Acta Psychologica 174 (2017) 9–16

Contents lists available at ScienceDirect

Acta Psychologica

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

An attentional mechanism for minimizing cross-modal distraction

Lauren D. Grant *, Daniel H. Weissman

Department of Psychology, University of Michigan, USA

ARTICLE INFO

Article history:
Received 12 September 2016
Accepted 9 January 2017
Available online 23 January 2017

Keywords:
Congruency sequence effect
Cross-modal distraction
Contingent attentional capture

ABSTRACT

Prior findings suggest that coping with distraction relies on cognitive control processes that increase attention to task-relevant processing and/or decrease attention to task-irrelevant processing. In line with this view, the size of the congruency effect in unimodal Stroop-like tasks, a popular measure of distraction, is typically reduced after more distracting incongruent trials relative to after less distracting congruent trials. It remains unclear, however, whether, and under what conditions, the control processes underlying this congruency sequence effect (CSE) minimize cross-modal distraction. The contingent attentional capture hypothesis predicts a cross-modal CSE when a distracter possesses a target-defining feature. In contrast, the perceptual conflict hypothesis predicts a cross-modal CSE when there is perceptual conflict between a distracter and a target. To distinguish between these hypotheses, we conducted two experiments wherein an auditory distracter word preceded a visual target that appeared in one of two formats (i.e., word or arrow). We observed robust, cross-modal CSEs. Moreover, the pattern of CSEs that we observed was more consistent with the contingent attentional capture hypothesis than with the perceptual conflict hypothesis. These findings reveal a novel attentional mechanism for minimizing cross-modal distraction.

1. Introduction

The ability to minimize distraction is crucial for completing everyday tasks. For example, consider an individual who is reading by mentally pronouncing each word. This individual may become distracted by two people speaking nearby, especially if those people say the words he or she is reading. Therefore, to read successfully, this individual must minimize distraction from the irrelevant conversation.

1.1. Laboratory studies of distraction

In the laboratory, researchers investigate distraction using distracter interference tasks such as the flanker, Stroop, Simon, and prime-probe tasks (Eriksen & Eriksen, 1974; Stroop, 1935; Simon & Rudell, 1967; Eriksen & Schultz, 1979). In each trial of such tasks, participants respond to a target while ignoring a distracter. For example, in the "directional word" version of the prime-probe task, participants indicate the direction specified by a target word (e.g., left, right, up, or down) while ignoring a preceding distracter word (e.g., left, right, up, or down) (Schmidt & Weissman, 2014). In congruent trials, the distracter and target indicate the same direction and thus engender the same response. In incongruent trials, they indicate different directions and thus engender different responses. Typically, participants are slower to respond in incongruent than in congruent trials. This phenomenon, which indexes distractibility, is called the *congruency effect*.

Interestingly, the congruency effect is smaller when the previous trial was incongruent than when it was congruent. This *congruency sequence effect* (CSE) is thought to index a mixture of top-down attentional control processes that minimize distraction (e.g., Gratton, Coles, & Donchin, 1992; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Kerns et al., 2004) and feature integration and contingency learning processes that are often confounded with trial congruency (Hommel, Proctor, & Vu, 2004; Mayr, Awh, & Laurey, 2003; Schmidt & De Houwer, 2011). However, the use of confound-minimized paradigms has revealed robust CSEs without such confounds (Schmidt & Weissman, 2014; Weissman, Egner, Hawks, & Link, 2015; Weissman, Jiang, & Egner, 2014). Researchers may therefore use these paradigms to investigate the "control-driven" component of the CSE.

1.2. A role for contingent attentional capture in triggering the CSE

Recent findings indicate that contingent attentional capture plays a key role in triggering the control-driven CSE. Contingent attentional capture is a phenomenon in which distracters that possess target-defining perceptual and/or categorical features involuntarily capture attention (Moore & Weissman, 2010; Serences et al., 2005). For instance, when participants are asked to identify red targets that appear at a pre-specified location, they are more highly distracted by red distracters that appear at a different location than by other-colored (e.g., blue) distracters (Folk, Leber, & Egeth, 2002). Such capture is thought to occur because participants maintain perceptual and/or categorical

* Corresponding authors at: Department of Psychology, 530 Church Street, Ann Arbor, MI 48109, USA.
E-mail addresses: ldgran@umich.edu (L.D. Grant), danweiss@umich.edu (D.H. Weissman).

http://dx.doi.org/10.1016/j.actpsy.2017.01.003
0001-6918/

representations of target-defining features in working memory. While the maintenance of such an *attentional set* effectively guides attention toward targets, it also permits irrelevant stimuli that possess target-defining features to attract, or capture, attention.

Given that distracters often possess target-defining features in studies of the CSE, Schmidt and Weissman (2015) hypothesized that contingent attentional capture plays a role in triggering this phenomenon. To test this hypothesis, the authors conducted two experiments. In each experiment, participants identified the direction specified by a target (left, right, up, or down) while ignoring the direction specified by a preceding distracter (left, right, up, or down). Critically, the authors varied whether a distracter possessed a target-defining feature. As we describe next, the findings from these experiments indicated an important role for contingent attentional capture in triggering the CSE.

In Experiment 1, the authors varied both the distracter's presentation format (word or arrow) and the target's presentation format (word or arrow) across four blocks. In two blocks, these stimuli appeared in the same format (i.e., both words or both arrows). In the other two blocks, they appeared in different formats (i.e., one word and one arrow). Thus, in each block, participants could adopt an attentional set for just one target format (e.g., words). The authors hypothesized that the distracter would possess a target-defining format, and thereby capture attention, only when it appeared in the same format as the target. Consistent with their hypothesis, the authors observed a larger CSE in blocks wherein the distracter and target appeared in the same format than in blocks wherein these stimuli appeared in different formats. In fact, the CSE was completely absent in the latter condition. These findings provided novel support for the view that contingent attentional capture plays a role in triggering the CSE.

In Experiment 2, the authors randomly varied the format of the distracter and target on a trial-by-trial basis. Since the target in each trial could be either a word or an arrow, participants had to adopt an attentional set for both target formats. The authors therefore hypothesized that a distracter in either format would possess a target-defining feature that captures attention. Consistent with their hypothesis, and unlike in Experiment 1, the authors observed equivalent, robust CSEs in the same and different format conditions. These findings provided further support for the view that contingent attentional capture triggers the CSE.

1.3. Does contingent attentional capture trigger a cross-modal CSE?

As in the example of reading described earlier, irrelevant distracters often appear in a different sensory modality than relevant targets. To our knowledge, however, researchers have yet to investigate whether control processes underlying the CSE minimize cross-modal distraction. Investigating this hypothesis is important for three reasons. First, multisensory interactions are linked to unique distraction effects (Zimmer, Roberts, Harshbarger, & Woldorff, 2010). Second, some methods for reducing unimodal distraction (e.g., increasing the perceptual load of a visual display) do not minimize cross-modal distraction (Tellinghuisen & Nowak, 2003). Third, such a finding would expand the domain over which control processes underlying the CSE are known to operate.

Based on our prior findings (Schmidt & Weissman, 2015), we hypothesize that the CSE will minimize cross-modal distraction when a distracter possesses a target-defining feature that captures attention. Along these lines, words that appear in the auditory modality share perceptual and/or categorical features (i.e., auditory-verbal representations) with words that appear in the visual modality. For example, recent findings from functional magnetic resonance imaging (fMRI) indicate that auditory voice-selective regions are activated during silent reading (Yao, Belin, & Scheepers, 2011). These findings suggest that the attentional set people adopt to read visual words includes auditory-verbal representations of those words. Thus, a word spoken in headphones (e.g., "Left") should capture attention if it is a potential target that can appear on a computer screen (e.g., "Left"). The *contingent attentional capture* hypothesis therefore predicts that auditory distracter words will trigger a CSE when participants adopt an attentional set that specifies analogous visual target words.

Interestingly, this hypothesis further predicts that auditory distracter words will not capture attention when participants can adopt a non-verbal attentional set for visual targets. For example, according to this hypothesis, auditory distracter words should not trigger a control-driven CSE when the visual target is always an arrow, because task-relevant arrows are represented with spatial, rather than with auditory-verbal, features (Miles & Proctor, 2011). In this situation, the attentional set that participants adopt to search for target arrows should not include auditory-verbal representations that are activated by auditory distracter words. Thus, auditory distracter words should not trigger a CSE when the visual target is always an arrow.

1.4. Does perceptual conflict trigger a cross-modal CSE?

Contrary to the contingent attentional capture hypothesis, the *perceptual conflict* hypothesis posits that the CSE is triggered by perceptual conflict between the distracter and the target in incongruent trials (Verbruggen, Notebaert, Liefooghe, & Vandierendonck, 2006). For instance, perceptual conflict may occur when an auditory distracter word and a visual target word differ, because these stimuli activate distinct auditory-verbal representations. However, such conflict is unlikely to occur between an auditory distracter word and a visual target arrow, because the auditory-verbal representations that are employed to identify an auditory word are unlikely to conflict at perceptual levels with the spatial representations that are employed to identify a visual target arrow. Thus, according to the perceptual conflict hypothesis, a cross-modal CSE should be observed when an auditory distracter and a visual target are both words, but not when the distracter is a word and the target is an arrow. While findings from unimodal visual-modality tasks weigh against the perceptual conflict hypothesis (Schmidt & Weissman, 2015), to our knowledge no prior work has investigated this hypothesis in a cross-modal context.

1.5. The present study

The goal of the present study was to investigate, and ultimately distinguish between, the contingent attentional capture and perceptual conflict hypotheses. To this end, we conducted two experiments. Both experiments were similar in design to those of Schmidt and Weissman (2015). In each experiment, however, the distracter and target appeared in different sensory modalities. The distracter was an auditory word ("Left", "Right", "Up", or "Down"). The target was either a visual word (Left, Right, Up, or Down) or a visual arrow that pointed left, right, up, or down.

In Experiment 1, we varied the target format (word or arrow) across blocks. Thus, in each block, participants could adopt an attentional set for just one target format (word *or* arrow). In this situation, auditory distracter words should possess target-defining perceptual or categorical features when the visual targets are analogous words but not when they are arrows. Similarly, perceptual conflict between the auditory distracter and the visual target should be greater in incongruent than in congruent trials when both stimuli are words, but not when the distracter is a word and the target is an arrow. Thus, both hypotheses predict the CSE will be larger in word-target trials than in arrow-target trials. The goal of Experiment 1 was to test this hypothesis and thereby establish whether a cross-modal CSE is observed only when auditory distracters activate the same auditory-verbal representations as potential visual targets. Such an outcome would reveal an important boundary condition for observing a cross-modal CSE.

In Experiment 2, we randomly varied the target format (word or arrow) on a trial-by-trial basis. Thus, in each block, participants were required to adopt an attentional set for both target formats (word *and* arrow). In this situation, auditory distracter words should always possess target-defining perceptual or categorical features. The contingent