findings are most compatible with a general re-orienting reflex most likely due to unexpected events (Notebaert et al., 2009), which temporarily reduces perceptual sensitivity and causes motor inhibition, and is later followed by adjustments to enhance task-specific performance (Ullsperger, & Danielmeier, 2016) such as CSE. Importantly, both the behavioral response (slowing down) and the origin of the effect (temporary disengagement from the task) are distinct from the CSE.

We can thus conclude that our findings regarding cross-task CSE are aligned with previous studies (Kiesel et al., 2006; Notebaert & Verguts, 2008), and together with the pattern obtained in within-task CSE, support a learning account of adaptive control in language production.

4.1. Theoretical implications

The results of the study speak to two bodies of literature: the cognitive control and the language production literature. As reviewed in the Introduction, there is considerable debate over the mechanisms underlying CSE. The learning account is relatively novel, and its predictions, such as the longevity of the CSE, have not been broadly tested. The few studies that have examined such issues have led to inconsistent results (e.g., Egner et al., 2010; Wühr and Ansorge, 2005). Language production provides an excellent testbed for theories of cognitive control, and might offer unique insights into how control is implemented in a generative system that is highly experienced at resolving competition at different levels. The experiments reported here take advantage of the properties of language and examine CSE in tasks that tap into the same cognitive functions (e.g., picture naming or sentence comprehension), but with non-repeating materials on each trial, thus minimizing stimulus-specific effects. Despite the removal of these low-level effects, we show lasting task-specific control effects that lend credibility to learning models of adaptive control (e.g., Verguts, & Notebaert, 2008).

The second literature for which these results have critical implications is, of course, language production. Models of word production (Dell, 1986; Levelt, Roelofs, & Meyer, 1999; see Dell, Nozari & Oppenheim, 2014 for a review) have contributed tremendously to our understanding of how concepts are translated into speech. While some of these models have included extensive discussions of monitoring and control (e.g., Levelt, 1983), they have not

considered such processes to be an integral part of the production system itself. For example, in the classic account of self-monitoring, the perceptual loop theory (Levelt, 1983), monitoring is viewed as a task for the *comprehension* system. In other words, monitoring and control are viewed as operations of a separate system that influence production output. Recent models of monitoring, on the other hand, posit a much stronger role for the production system itself in speech monitoring. For instance, forward models of monitoring (e.g., Pickering & Garrod, 2013) propose a critical role for the production system in generating information useful for error monitoring (see Nozari & Novick, 2017 for a review of monitoring models in production).

Similarly, Nozari, Dell and Schwartz (2011) proposed that information such as conflict between representations during lexical and segmental selection is monitored by a domain-general monitoring center (e.g., the Anterior Cingulate Cortex, Botvinick et al., 2001; see Ullsperger et al., 2014 for alternative views), which releases an error signal based on the likelihood of an error. Such a monitor is more successful than a pure comprehension-based monitor in explaining electrophysiological signatures of error detection (Ganushchak & Schiller, 2008; Riès, et al., 2011), as well as error detection performance in individuals with post-stroke aphasia (Nozari et al., 2011) and children (Hanley et al., 2016), which point to the direct involvement of the production system in monitoring.

Much less work has focused on the nature of the control operations in the production system that follow monitoring. This is the first attempt, to our knowledge, to investigate whether the representations and operations within the production system are directly involved in the implementation of control, or whether they are only regulated online via a separate control system. The learning and activation accounts capture these two positions well: a learning account views control as making lasting changes to the production system itself. Activation accounts, on

the other hand, propose a transient influence of a Central controller over the production system with no persistent changes within the production system. The current findings provide strong support for the former. Moreover, they align well with production-based views of monitoring in that information generated within the production system during monitoring is used to change the same production system in ways that would make production more efficient in similar contexts.

This view dovetails with recent perspectives on how various parts of the production system show evidence of quick, implicit incremental learning in order to facilitate future performance based on current experience. For example, Warker and Dell (2006) showed that participants' error patterns reveal learning of experimentally induced phonotactic constraints (e.g., /s/ always occurring in onset but never in coda position) after only nine trials. Similarly, Oppenheim, Dell and Schwartz (2010) showed that a learning model provides the best account of why participants' production slows after the first cycle of naming semantically related pictures. More recently, we have proposed that a comparable learning mechanism leads to interference after the first cycle of naming pictures with segmentally related names (Breining et al., 2016; Breining, Nozari, & Rapp, under review; Nozari et al., 2016). The current results suggest that incremental learning also provides a viable mechanism for the implementation of control in the production system.

4.1.1. Domain-general vs. domain-specificity of monitoring and control

The question of domain-general vs. domain-specificity of monitoring and control processes have been central to the cognitive control literature, and more recently, to the language processing literature as well. While most of this literature has focused on whether the same brain region is involved in implementing control in different domains (e.g., Novick et al., 2005; Nozari & Thompson-Schill, 2015), we have recently argued that domain-general vs. domain-specificity may be defined at several levels (Nozari

& Novick, 2017). The first level is domain-generality as *shared computational principles*. For example, both forward models and production-based models of monitoring assume that very similar computational mechanisms underlie the generation of the error signal in the language production system, as well as in others such as motor or vision, even though these mechanisms operate on domain-specific representations (see also Hickok, 2012). A language-specific account, such as the perceptual loop monitor, however, does not follow domain-general computational principles for the detection of speech errors.

The current findings, along with other studies of CSE in language production, suggest that implementation of control in language production is likely to follow computational principles similar to that which other non-linguistic systems obey—namely, quick adjustments of performance through monitoring information generated within the production system. Conflict may be an example of such information (e.g., Botvinick et al., 2001), but it is possible that other information is being used by the monitor to adjust performance (Lamers & Roelofs, 2011). As far as the learning account is concerned, what is produced on the current trial is learned, and thus reinforced, on the next trial.

The second level of domain-generality described by Nozari and Novick (2017) is *shared neural implementations*. As alluded to before, this aspect of domain-generality has been discussed extensively elsewhere; here, we will only mention that the current body of evidence points to both shared and distinct neural substrates for monitoring and control in linguistic and non-linguistic tasks (e.g., De Zubicaray et al., 2006; De Zubicaray, McMahon, & Howard, 2015; Gauvin, De Baene, Brass, & Hartsuiker, 2016; Jiang & Egner, 2014; Piai et al., 2013; Riès et al., 2015). Finally, the third level of domain-generality is domain generality as *cross-task adjustment*

in control, that is, whether an increase in the control demand in one task leads to better implementation of control in a different task.

The evidence for this third level of domain-generalness is mixed. Several studies have reported no cross-task adjustment in control (e.g., Akçay & Hazeltine, 2011; Boy et al., 2010; Egner et al., 2007; Funes, Lupiáñez, & Humphreys, 2010a; Kim et al., 2012; Kunde, Augst, & Kleinsorge, 2012; Kunde & Stöcker, 2002; Kunde & Wühr, 2006; Schlaghecken et al., 2011; Verbruggen, Liefoghe, Notebaert, & Vandierendonck, 2005; Wendt et al., 2006; Wühr et al., 2015), while a few have reported such cross-task transfer of control (Freitas et al., 2007; Hsu & Novick, 2016; Kan et al., 2013; Kleiman, Hassin, & Trope, 2014). This discrepancy may have several origins, such as the lack of statistical power in detecting small effects in the former set of studies, or a special status of certain domains such as language comprehension (e.g., Kan et al., 2013; Hsu & Novick, 2016) that may lend themselves to domain-general control better than other domains. The design of the current experiment aimed to cover both the issue of statistical power, by including demonstrations of a robust 2-back within-task adaptation in all tasks, as well as the issue of domain, by using the same domains used in two of the studies that previously reported cross-task transfer. Similar to the larger body of evidence, we found no support for cross-task adjustment in control.

How can these discrepant results be reconciled? A possible explanation is that learning constitutes the basis of the CSE. Because learning concerns connections to task-specific representations, the majority of studies yield no CSE transfer between different tasks. However, CSE *could* also result from the short-term priming of a domain-general component of the control center, even if the rest of the control network is task-specific (e.g., Jiang & Egner, 2014). This activation component explains the results of those studies that obtained cross-task CSE, as well

as the decrease in the size of CSE observed in Egner et al. (2010) and Wühr and Ansorge (2005). However, priming-induced CSE is clearly not observed under all circumstances, including the current tasks. Future research might shed light on the specific situations where activation-based CSE contributes to adaptive control. What we can conclude with certainty given the current results, however, is that learning plays a prominent role in adaptive control in language production, resulting in domain-specific control regulation.

In summary, it appears that monitoring and control processes in language production follow the same general principles as in other domains (Nozari et al., 2011; Riès et al., 2011), and are implemented by neural substrates that are at least partially shared with other domains. However, when it comes to online monitoring and regulation of control, the production system shows specificity on both accounts. In earlier work (Hanley et al., 2016; Nozari et al., 2011; Nozari & Novick, 2017) we have described the origin of this specificity for monitoring. The current study has identified the root of this specificity for control implementation in a learning mechanism. This is important for theoretical models of language production, because it means that in order to conceive a holistic “language production network”, it is necessary to include not only central monitoring and control regions, but also their specific links to the representations and operations within the production system. The domain-specificity of adaptive control also has critical implications for clinical practice, which we discuss in the next section.

4.2. Clinical implications

Including monitoring and control processes in the language production network is not only a matter of theoretical completeness, but has a critical impact on diagnosis and treatment of language disorders such as post-stroke aphasia. For example, Jefferies and Lambon Ralph (2006) identified a specific class of individuals with aphasia whose primary deficit was in suppressing a

competitor during lexical selection (see also Robinson, Shallice, Bozzali, & Cipolotti, 2010). It is reasonable to deduce from this finding that treatment in such individuals must focus on strengthening inhibitory control. But what kind of inhibitory control training would be useful? Generally speaking, two approaches to cognitive control training have been proposed. One approach claims wide transfer of benefits from training control in a given task to other tasks, domains, and even functions such as working memory (e.g., Au et al., 2015; Jaeggi et al., 2011; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). A more moderate version of this approach claims transfer between tasks in different domains but not between different functions (Hussey et al., 2016). Against these claims, several studies have questioned the reliability of such transfer, in favor of a more task-specific view of training (e.g., Owen et al., 2010; Redick et al., 2013; Shipstead, Redick, & Engle, 2010, 2012; see Simons et al., 2016, for a comprehensive review).

Our results have direct implications for this debate: we found no evidence that implementation of control in a non-production (visuospatial or comprehension) task benefits the implementation of control in a word production task. This finding, which was discussed in the previous section as domain-specific adjustments in control, predicts that practicing control implementation in non-production tasks would have little benefit for resolving competition in the language production system. On the other hand, the results show much promise for the efficacy of a cognitive control training method focused on resolving competition in language production. Control implemented on one production trial facilitated performance on the next control-demanding production trial, even when participants performed other tasks in-between, suggesting that practicing control could have lasting effects. Moreover, the effect was not item-specific; it transferred across 120 unique items in PWI. This finding suggests that practicing inhibitory control in a task with a limited number of items may have benefits for a much larger of

set of items. Recall that Shitova et al. (2017) localized the competition in PWI to the level of lexical selection. Thus, training control using PWI—or a similar—task can theoretically be expected to benefit control during natural word production for individuals with inhibitory control deficit. Importantly, the best outcome for language production can be expected when inhibitory control is trained on a language production task, as opposed to other tasks that do not engage production processes.

4.3. Conclusion

Regardless of whether the task involves language production, comprehension, or neither, the congruency sequence effect is persistent and specific to the task at hand, yet abstract enough to generalize across particular items. In language production, these properties are well captured by an adaptive control mechanism mediated by learning, in which the change is integral to the production system itself. This learning perspective links adaptive control to other incremental learning mechanisms in language production and provides an angle from which one can evaluate the potential efficacy of training methods.

5. Acknowledgements

This work was supported in part by the NSF grant 1631993 awarded to Nazbanou Nozari, and in part by the Cognitive Neuroscience Fund awarded to the Cognitive Neurology division of the Neurology Department at Johns Hopkins School of Medicine. We would like to thank Aaron Apple for his help with the design, Chris Hepner for his comments on an earlier version of the manuscript, and Jared Novick, Ardi Roelofs and Natalia Shitova for useful discussions.

6. Supplemental material

All data reported here are archived on Open Science Framework and are publicly accessible via the following link: osf.io/z2cwn.

7. References

- Akçay, Ç., & Hazeltine, E. (2011). Domain-specific conflict adaptation without feature repetitions. *Psychonomic Bulletin & Review*, *18*(3), 505–511. doi:10.3758/s13423-011-0084-y
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377. doi:10.3758/s13423-014-0699-x
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. Retrieved from <http://arxiv:abs/1406.5823>
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, *58*(4), 667-692. doi: 10.1080/02724980443000142
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Review*, *108*(3), 624–652. doi:10.1037/0033-295X.108.3.624
- Boy, F., Husain, M., & Sumner, P. (2010). Unconscious inhibition separates two forms of cognitive control. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 11134–9. doi:10.1073/pnas.1001925107

- Braem, S., Abrahamse, E. L., Duthoo, W., & Notebaert, W. (2014). What determines the specificity of conflict adaptation? A review, critical analysis, and proposed synthesis. *Frontiers in Psychology*, 5, 1134. doi:10.3389/fpsyg.2014.01134
- Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic bulletin & review*, 23(2), 500-506. doi: 10.3758/s13423-015-0900-x
- Breining, B., Nozari, N., & Rapp, B. (under review). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42, 368–407. doi: 10.1006/cogp.2001.0752
- Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia*, 49(7), 1953–1961. doi:10.1016/j.neuropsychologia.2011.03.023
- Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, 12(1), 125–131. doi:10.3758/BF03196357
- Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, 2, 233. doi:10.3389/fpsyg.2011.00233
- De Zubicaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory

effects in picture naming. *Human Brain Mapping*, 27(11), 864–873. doi:10.1002/hbm.20227

De Zubicaray, G., McMahon, K., & Howard, D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in spoken word production. *Language, Cognition and Neuroscience*, 30, 261–272. doi: 10.1080/01690965.2013.848990

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. doi:10.1037/0033-295X.93.3.283

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Lexical access: Behavioral and computational considerations. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 88-104). Oxford University Press.
doi:10.1093/oxfordhb/9780199735471.013.014

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27(4), 429–446. doi:10.1016/0749-596X(88)90066-6

Durston, S., Davidson, M. C., Thomas, K. M., Worden, M. S., Tottenham, N., Martinez, A., ... & Casey, B. J. (2003). Parametric manipulation of conflict and response competition using rapid mixed-trial event-related fMRI. *Neuroimage*, 20(4), 2135-2141. doi: 10.1016/j.neuroimage.2003.08.004

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The Congruency Sequence Effect 3.0: A Critical Test of Conflict Adaptation. *PLoS ONE*, 9(10). doi: 10.1371/journal.pone.0110462

- Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454-465. doi: 10.3758/s13414-011-0243-2
- Egner, T. (2014). Creatures of habit (and control): A multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, doi:10.3389/fpsyg.2014.01247
- Egner, T., Delano, M., & Hirsch, J. (2007). Separate conflict-specific cognitive control mechanisms in the human brain. *NeuroImage*, *35*(2), 940–948. doi:10.1016/j.neuroimage.2006.11.061
- Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428-16433. doi: 10.1073/pnas.1112937108
- Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616-16621. doi: 10.1073/pnas.1315235110
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. doi:10.1016/0749-596X(86)90006-9
- Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555–68. doi: 10.1037/0278-7393.16.4.555
- Ferreira, V.S. & Slevc, L.R. (2007). Grammatical encoding. In M.G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 453-469). Oxford: Oxford University Press.
- Freitas, A. L., & Clark, S. L. (2015). Generality and specificity in cognitive control: conflict

adaptation within and across selective-attention tasks but not across selective-attention and Simon tasks. *Psychological Research*, 79(1), 143–162. doi: 10.1007/s00426-014-0540-1

Freitas, A. L., Bahar, M., Yang, S., & Banai, R. (2007). Contextual in Adjustments Control Across Tasks Cognitive Control Across Tasks. *Psychological Science*, 18(12), 1040–1043.

Freund, M., Gordon, B., & Nozari, N. (2016). Conflict-based regulation of control in language production. In A. Papafragou, D. Grodner, D. Mirman, & J.C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1625-1630). Austin, TX: Cognitive Science Society.

[dataset] Freund, M., & Nozari, N. (2018, March 5). Freund and Nozari (2018) Cognition. Retrieved from osf.io/z2cwn

Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Analyzing the generality of conflict adaptation effects. *Journal of Experimental Psychology. Human Perception and Performance*, 36(1), 147–61. doi:10.1037/a0017598

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *Neuroimage*, 39(1), 395-405.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. a. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, 37(1), 58–93. doi:http://dx.doi:10.1006/jmla.1997.2512

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage*, 126, 96–105. doi:10.1016/j.neuroimage.2015.11.037

- Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, *54*(3), 363–388.
doi:10.1016/j.jml.2005.12.005
- Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology. General*, *121*(4), 480–506. doi: 10.1037/0096-3445.121.4.480
- Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic Complexity in Ambiguity Resolution. *Journal of Memory and Language*, *46*(2), 267–295. doi:10.1006/jmla.2001.2808
- Hanley, J. R., Cortis, C., Budd, M. J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of experimental child psychology*, *142*, 36-47. doi: 10.1016/j.jecp.2015.09.008
- Hengstler, M., Holland, R., Steenbergen, H., & Knippenberg, A. (2014). The influence of approach–avoidance motivational orientation on conflict adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 548–560. doi: 10.3758/s13415-014-0295-6
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145. doi: 10.1038/nrn3158
- Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405.
doi:10.1080/01690965.2010.492642
- Hommel, B. (1998). Event Files: Evidence for Automatic Integration of Stimulus-Response Episodes. *Visual Cognition*, *5*(1–2), 183–216. doi: 10.1080/713756773

- Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, 68(1), 1–17. doi: 10.1007/s00426-003-0132-y
- Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and language*, 166, 63-77. doi: 10.1016/j.bandl.2016.12.006
- Hsu, N. S., & Novick, J. M. (2016). Dynamic Engagement of Cognitive Control Modulates Recovery From Misinterpretation During Real-Time Language Processing. *Psychological Science*, 27(4), 572–82. doi:10.1177/0956797615625223
- Hussey, E. K., Harbison, J. I., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2016). Memory and Language Improvements Following Cognitive Control Training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000283
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833. doi:10.1073/pnas.0801268105
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., Shah, P., Morrison, A. B., & Chein, J. M. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25), 46–60. doi:10.1073/pnas.1103228108

- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(8), 2132-2147. doi: 10.1093/brain/awl153
- Jiang, J., & Egner, T. (2014). Using neural pattern classifiers to quantify the modularity of conflict-control mechanisms in the human brain. *Cerebral Cortex*, *24*(7), 1793-1805. doi:10.1093/cercor/bht029
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, *111*(2), 228–38. doi: 10.1037/0096-3445.111.2.228
- Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637–651. doi:10.1016/j.cognition.2013.09.001
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023-1026. doi: 10.1126/science.1089910
- Kiesel, A., Kunde, W., & Hoffmann, J. (2006). Evidence for task-specific resolution of response conflict. *Psychonomic Bulletin & Review*, *13*(5), 800-806. doi: 10.3758/BF03194000
- Kim, C., Chung, C., & Kim, J. (2012). Conflict adjustment through domain-specific multiple cognitive control mechanisms. *Brain Research*, *1444*, 55–64. doi:10.1016/j.brainres.2012.01.023
- Kleiman, T., Hassin, R. R., & Trope, Y. (2014). The control-freak mind: stereotypical biases are

- eliminated following conflict-activated cognitive control. *Journal of Experimental Psychology. General*, *143*(2), 498–503. doi:10.1037/a0033047
- Kunde, W., & Stöcker, C. (2002). A Simon effect for stimulus-response duration. *Quarterly Journal of Experimental Psychology*, *55A*(2), 581–592. doi:10.1080/0272498014300043
- Kunde, W., & Wühr, P. (2006). Sequential modulations of correspondence effects across spatial dimensions and tasks. *Memory & Cognition*, *34*(2), 356–367. doi:10.3758/BF03193413
- Kunde, W., Augst, S., & Kleinsorge, T. (2012). Adaptation to (non)valent task disturbance. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(4), 644–660. doi:10.3758/s13415-012-0116-8
- Lamers, M. J., & Roelofs, A. (2011). Attentional control adjustments in Eriksen and Stroop task performance can be independent of response conflict. *The Quarterly Journal of Experimental Psychology*, *64*(6), 1056–1081. doi: 10.1080/17470218.2010.523792
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. doi:10.1016/0010-0277(83)90026-4
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. doi: 10.3115/992628.992631
- Macdonald, M. C. (1993). The Interaction of Lexical and Syntactic Ambiguity. *Journal of Memory and Language*, *32*(5), 692–715. doi:10.1006/jmla.1993.1035
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the

picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 503. doi: 10.1037/0278-7393.33.3.503

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, 6(5), 450–452. doi:10.1038/nn1051

Mordkoff, J. T. (2012). Observation: Three reasons to avoid having half of the trials be congruent in a four-alternative forced-choice experiment on sequential modulation. *Psychonomic Bulletin & Review*, 19(4), 750–757. doi:10.3758/s13423-012-0257-3

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61–64. doi:10.3758/s13414-012-0291-2

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation*, 12(1), 16–37. doi:10.3758/BF03208320

Nieuwenhuis, S., Stins, J. F., Posthuma, D., Polderman, T. J. C., Boomsma, D. I., & de Geus, E. J. (2006). Accounting for sequential trial effects in the flanker task: Conflict adaptation or associative priming? *Memory & Cognition*, 34(6), 1260–1272. doi: 10.3758/BF03193270

Notebaert, W., Gevers, W., Verbruggen, F., & Liefoghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review*, 13(1), 112–117. doi: 10.3758/BF03193821

Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, 111(2), 275-279. doi:

10.1016/j.cognition.2009.02.002

Notebaert, W., & Verguts, T. (2008). Cognitive control acts locally. *Cognition*, *106*(2), 1071-1080.

doi: 10.1016/j.cognition.2007.04.011

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527-567. doi:

10.1080/02643290903519367

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263-281. doi: 10.3758/CABN.5.3.263

Nozari, N. & Thompson-Schill, S.L. (2015). Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. In G. Hickok & S.L. Small (Eds.), *The Neurobiology of Language* (pp. 569-588). Waltham, MA: Academic Press. doi: 10.1016/B978-0-12-407794-2.00046-8

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. doi:10.1016/j.cogpsych.2011.05.001

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886-903. doi: 10.1080/23273798.2016.1157194

Nozari, N. & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403-410. doi: 10.1177/09637214177024

- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–52. doi:10.1016/j.cognition.2009.09.007
- O'seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human perception and performance*, *26*(1), 57. doi: 10.1037/0096-1523.26.1.57
- Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., ... Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–8. doi:10.1038/nature09042
- Piai, V., Riès, S. K., & Swick, D. (2016). Lesions to Lateral Prefrontal Cortex Impair Lexical Interference Control in Word Production. *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2015.00721
- Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013). Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience*. doi:10.3389/fnhum.2013.00832
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347. doi: 10.1017/S0140525X12001495
- Purcell, B. A., & Kiani, R. (2016). Neural mechanisms of post-error adjustments of decision policy in parietal cortex. *Neuron*, *89*(3), 658-671. doi: 10.1016/j.neuron.2015.12.027
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle,

- R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology. General*, *142*(2), 359–79. doi:10.1037/a0029082
- Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language*, *149*, 135–147. doi:10.1016/j.bandl.2015.07.007
- Riès, S., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, *23*(6), 1419–1436. doi:10.1162/jocn.2010.21467
- Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: Neuropsychological evidence from a focal frontal group. *Neuropsychologia*, *48*(6), 1652–1663. doi: 10.1016/j.neuropsychologia.2010.02.010
- Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive psychology*, *68*, 33-58. doi: 10.1016/j.cogpsych.2013.10.001
- Schlaghecken, F., Refaat, M., & Maylor, E. a. (2011). Multiple systems for cognitive control: evidence from a hybrid prime-Simon task. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(5), 1542–53. doi:10.1037/a0024327
- Schmidt, J. R., & De Houwer, J. (2011). Now you see it, now you don't: Controlling for contingencies and stimulus repetitions eliminates the Gratton effect. *Acta Psychologica*, *138*(1), 176–186. doi:10.1016/j.actpsy.2011.06.002

- Schmidt, J. R., & Weissman, D. H. (2015). Contingent attentional capture triggers the congruency sequence effect. *Acta Psychologica, 159*, 61–68. doi:10.1016/j.actpsy.2015.05.007
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language, 75*, 27–44. doi:10.1016/j.jml.2014.04.006
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences of the United States of America, 106*(1), 322–7. doi:10.1073/pnas.0805874106
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102. doi:10.1016/0749-596X(90)90011-N
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of experimental psychology: General, 125*(1), 4. doi: 10.1037/0096-3445.125.1.4
- Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition, 41*(8), 1200–11. doi: 10.3758/s13421-013-0332-7
- Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology, 65*(10), 1927–1944. doi: 10.1080/17470218.2012.670252
- Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming

performance in semantic blocking, picture-word interference, and color-word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1806–1820. doi: 10.1037/a0039363

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50(3–4), 245. doi:10.5334/pb-50-3-4-245

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. doi:10.1037/a0027473

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, 92, 289-303. doi: 10.1016/j.cortex.2017.04.017

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work?. *Psychological Science in the Public Interest*, 17(3), 103-186. doi:10.1177/1529100616661983

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 28(6), 1345. doi: 10.1037/0096-1523.28.6.1345

Szekely, A., Jacobsen, T., D’Amico, S., Devescovi, A., Andonova, E., Herron, D., ... Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247–250. doi:10.1016/j.jml.2004.03.002

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on

sentence processing. *Journal of Memory and Language*, 50(4), 355–370.

doi:10.1016/j.jml.2004.01.001

Thomas, S. J., Gonsalvez, C. J., & Johnstone, S. J. (2009). Sequence effects in the Go/NoGo task: Inhibition and facilitation. *International Journal of Psychophysiology*, 74(3), 209-219. doi: 10.1016/j.ijpsycho.2009.09.002

Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797. doi:10.1073/pnas.94.26.14792

Trude, A., & Nozari, N. (2017). Inhibitory Control Supports Referential Context Use in Language Production and Comprehension. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1218-1223). Austin, TX: Cognitive Science Society.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*. doi:10.1006/jmla.1994.1014

Ullsperger, M., & Danielmeier, C. (2016). Reducing speed and sight: how adaptive is post-error slowing?. *Neuron*, 89(3), 430-432. doi: 10.1016/j.neuron.2016.01.035

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological reviews*, 94(1), 35-79. doi: 10.1152/physrev.00041.2012

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal

dynamics of performance monitoring. *Trends in cognitive sciences*, 18(5), 259-267. doi:
10.1016/j.tics.2014.02.009

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285–316. doi:10.1016/S0749-596X(03)00081-0

Van Maanen, L., & Van Rijn, H. (2010). The locus of the Gratton effect in picture word interference. *Topics in Cognitive Science*, 2, 168-180. doi: 10.1111/j.1756-8765.2009.01069.x

Verbruggen, F., Liefoghe, B., Notebaert, W., & Vandierendonck, A. (2005). Effects of stimulus-stimulus compatibility and stimulus-response compatibility on response inhibition. *Acta Psychologica*, 120(3), 307–326. doi:10.1016/j.actpsy.2005.05.003

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological Review*, 115(2), 518–25. doi: 10.1037/0033-295X.115.2.518

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: a learning account of cognitive control. *Trends in Cognitive Sciences*, 13(6), 252–257. doi:10.1016/j.tics.2009.02.007

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 387. doi: 10.1037/0278-7393.32.2.387

Weissman, D. H., Egnér, T., Hawks, Z., & Link, J. (2015). The congruency sequence effect emerges when the distracter precedes the target. *Acta Psychologica*, 156, 8–21.
doi:10.1016/j.actpsy.2015.01.003

Wendt, M., Kluwe, R. H., & Peters, A. (2006). Sequential modulations of interference evoked by processing task-irrelevant stimulus features. *Journal of Experimental Psychology. Human Perception and Performance*, 32(3), 644–67. doi:10.1037/0096-1523.32.3.644

Wühr, P., & Ansorge, U. (2005). Exploring trial-by-trial modulations of the Simon effect. *The Quarterly Journal of Experimental Psychology*, 58(4), 705-731. doi:10.1080/02724980443000269

Wühr, P., Duthoo, W., & Notebaert, W. (2015). Generalizing attentional control across dimensions and tasks: Evidence from transfer of proportion-congruent effects. *The Quarterly Journal of Experimental Psychology*, 68(4), 779–801. doi:10.1080/17470218.2014.966729

7. Figure captions

Figure 1. Schematic of the learning and activation models of adaptive control. The important difference between the two models is in the locus of the effect (black arrows), which is changes to the weights in the learning model and increased activation of the Control center in the activation model. Both models predict CSE between task A trials without an intervening task (single-task CSE), but make different predictions about CSE during task switching (see text for predictions). Canonical CSE is marked by +. Absence of the canonical CSE is marked by –.

Figure 2. Schematic of the general design for Experiments 1 and 2. PWI trials are interleaved with trials from a different task—prime-probe in Experiment 1 and sentence reading in Experiment 2—in a predictable ABAB fashion. Prime-probe trials induced a gap of 3666 ms between the two PWI trials, and this gap was more than doubled in the sentence reading trials. Changes to performance on the current trial as a function of the previous trial in the same task

(i.e., a 2-back trial) constitute within-task CSE (longer arrows on the left side of the figure). A learning model predicts robust within-task CSE despite the intervening trial (prediction 1). Changes to performance on the current trial as a function of the previous trial belonging to a different task (i.e., 1-back trial) constitute cross-task CSE. A learning model predicts no cross-task CSE or a reversed cross-task CSE.

Figure 3. Within-task CSEs in Experiment 1. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 2-back PWI trial. Mean RT (c) and percent error (d) for congruent and incongruent prime-probe trials as a function of the congruency of a 2-back prime-probe trial. Error bars indicate 95% confidence intervals of within-participant variability (Morey, 2008).

Figure 4. Cross-task CSEs in Experiment 1. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 1-back prime-probe trial. Mean RT (c) and percent error (d) for congruent and incongruent prime-probe trials as a function of the congruency of a 1-back PWI trial. Error bars indicate 95% confidence intervals of within-participant variability.

Figure 5. Ambiguity effects in Experiment 2. Mean residual log reading time for the disambiguating (a) and “spillover” (b) regions of semantically and syntactically ambiguous sentences. Error bars indicate 95% confidence intervals of within-participant variability.

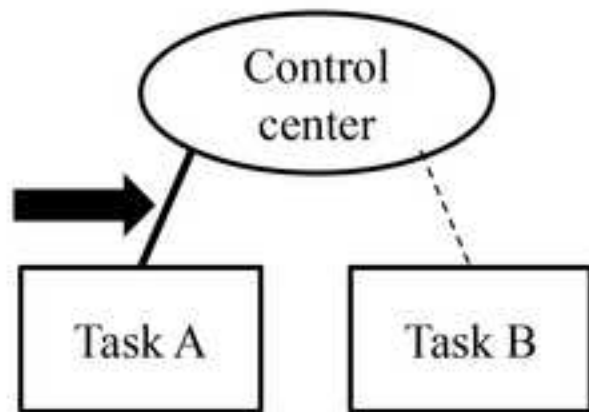
Figure 6. Within-task CSEs in PWI in Experiment 2. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the congruency of a 2-back PWI trial with an intervening semantically ambiguous sentence or its control. Mean RT (c) and percent error (d) for congruent and incongruent PWI trials as trials as a function of the congruency of a 2-back PWI trial with an intervening syntactically ambiguous sentence or its control. Error bars indicate 95% confidence intervals of within-participant variability. Note the consistency of the pattern of results across the four graphs in the direction predicted by the canonical CSE.

Figure 7. Within-task CSEs in sentence reading in Experiment 2. Mean residual log reading times for semantically ambiguous sentences and their controls as a function of the ambiguity of a 2-back sentence with an intervening PWI trial (a). Mean residual log reading times for syntactically ambiguous sentences and their controls as a function of the ambiguity of a 2-back sentence with an intervening PWI trial (b). Error bars indicate 95% confidence intervals of within-participant variability. Note the consistency of the pattern of results across the two sentence types in the direction predicted by the canonical CSE.

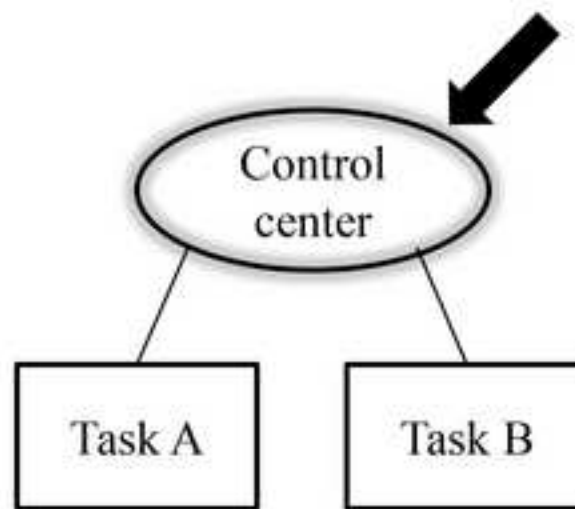
Figure 8. Cross-task CSEs in PWI in Experiment 2. Mean RT (a) and percent error (b) for congruent and incongruent PWI trials as a function of the ambiguity of a 1-back semantically ambiguous sentence or its control. Mean RT (c) and percent error (d) for congruent and incongruent PWI trials as a function of the ambiguity of a 1-back syntactically ambiguous sentence or its control. Error bars indicate 95% confidence intervals of within-participant variability. Note that there is no evidence of the canonical CSE in any of the four plots.

Figure 9. Cross-task CSEs in sentence reading in Experiment 2. Mean residual log reading times for semantically ambiguous sentences and their controls as a function of the ambiguity of a preceding PWI trial (a). Mean residual log reading times for syntactically ambiguous sentences and their controls as a function of the ambiguity of a preceding PWI trial (b). Error bars indicate 95% confidence intervals of within-participant variability. Note the consistent pattern across the two sentence types pointing to a reversed CSE.

Figure1
[Click here to download high resolution image](#)



Learning model



Activation model

*Single-task
CSE*

$$A^+ \text{---} A$$

$$A^+ \text{---} A$$

*Task-switching
CSE*

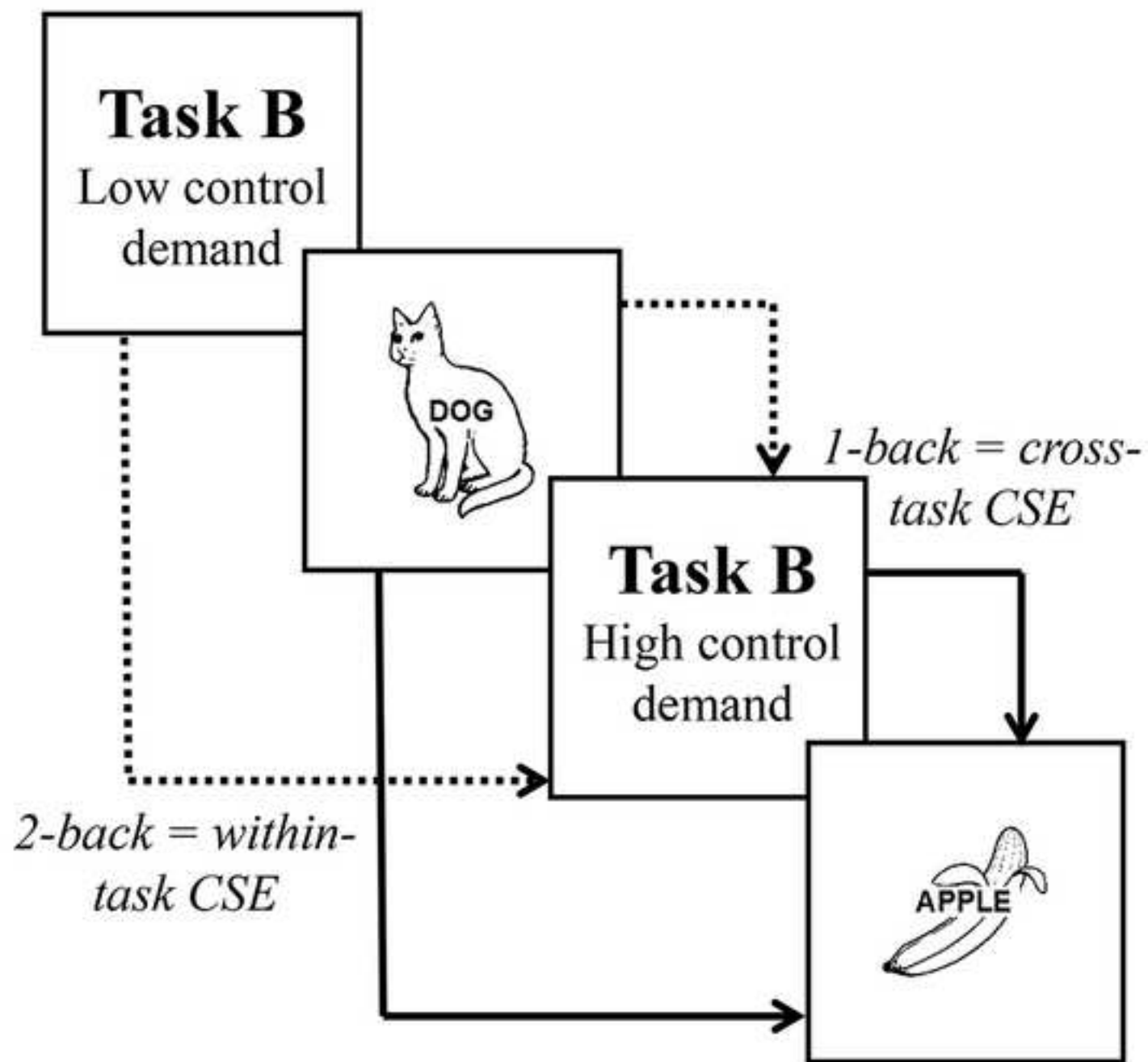
$$\begin{array}{c} + \\ \text{---} \\ A^- \text{---} B^- \text{---} A^- \text{---} B \\ \text{---} \\ + \end{array}$$

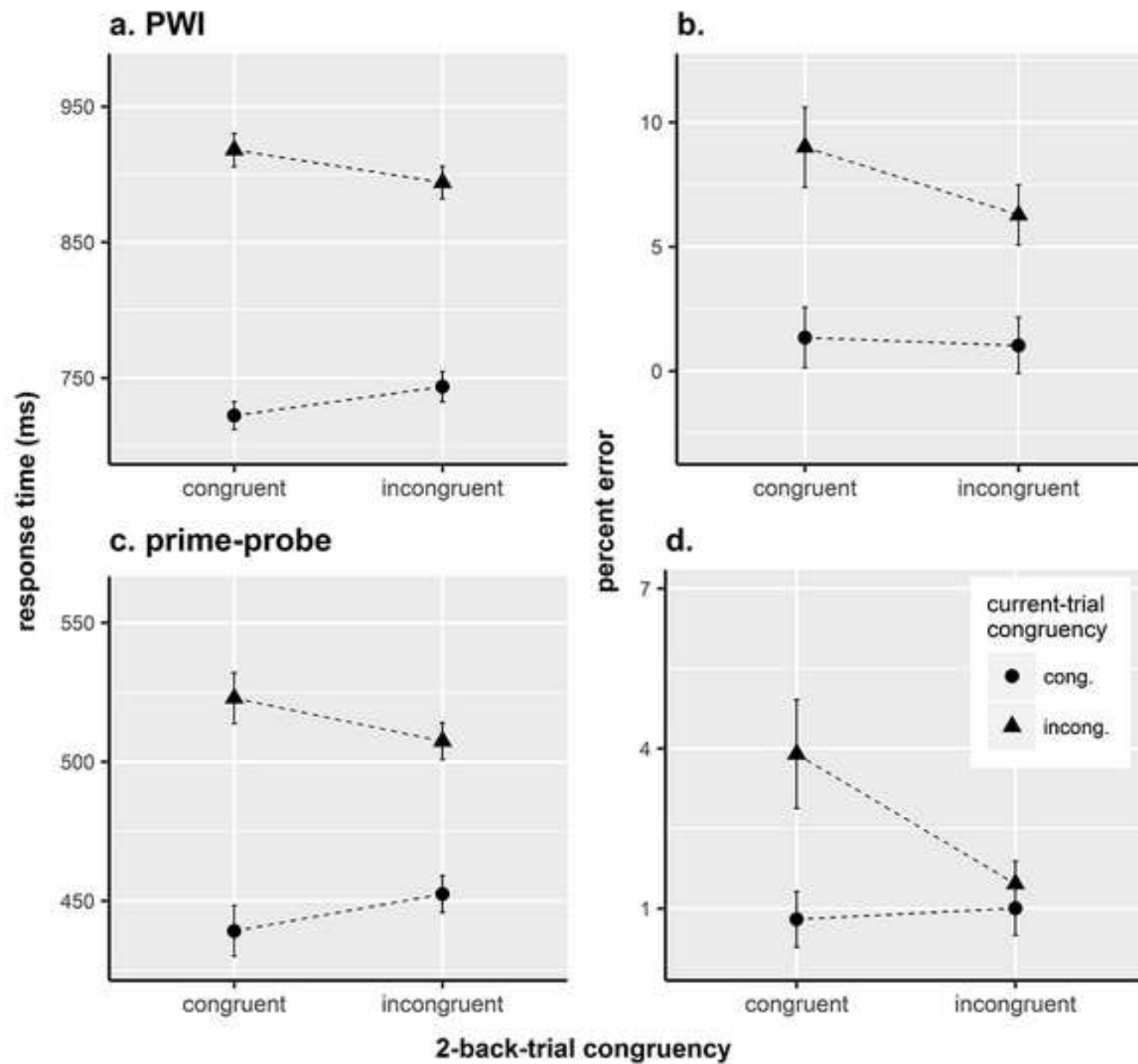
$$\begin{array}{c} + \\ \text{---} \\ A^+ \text{---} B^+ \text{---} A^+ \text{---} B \\ \text{---} \\ + \end{array} \quad \text{General controller}$$

$$\begin{array}{c} - \\ \text{---} \\ A^- \text{---} B^- \text{---} A^- \text{---} B \\ \text{---} \\ - \end{array} \quad \text{Task-specific controller}$$

Figure(s)

[Click here to download high resolution image](#)





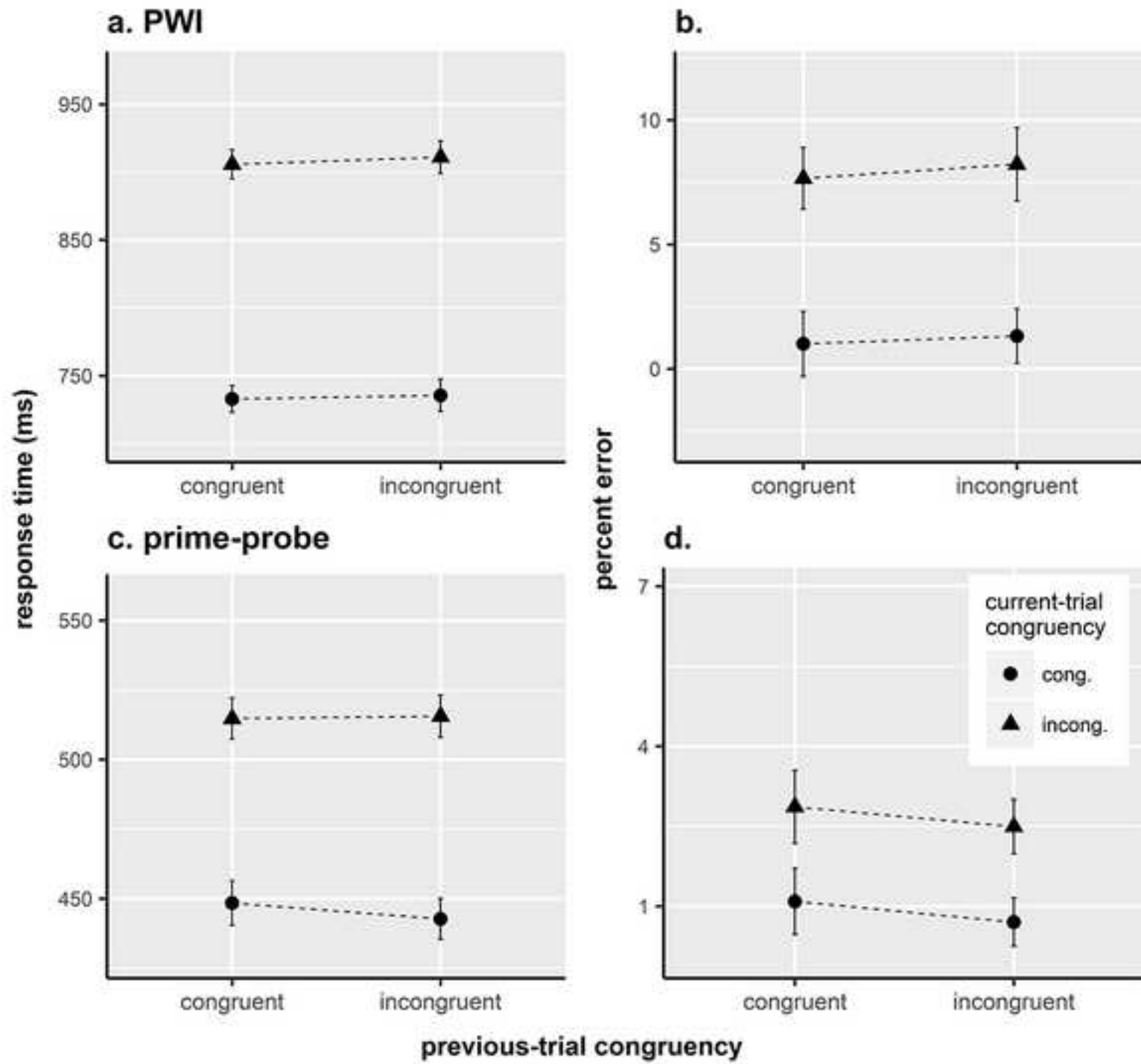
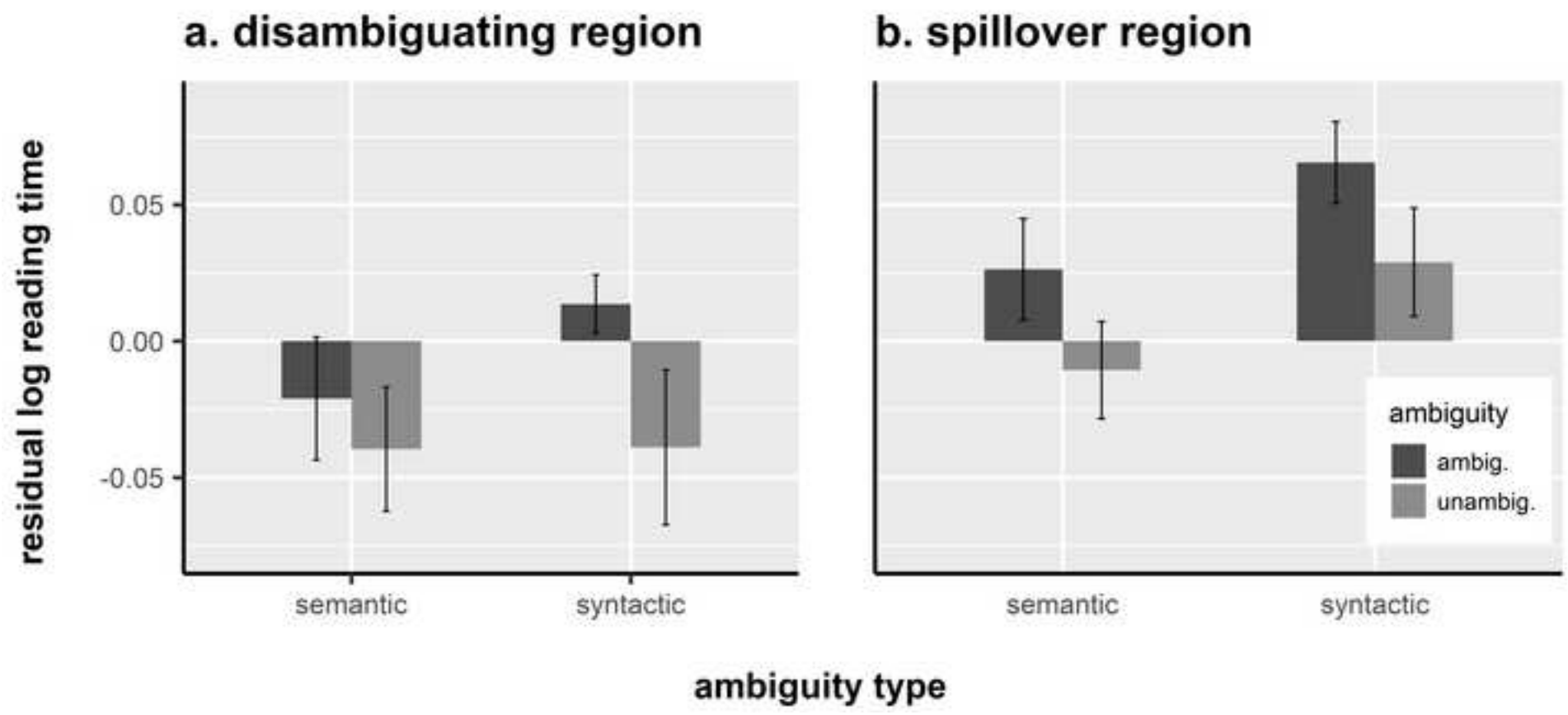


Figure5

[Click here to download high resolution image](#)



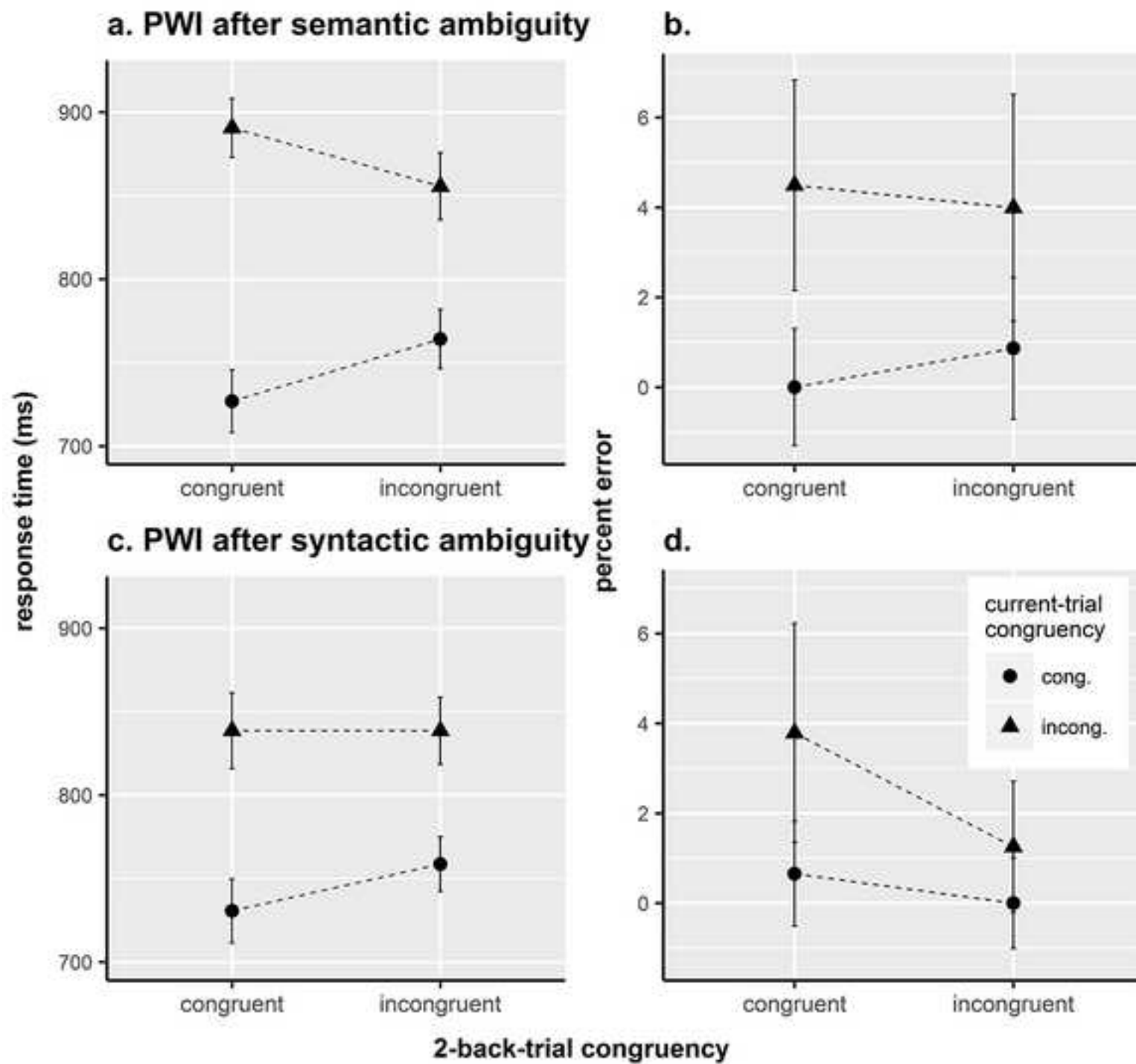
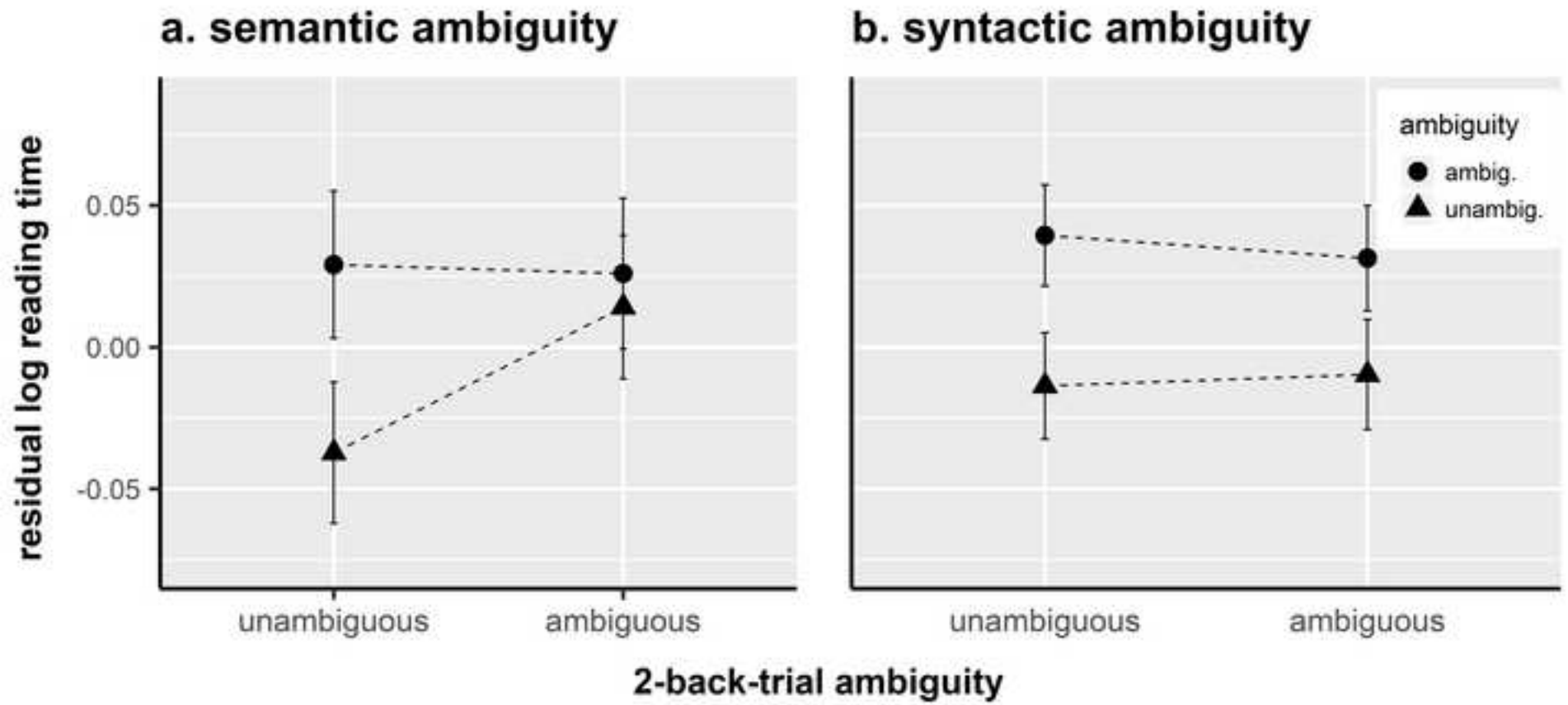


Figure7
[Click here to download high resolution image](#)



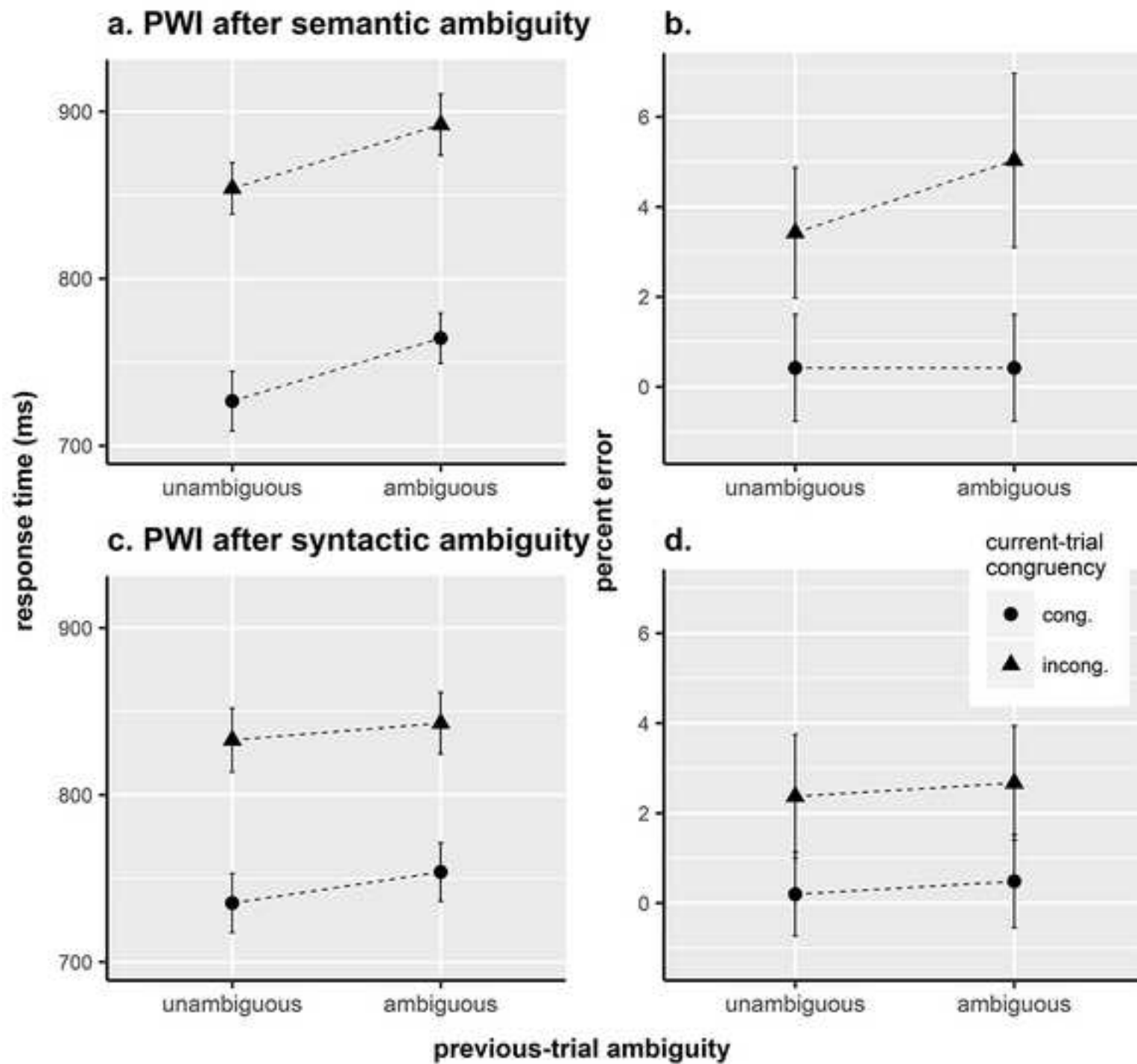
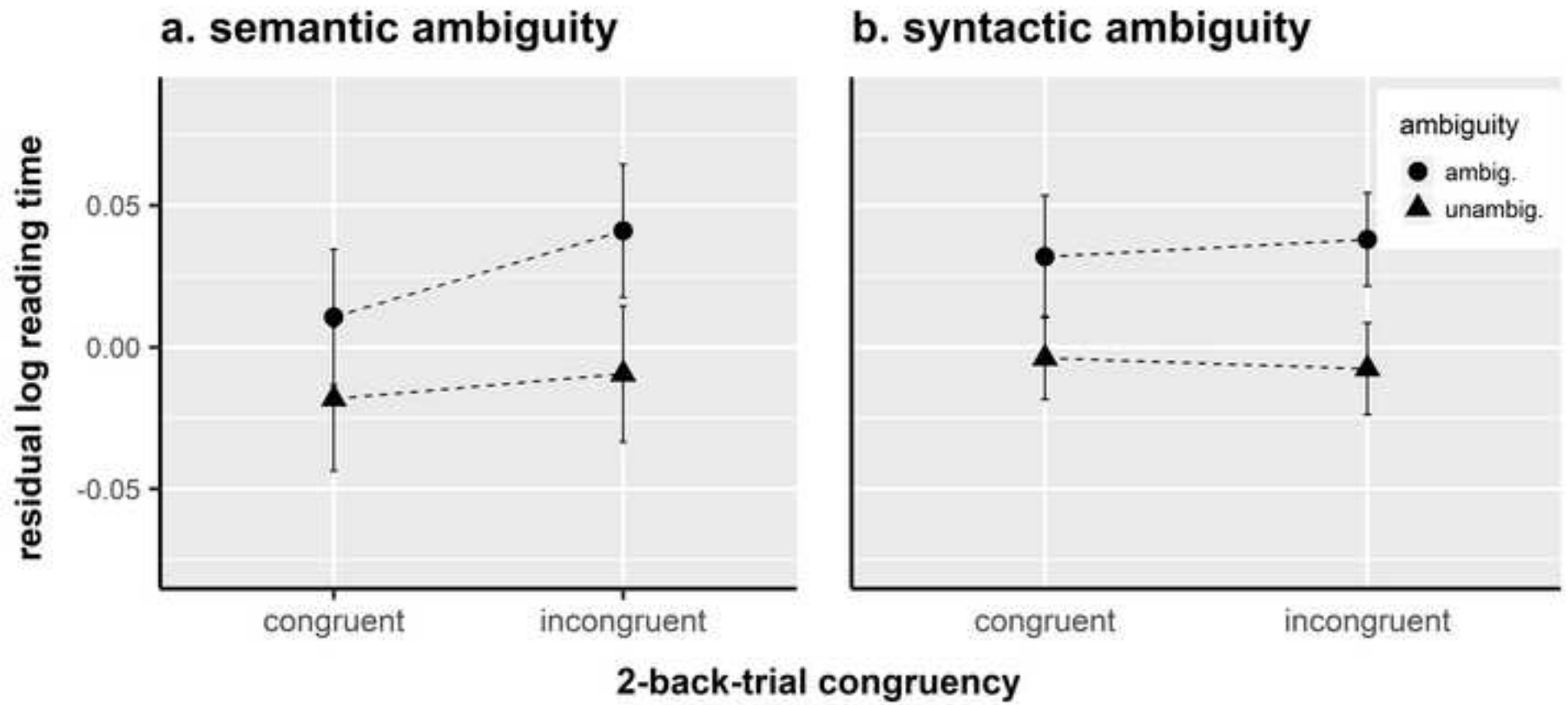


Figure9
[Click here to download high resolution image](#)



Appendix A

PWI target and distractor words in Experiment 1.

target	distractor	target	distractor	target	distractor	target	distractor
bone	muscle	bench	chair	*chain	rope	deck	patio
ear	nose	couch	futon	mallet	hammer	*fence	gate
elbow	knee	*desk	table	*match	lighter	shed	barn
*finger	thumb	*drawer	shelf	mop	broom	wheelbarrow	lawnmower
*foot	hand	accordion	bagpipe	nail	screw	arrow	harpoon
hoof	paw	banjo	guitar	*rake	shovel	rifle	pistol
horn	tusk	drum	bongo	saw	ax	rocket	bomb
whiskers	antennae	flute	oboe	wrench	pliers	sword	knife
*cat	dog	harmonica	tambourine	bicycle	scooter	anchor	mast
caterpillar	millipede	piano	keyboard	*bridge	tunnel	backpack	suitcase
*fly	bee	saxophone	trumpet	bus	car	baseball	tennis
*lion	tiger	violin	cello	*helicopter	airplane	*binoculars	microscope
raccoon	skunk	crown	tiara	kayak	rowboat	*book	movie
*snake	lizard	diamond	sapphire	*train	subway	*box	crate
swan	duck	medal	badge	van	truck	chimney	fireplace
turkey	chicken	necklace	earring	wheel	tire	flag	banner
comb	razor	cliff	hill	*apron	napkin	*globe	map
deodorant	shampoo	moon	sun	bowl	plate	hanger	clothespin
shower	bathtub	snow	rain	fridge	pantry	key	lock
*toilet	sink	tornado	hurricane	grill	stove	ladder	stairs
laptop	tablet	clipboard	notepad	pitcher	bottle	newspaper	magazine
speaker	microphone	glue	tape	spatula	ladle	*pipe	cigar
tv	radio	package	envelope	*spoon	fork	plug	outlet
typewriter	computer	paperclip	safety pin	toaster	microwave	pyramid	sphinx
*banana	apple	girl	boy	dress	skirt	robot	alien
carrot	radish	*nurse	doctor	*lipstick	mascara	*tent	cabin
onion	garlic	queen	knight	mitten	glove	towel	blanket
peanut	almond	sailor	pilot	*scarf	bandana	*treadmill	elliptical
*sandwich	hamburger	branch	stem	*shoe	boot	walker	cane
spaghetti	lasagna	sunflower	dandelion	vest	jacket	whip	lasso
waffle	pancake	tree	bush	watch	clock	yarn	twine
watermelon	grapefruit	wheat	hay	*zipper	button	yoyo	slinky

* These items were used as stimuli in Experiment 2.

Appendix B

Full results of RT models reported in the manuscript.

Table B1- Within-task (2-back) CSE in the PWI task in Experiment 1

Fixed effects	Coefficient	SE	t	p value
intercept	6.55	0.02	329.34	<0.001
current congruency	0.24	0.01	21.26	<0.001
2-back congruency	0.04	0.01	6.03	<0.001
current x 2-back congruency	-0.06	0.01	-5.97	<0.001
Random effects	Variance			
subject intercept	0.0113			
current congruency subject	0.0021			
2-back congruency subject	<0.0001			
current x 2-back congruency subject	0.0003			
item intercept	0.0036			
current congruency item	0.0046			
2-back congruency item	0.0004			
current x 2-back congruency item	0.0014			

Table B2 - Within-task (2-back) CSE in the prime-probe task in Experiment 1

Fixed effects	Coefficient	SE	t	p value
Intercept	6.06	0.03	175.21	<0.001
current congruency	0.18	0.02	11.13	<0.001
2-back congruency	0.03	0.01	4.27	0.005
current x 2-back congruency	-0.06	0.01	-5.69	<0.001
Random effects	Variance			
subject intercept	0.0226			
current congruency subject	0.0057			
2-back congruency subject	0.0005			
current x 2-back congruency subject	0.0012			
item intercept	0.0019			
current congruency item	0.0003			
2-back congruency item	0.0001			
current x 2-back congruency item	0.0002			

Table B3 - Cross-task (1-back) CSE in the PWI task in Experiment 1

Fixed effects	Coefficient	SE	t	p value
intercept	6.57	0.02	326.24	<0.001
current congruency	0.21	0.01	18.78	<0.001
1-back congruency	0.00	0.01	0.02	0.98
current x 1-back congruency	0.01	0.01	0.75	0.46
Random effects	Variance			
subject intercept	0.0114			
current congruency subject	0.0017			
1-back congruency subject	<0.0001			
current x 1-back congruency subject	0.0001			
item intercept	0.0051			
current congruency item	0.0068			
1-back congruency item	0.0015			
current x 1-back congruency item	0.0017			

Table B4 - Cross-task (1-back) CSE in the prime-probe task in Experiment 1

Fixed effects	Coefficient	SE	t	p value
intercept	6.08	0.03	185.79	<0.001
current congruency	0.14	0.01	10.40	<0.001
1-back congruency	-0.01	0.00	-2.87	0.03
current x 1-back congruency	0.01	0.01	2.33	0.05
Random effects	Variance			
subject intercept	0.0216			
current congruency subject	0.0041			
1-back congruency subject	<0.0001			
current x 1-back congruency subject	<0.0001			
item intercept	0.0016			
current congruency item	0.0002			
1-back congruency item	<0.0001			
current x 1-back congruency item	<0.0001			

Table B5- Within-task (2-back) CSE in the PWI task in Experiment 2

Fixed effects	Coefficient	SE	t	p value
intercept	6.57	0.02	276.57	<0.001
current congruency	0.18	0.02	9.74	<0.001
2-back congruency	0.05	0.01	4.20	<0.001
current x 2-back congruency	-0.06	0.01	-5.05	<0.001
Random effects	Variance			
subject intercept	0.014			
current congruency subject	0.0045			
2-back congruency subject	0.0006			
current x 2-back congruency subject	0.0006			
item intercept	0.003			
current congruency item	0.004			
2-back congruency item	0.0011			
current x 2-back congruency item	0.0003			

Table B6- Within-task (2-back) CSE in the reading task in Experiment 2

Fixed effects	Coefficient	SE	t	p value
intercept	-0.02	0.01	-3.13	<0.001
current ambiguity	0.06	0.01	6.04	<0.001
2-back ambiguity	0.02	0.01	2.13	0.03
current x 2-back ambiguity	-0.03	0.01	-2.06	0.04
Random effects	Variance			
subject intercept	<0.0001			
current ambiguity subject	<0.0001			
2-back ambiguity subject	<0.0001			

Table B7 - Cross-task (1-back) CSE in the PWI task in Experiment 2

Fixed effects	Coefficient	SE	t	p value
intercept	6.57	0.02	269.93	<0.001
current congruency	0.15	0.02	8.78	<0.001
1-back congruency	0.04	0.01	3.19	0.003
current x 1-back congruency	-0.01	0.01	-1.06	0.30
Random effects	Variance			
subject intercept	0.0146			
current congruency subject	0.0041			
1-back congruency subject	0.0013			
current x 1-back congruency subject	0.0001			
item intercept	0.0034			
current congruency item	0.0035			
1-back congruency item	0.0011			
current x 1-back congruency item	0.0015			

Table B8 - Cross-task (1-back) CSE in the reading task in Experiment 2

Fixed effects	Coefficient	SE	t	p value
intercept	-0.01	0.01	-1.28	0.2
current ambiguity	0.03	0.01	3.63	<0.001
1-back ambiguity	0	0.01	0.02	0.99
current x 1-back ambiguity	0.01	0.01	1.22	0.22
Random effects	Variance			
subject intercept	0.0005			
current ambiguity subject	0.0005			
1-back ambiguity subject	0.0004			

Appendix C

List of the sentences used in Experiment 2.

Semantically Ambiguous Sentences

Homographs are in bold. Regions of disambiguation and spillover are bracketed with square and curly brackets, respectively. Disambiguated sentences contained the parenthetical monosemic word instead of the homograph. Sentences 1–16 were adopted from Duffy, Morris, and Rayner (1988; non-equibiased sentences only). Sentences 17–64 were created from homographs in Nelson, McEvoy, Walling, and Wheeler (1980). Sentences not containing a highly biased homograph (i.e. > 0.9 for the most preferred meaning) also contained a semantic prime for the most preferred meaning.

1. Of course the **port** (soup) was a great success when she finally [served] {it to her} guests.
2. Once the **scale** (stone) was removed, the [dragon] {was} no longer in pain.
3. He found that the **table** (total) was too large after starting to [copy] {it into his} notebook.
4. The **bill** (hair) was not quite right although the head of the [animal] {had been} well sketched.
5. The old **pen** (zoo) was replaced because it was too [small to hold] {all} the new animals.
6. This time the **ball** (test) was moved because it was always so [well attended] {by the} students.
7. Usually the **bank** (edge) is not the place to start if you want to [catch a fish] {in this} stream.
8. Today the **wire** (pope) was [received] {with smiles} because it brought such good news.

9. Yesterday the **horn** (tail) was mounted on the wall after it was [cut off] {the dead} animal.
10. Actually the **yarn** (tale) was much too long given the [age of the children] {who were} listening.
11. Last year the **mint** (jail) was well advertised because it was [hiring] {twenty} new employees.
12. Of course the first **coach** (cabin) was rejected because [it was] {much too small} for them to sleep in.
13. Last night the **poker** (sword) was abandoned after [it accidentally] {fell} into the fire.
14. Of course the **boxer** (puppy) was exhausted by the time they got [it back] {on its} leash.
15. At last the **cabinet** (tourist) was finished after attentively [listening] {to the president's} speech.
16. Unfortunately the **band** (gold) was lost after [it suddenly] {fell} off her finger.
17. The artist **drew** (pulled) the table [out into] {the hall} to make room for his sculpture.
18. Although Mark said it was a **lot** (parking lot), it was technically a [garage] {because} of the roof.
19. Though Yusef claimed his contribution was **rare** (undercooked), it was the most [overcooked food] {at the potluck} that day.
20. Although the initial **story** (floor) was her favorite, the [museum's second floor] {was} the most popular.
21. Bernie didn't like the **game** (meat), even though it [tasted] {just like} chicken.
22. The fisherman thought the **net** (earnings) would be large enough, but his [profit estimates] {didn't account} for the tariff.

23. The professor cleaned her **mug** (face) every morning, scrubbing her [cheeks with soap] {and water.}
24. He struggled to conceal his **quiver** (bag), as it [contained arrows] {that were} quite long.
25. The carpenter loved his **drill** (routine), [in which] {he would practice} evacuating his shop.
26. The explorer approached the old **mine** (bomb) and [defused] {it before} continuing his journey.
27. The **count** (recitation) was always late, as the [kindergartener's voices] {were} totally out of sync.
28. Many people found the **tie** (score) off-putting, as the [football teams] {were} bitter rivals.
29. Even though Albert knew the business was **on the left** (liberal), their [public endorsement of Communism] {astonished} him.
30. It was a brand new **racket** (scheme), but the [criminals] {had years} of experience.
31. As the **shells** (bullets) all looked similar, they were probably [shot] {from} the same gun.
32. After putting on a **coat** (varnish), Andy realized that a darker [paint] {would have} looked better.
33. While the large **fan** (man) was loud, [his cheers were] {drowned} out by the roar of the crowd.
34. The **bluff** (mountain) deceived everyone, as it [looked] {much steeper} than it really was.
35. The cashier **charged** (attacked) Aaron, [knocking him] {to the ground} in anger.
36. Although the **shower** (ceremony) was long, the [expectant mother] {received} only four gifts.

37. Because the **runners** (blades) were out of shape, the [snow sled wobbled] {the entire} way down the hill.
38. The retiree often flaunted his **arms** (guns), as his large [collection of weapons] {usually} impressed his guests.
39. The **key** (island) was frequently used to [dock ships] {before voyages} to Australia.
40. The **bark** (tree) had captured Henry's attention, as the [tree] {was} unusually white.
41. The mathematician considered the **digit** (finger) and realized it was [broken] {after} the bookshelf fell on his hand.
42. The guests enjoyed the **jam** (music), even though the [musicians] {needed} more practice.
43. The former lawyer wandered into the **court** (arena), interrupting the [tennis match] {during} a serve.
44. When the roller coaster dropped, Megan was startled by the **jerk** (boy) because [he screamed] {in her} face.
45. The new **diamond** (stadium) impressed Mary, even though [baseball season] {was} months away.
46. The man assumed the **crooks** (turns) would straighten out, but the [road zigzagged] {for many} miles.
47. The **dive** (pub) was unrefined, but [its patrons] {were} loyal.
48. When Emily held up her **hand** (cards), her [royal flush] {took} everyone by surprise.
49. Jamie thought the **toast** (speech) was disgusting because the [groom only spoke] {about} himself.

50. Catherine examined the **page** (servant), then dismissed [him] {to his chambers} for the night.
51. Many **stars** (celebrities) are visible at night, especially in the [bars] {on Sunset} Boulevard.
52. Martin used the **log** (logarithm) to [calculate] {the derivative} on the exam.
53. The **sock** (uppercut) that Jeremy received at the Christmas party [broke] {his jaw} in three places.
54. Quinn deliberately left the **pack** (group) at the hotel, as his [loud companions] {exhausted} him.
55. The **ruler** (emperor) was warped, but [his kingdom] {was} peaceful.
56. It was a tough year for the **bar** (exam), as no one [passed] {the essay} section.
57. Anna received a bad **check** (hit), but the [referee] {allowed} the hockey game to continue.
58. The teacher's **pupils** (eyes) were giving her trouble, but the [eye doctor] {hesitated} to write her a new lens prescription.
59. The prisoner was going to use his **spade** (ace), but he played his [flush of hearts] {and won} the round.
60. While Owen had considerable knowledge of **planes** (triangles), it was the only [concept of geometry] {he truly} understood.
61. Reggie enjoyed his first **date** (raisin) so much that he [ate] {seven} more.
62. The gardener despised the **plant** (refinery), as it [manufactured] {many} chemicals harmful for the environment.

63. Before the speech concluded, the **speaker** (PA system) went silent for a moment until [it was] {plugged} back in.
64. The journalist reviewed the **draft** (selections), scrutinizing each [baseball player] {chosen} by the Orioles.

Syntactically Ambiguous Sentences

Disambiguated versions of each sentence contained the parenthetical phrase.

Homographs are in bold. Regions of disambiguation and spillover are bracketed with square and curly brackets, respectively.

Prepositional phrase attachment ambiguity.

Ferreira and Clifton (1986).

65. Katie laid the dress (that was) on the floor [onto the] {bed.}
66. George placed the record (that was) on the shelf [onto the] {turntable.}
67. Leslie positioned the dress (that was) on the rack [onto the] {display.}
68. Laura dragged the doll (that was) behind the bed [into the] {closet.}
69. The clerk put the receipt (that was) in the bag [into her] {hand.}
70. Mary set the flowers (that were) on the table [onto the] {cabinet.}
71. The sheriff locked the suspect (that was) in his office [into the] {jail} cell.
72. Sam loaded the boxes (that were) on the cart [onto the] {van.}

Noun-noun–relative clause ambiguity.

Grodner et al. (2002).

73. The alley (which) mice run rampant in [is damp] {and dimly} lit.
74. The kitchen (which) lamps shine brightest in [is one] {with white} tile.

75. The river (which) kayaks float slowly down [is broad] {and contains} a large volume of water.
76. The highway (which) billboards are placed along [gets extremely congested] {during rush} hour.
77. The jacket (which) pockets are sewn on [keeps your] {hands warm} though isn't very fashionable.
78. The restaurant (which) tables are placed behind [is trying] {to gain} more business with outside seating.
79. The juice (which) blenders are corroded by [is highly] {acidic} and can cause stomach problems.
80. The sidewalk (which) stones are piled near [will be] {torn up} by construction workers.

Past participle–past-tense verb ambiguity.

Tabor, Galantucci, and Richardson (2004).

81. We saw a movie about an artist (who was) [painted] {a picture} by her father.
82. One should respect a man (who was) [told] {his sins} by his own god.
83. The foreman yelled at a carpenter (who was) [cut] {a board} by his buddy.
84. The manager watched a waiter (who was) [served] {pea soup} by a trainee.
85. The prophet spoke of a man (who was) [planted] {a tree} by his daughter.
86. The agent photographed the man (who was) [recognized] {the previous} day by the spy.
87. The activist admired the speaker (who was) [proposed] {the first} time by the group.
88. Joseph forgot about the mailman (who was) [expected] {the next} day by the secretary.

Lexical category ambiguity.

Macdonald (1993).

89. The townspeople were pleased that the new prison guards (prisons guard) [the community] {from dangerous} criminals.
90. The doctor refused to believe that the miracle cures (miracles cure) [people] {of many} fatal diseases.
91. It says in the manual that the computer programs (computers program) [the printer] {to use} wide margins.
92. The efficiency experts reported that the office supplies (offices supply) [more] {than their} share of effort.

Trueswell, Tanenhaus, and Garnsey (1994).

93. The workers (who were) lifted [by the crane] {were} deposited on the roof.
94. The troops (who were) attacked [by the terrorists] {suffered} heavy losses.
95. The speaker (who was) proposed [by the group] {would} work perfectly for the program.
96. The teacher (who was) loved [by the class] {was} very easy to understand.

Direct object–subordinate clause ambiguity.

Christianson, Hollingworth, Halliwell, and Ferreira (2001).

97. While the man hunted (the deer) [ran] {into} the woods.
98. While the skipper sailed (the boat) [veered] {off} course.
99. As Henry whittled (the stick) [broke] {in} half.
100. While Rick drove (the car) [veered] {into} a ditch.
101. As the man walked (the poodle) [barked] {loudly} at him.
102. As the cowboy rode (the horse) [sweated] {profusely} and neighed.
103. While the chef stirred (the soup) [boiled] {vigorously} on the stove.
104. As Bill ate (the turkey) [sat] {on} the table.

Direct object–sentential compliment ambiguity.

Garnsey, Pearlmutter, Myers, and Lotocky (1997).

105. The scuba diver discovered that the wreck [was caused] {by a massive} collision.
106. The CIA director confirmed that the rumor [should have been] {stopped} sooner.
107. The trained referees warned that the spectators [would probably] {get} too rowdy.
108. The primary suspect established that the alibi [had been] {a total} lie.

Van Dyke and Lewis (2003).

109. The greedy dictator denied that the law [was justified] {by the crisis.}
110. The math professor proved that the theorem [was easy] {for the students.}
111. The television anchorman reported that the story [had broken] {this} morning.
112. The political scientist read that the book [was banned] {in Russia.}

Determiner–complementizer ambiguity.

Gibson (2006).

113. That popular [articles] {might be} plagiarized makes publishers nervous.
(That popular [article] {might be} plagiarized by an unethical author.)
114. That historical [novels] {would bring} the author acclaim was certain.
(That historical [novel] {would bring} the author acclaim and money.)
115. That strong [drugs] {would help} the patient was the doctor's opinion.
(That strong [drug] {would help} the patient if the doctor was correct.)
116. That defective [computers] {should be} replaced was the customer's claim.
(That defective [computer] {should be} replaced to keep the customer satisfied.)
117. That large [hedges] {should be} kept trimmed motivated the gardener.
(That large [hedge] {should be} kept trimmed according to the gardener.)

118. That large [donations] {are likely} to save the church encouraged the priest.
(That large [donation] {is likely} to save the church from being torn down.)
119. That lousy [scripts] {could probably} be revised gave the director hope.
(That lousy [script] {could probably} be revised with the director's help.)
120. That illegal [warrants] {were not} fair or just guided the judge's ruling.
(That illegal [warrant] {was not} fair or just in the mind of the judge.)

Active–reduced relative clause ambiguity.

Ferreira and Clifton (1986).

121. The man expected to die (but) [would not] {give} up easily.
122. The horse raced past the barn (and) [fell] {in a puddle.}
123. The woman told the joke (but) [didn't] {think} it was funny.
124. The man ordered the drink (but) [refused] {to drink} it.
125. The woman paid the money (and) [left] {the store} immediately.
126. The union sued for damages (but) [didn't] {expect} the settlement to be large.
127. The company awarded the contract (and) [was] {anxious} for the project to begin.
128. The troll brought the princess (and) [thought] {she} looked good enough to eat.