Working memory load dissociates contingency learning and item-specific proportion-congruent effects

Giacomo Spinelli, Kesheni Krishna, Jason R. Perry, Stephen J. Lupker

University of Western Ontario

Running header: Working memory and proportion-congruent effects

Corresponding Author: Giacomo Spinelli

Department of Psychology, University of Western Ontario

London, Ontario, N6A 5C2, Canada

Phone number: +1-226-448-5291

E-mail: gspinel@uwo.ca

Word count: 20,251

Abstract

A consistent finding in the Stroop literature is that congruency effects (i.e., the color-naming latency difference between words presented in incongruent vs. congruent colors) are larger for mostly congruent items (e.g., the word RED presented most often in red) than for mostly incongruent items (e.g., the word GREEN presented most often in yellow). This “item-specific proportion-congruent effect” might be produced by a conflict-adaptation process (e.g., fully focus attention to the color when the word GREEN appears) and/or by a more general learning mechanism of stimulus-response contingencies (e.g., respond “yellow” when the word GREEN appears). Under the assumption that limited-capacity resources are necessary for learning stimulus-response contingencies, we examined the contingency-learning account using both Stroop and nonconflict (i.e., noncolor words written in colors) versions of a color identification task while participants maintained a working-memory (WM) load. Consistent with the contingency-learning account, WM load modulated people’s ability to learn contingencies in the nonconflict task. In contrast, across three experiments, WM load did not affect the item-specific proportion-congruent effect in the Stroop task even though we employed a design (the “two-item set” design) in which contingency learning should be the dominant process. These results imply that the item-specific proportion-congruent effect is not merely a by-product of contingency learning but a manifestation of reactive control, a mode of control engagement that may be especially useful when WM resources are scarce.

Keywords: contingency learning, conflict adaptation, working memory, proportion-congruent effect, Stroop

Working memory load dissociates contingency learning and item-specific proportion-congruent effects

In the Stroop task (Stroop, 1935), participants are instructed to name the ink color of a word while ignoring the word itself. The term “congruency effect” refers to the finding that responses to congruent items (e.g., the word RED in red color, RED_{red}) are typically faster (and often more accurate) than responses to incongruent items (e.g., the word RED in blue color, RED_{blue}). Among the numerous investigations of the mechanisms involved in resolving and managing interference in this task (for a review, see MacLeod, 1991), manipulating the proportion of congruent items is an approach which has gained increasing research interest. The typical result of these proportion-congruent manipulations is that situations in which the proportion of congruent items is high elicit larger congruency effects than do situations in which the proportion of congruent items is low, a finding known as the “proportion-congruent effect” (e.g., Crump, Gong, & Milliken, 2006; Jacoby, Lindsay, & Hessels, 2003; Logan & Zbrodoff, 1979; for a review, see Bugg & Crump, 2012).

The classic proportion-congruent paradigm involves manipulating the proportion of congruent items in a list-wide fashion, allowing the comparison of performance on a list composed mainly of congruent items (a mostly-congruent list) with performance on a separate list composed mainly of incongruent items (a mostly-incongruent list). As noted above, larger congruency effects are generally obtained for the mostly-congruent list than for the mostly-incongruent list (e.g., Logan & Zbrodoff, 1979). The traditional explanation that has been offered for these proportion-congruent effects posits that attention to the task-relevant (i.e., the ink color) and task-irrelevant (i.e., the written word) dimensions is adjusted in response to the frequency of conflict from the task-irrelevant dimension (the *control account*: e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Bugg, Jacoby, & Toth, 2008). A situation in which conflict is frequent (i.e., a mostly-incongruent list) poses regular demands for the cognitive control system to adapt to the situation by directing attention to the relevant dimension. Interference from the irrelevant dimension will thus be minimized. On the other hand, a situation in which conflict is infrequent (i.e., a mostly-congruent list) biases attention toward the irrelevant dimension. As a result, interference from the irrelevant dimension on the few incongruent items will be especially problematic, a situation which typically results in a large congruency effect.

More recently, however, Jacoby et al. (2003) designed a new version of this paradigm that poses a challenge to the idea that proportion-congruent effects are due to the implementation of a list-wide, expectancy-based process as posited by the traditional control account. What Jacoby et al. demonstrated was an item-specific proportion-congruent effect. In their manipulation (the "two-item set" design), two color words (e.g., RED and BLUE) were presented mainly in their congruent color (mostly-congruent items, e.g., RED_{red} appearing more often than RED_{blue}) and two other color words (e.g., GREEN and YELLOW) were presented mainly in an incongruent color (mostly-incongruent items, e.g., $\text{GREEN}_{\text{yellow}}$ appearing more often than $\text{GREEN}_{\text{green}}$). The two sets of words were not permitted to cross (e.g., GREEN and YELLOW never appeared in either red or blue ink), and the two sets were intermixed such that in the list as a whole congruent and incongruent items were equally probable. Similar to the list-wide proportion-congruent effect, an item-specific proportion-congruent effect emerged, with a larger congruency effect for the mostly-congruent items than for the mostly-incongruent items. Because congruent and incongruent items were equally probable in the list as a whole, whatever process was being used that led to the item-specific proportion-congruent effect could not have been one that was based on the overall congruency proportion of the list. Rather, this process must have been an item-specific one, based on the congruency proportion assigned to each item in the list, that is, a process that is initiated in response to the nature of the specific item appearing on a given trial.

The control account of the item-specific proportion-congruent effect

The presence of an item-specific proportion-congruent effect has led researchers in the area of cognitive control to reconsider the original idea that *adaptation to conflict frequency*, or *conflict adaptation*, is the result of a single process of conflict-triggered adjustment (e.g., Botvinick et al., 2001). Although a more general conflict-adaptation account could potentially explain both list-wide and item-specific proportion-congruent effects (Bugg & Crump, 2012), the two effects are now thought to involve distinct processes of control engagement (Gonthier, Braver, & Bugg, 2016). A useful framework for interpreting these effects is the Dual Mechanisms of Control (DMC) account (Braver, 2012; Braver, Gray, & Burgess, 2007; see also Bugg & Crump, 2012), an account that, although somewhat more general, has many commonalities with an earlier account of Stroop interference (Kane & Engle, 2003).

The DMC framework proposes that control is engaged via two operating modes, proactive and reactive (roughly equivalent to Kane & Engle's, 2003, notions of "goal maintenance" and "conflict resolution", respectively). The proactive mode involves effortful, sustained maintenance of task-relevant items or goals in working memory (WM). For example, in the context of the Stroop task, participants might effortfully maintain the goal of naming colors and ignoring words throughout the task. Although such a process could be used in any situation in the Stroop task, its use would be favored in situations which repeatedly reinforce task relevance, e.g., in a mostly-incongruent list in the list-wide proportion-congruent manipulation. In this situation, because frequent conflict is expected between the word and the color, individuals would be prone to engage in a proactive process that minimizes interference from the word by constantly maintaining focus on the color-naming goal.

In contrast, the reactive mode relies on the stimuli in the environment for re-activation of task-relevant items or goals. The reactive mode can take more than one form. A basic form of reactive control is a process whereby the task goal is re-activated upon detection of a conflict between task-relevant and task-irrelevant dimensions (e.g., the color-naming goal is re-activated upon presentation of an incongruent word-color pair; Braver, 2012). This process would be favored in situations which rarely reinforce task relevance, e.g., a mostly-congruent list in the list-wide proportion-congruent manipulation. In this situation, because conflict between the word and the color is not expected, individuals would be prone to engage in a reactive process whereby the color-naming goal is frequently neglected and is only retrieved upon presentation of the infrequent incongruent words.

Reactive control can also take the form of a process that uses information about the stimulus to select a specific control process for dealing with that stimulus. This process is especially relevant in item-specific proportion-congruent manipulations. The reason is that these manipulations put participants in a situation in which they can use associations between words and their congruency to select the control process (e.g., relaxed vs. focused attention to the color) that would be best to apply to the presented word. Because those associations can only be used after the word has been presented, the use of those associations would require a form of reactive control. In this form of reactive control, early processing of specific words would regulate recruitment of appropriate control processes (Shedden, Milliken, Watter, & Monteiro, 2013; for a computational model of this mechanism, see Blais, Robidoux, Risko, & Besner, 2007). Specifically, the recognition of a mostly-incongruent word, e.g., GREEN, for example, may initiate a reactive control process favoring inhibition of word reading, with the result being reduced interference for this type of word. On the other hand, the recognition of a mostly-congruent word, e.g., RED, may initiate a reactive control process leading to relaxed attention, thus encouraging word processing in spite of the color-naming goal. The result will be large interference in the few instances in which the mostly-congruent word does conflict with the color (e.g., the word is RED but its color is blue rather than its usual red color).

Two aspects of the DMC account are worth noting. First, proactive and reactive modes of control are assumed to be partially independent, as demonstrated by their distinct neural signatures (e.g., Burgess & Braver, 2010; De Pisapia & Braver, 2006; Marini, Demeter, Roberts, Chelazzi, & Woldorff, 2016) and the fact that experimental manipulations can bias use of one or the other mode (for a review, see Braver, 2012). However, this account concedes that successful behavior likely depends on a mixture of proactive and reactive control engagement. For example, in the item-specific proportion-congruent paradigm, reliance on a reactive process regulating control in an item-specific manner would not necessarily prevent other processes from being invoked, although such processes may not be particularly encouraged by the task context. In particular, because in a typical item-specific proportion-congruent manipulation congruent and incongruent items are equally probable in the list as a whole, a proactive process of maintaining the task goal should not be encouraged to the same extent as it should be in a situation in which incongruent items are very frequent (i.e., a mostly-incongruent list in a list-wide proportion-congruent manipulation). Nonetheless, at least some individuals in an item-specific proportion-congruent manipulation might prefer to engage in this sort of proactive process instead of applying, or while concurrently applying, a reactive process of adaptation to item-specific conflict frequency. In sum, it is reasonable to hypothesize that reactive control may be the more prominent process, but not necessarily the only process that individuals could employ in an item-specific proportion-congruent manipulation.[1]

The second aspect worth nothing about the DMC account is that this account does not necessarily negate the possibility that non-control processes may also have an important role in phenomena such as the item-specific proportion-congruent effect. Indeed, Bugg and colleagues (Bugg, 2015; Bugg et al., 2011; Bugg & Hutchison, 2013) have proposed that the item-specific proportion-congruent effect reflects the action of a control-based process only when this effect is obtained in circumstances that prevent learning of associations between task-irrelevant information and responses (i.e., contingency learning, reviewed in the next section). When the experimental situation favors learning of contingencies, the item-specific proportion-congruent effect has been argued to mainly reflect the action of that (non-control) learning process instead.

The contingency-learning account of the item-specific proportion-congruent effect

Although control accounts have had good success in explaining data in interference tasks, recent years have witnessed a growing concern among researchers about the validity of conflict adaptation as an explanation for proportion-congruent effects (Schmidt, 2013b; Schmidt, Notebaert, & van den Bussche, 2015). This concern is motivated by the realization that, in speeded tasks, responding might be influenced by learning associations, or contingencies, between a stimulus and a motor response as opposed to learning associations between a particular word and a control process (Schmidt, Crump, Cheesman, & Besner, 2007). In nonconflict color identification tasks, contingency learning had been demonstrated by the finding that color identification is faster for a frequent word-color pair (= high-contingency item, e.g., the word BRAG presented in green color 75% of the time) than for an infrequent word-color pair (= low-contingency item, e.g., the word BRAG presented in yellow color 25% of the time). This effect, which is found for color words and color-unrelated words alike (Hutchison, 2011; Schmidt et al., 2007; see also Musen & Squire, 1993), is thought to reflect the fact that participants implicitly learn that specific words predict specific color responses (e.g., BRAG predicts green; Schmidt et al., 2007; see also Forrin & MacLeod, 2017; Lin & MacLeod, 2017).

Contingency learning provides a potential alternative explanation for proportion-congruent effects since manipulating the proportion of congruent items in the Stroop task typically involves altering the frequency of specific word-color pairs as well. Consider an item-specific proportion-congruent manipulation as an example. If the mostly-incongruent word GREEN appears most often in yellow, individuals may learn to associate the word GREEN with the (incongruent) yellow response. Conversely, if the mostly-congruent word RED appears most often in red, that would allow participants to learn that RED predicts the (congruent) red response. Crucially, if frequent word-color pairs elicit faster responses, relatively fast responding to the high-contingency incongruent item GREEN$_{\text{yellow}}$ will lead to a relatively small congruency effect for mostly-incongruent items, whereas fast responding to the high-contingency congruent item RED$_{\text{red}}$ will lead to a relatively large congruency effect for mostly-congruent items. Similar observations can be made for list-wide proportion-congruent manipulations (Schmidt, 2013b). This explanation, known as the *contingency-learning account* of proportion-congruent effects, suggests that learning of word-color contingencies, rather than adaptation to conflict frequency via control processes, might be responsible for the difference in the magnitude of congruency effects that is typically found in proportion-congruent manipulations in the Stroop task (Schmidt & Besner, 2008). Essentially, the item-specific proportion-congruent effect would have “everything to do with contingency” (Schmidt & Besner, 2008, p. 514).

Is control involved in the item-specific proportion-congruent effect?

The control account and the contingency learning account of proportion-congruent effects are fundamentally different in that the former invokes an interference-driven mechanism of conflict adaptation whereas the latter argues for a facilitative mechanism where conflict plays no role in modulating the congruency effect. Although conflict-adaptation and contingency-learning mechanisms are not necessarily mutually exclusive and could be integrated within a common theoretical framework (Abrahamse, Braem, Notebaert, & Verguts, 2016; Egner, 2014), in recent years there has been a debate about whether contingency learning alone may be a sufficient explanation for proportion-congruent effects, that is, whether these effects can be explained by an account that does not require invoking a mechanism of adaptation to conflict frequency at all (e.g., Atalay & Misirlisoy, 2012,2014; Bugg, 2014; Bugg et al., 2011; Bugg & Hutchison, 2013; Hazeltine & Mordkoff, 2014; Hutchison, 2011; Schmidt, 2013a, 2013b, 2013c; Schmidt & Besner, 2008; Schmidt et al., 2014). More recently, however, some evidence has emerged suggesting that list-wide proportion-congruent effects do persist when controlling for both contingency learning (Bugg, 2014; Bugg & Chanani, 2011; Gonthier et al., 2016; Hutchison, 2011; Spinelli & Lupker, 2020; Spinelli, Perry, & Lupker, 2019) and learning of list-wide temporal expectancies, another nonconflict learning mechanism thought to contribute to generating list-wide proportion-congruent effects (Cohen-Shikora, Suh, & Bugg, 2019; Spinelli et al., 2019). These results support the claim that humans do have access to a proactive mechanism of adaptation to list-wide frequency of conflict (for counterarguments, see Schmidt, 2013c, 2014, 2017).

With respect to the item-specific proportion-congruent effect, however, the situation is a bit different. The fundamental difference between the control-based account and a contingency-learning account of the item-specific proportion-congruent effect is that the former assumes that participants in an item-specific proportion-congruent manipulation associate words with control processes (e.g., inhibit word reading upon presentation of the mostly-incongruent word GREEN) while the latter assumes that they associate words with specific responses (e.g., predict a yellow response upon presentation of the mostly-incongruent word GREEN). While both mechanisms might be used, researchers who have tried to directly dissociate the two accounts have mostly found support for contingency-learning processes (Hazeltine & Mordkoff, 2014; Schmidt, 2013a; but see Spinelli & Lupker, 2019). For example, Schmidt (2013a) constructed a Stroop task in which item-specific conflict frequency and contingency learning were manipulated partially independently. Using this design, he was able to compare mostly-congruent words and mostly-incongruent words on what were “contingency matched” incongruent trials. For example, the color blue was a low-contingency and equally probable color for both the mostly-congruent word RED and the mostly-incongruent word GREEN. According to the control-based account, because mostly-congruent words should induce relaxed attention whereas mostly-incongruent words should induce focused attention to the color, the mostly-congruent word RED should produce more interference than the mostly-incongruent word GREEN when those words are presented in blue. However, performance on mostly-congruent and mostly-incongruent words was equivalent when those words appeared in the critical incongruent colors, suggesting that no conflict-adaptation process was in use. Based on these results, Schmidt (2013a) concluded that contingency learning is the sole source of item-specific proportion-congruent effects, with conflict adaptation playing no role at all.

As noted, this conclusion has gained at least some credence even among proponents of control accounts (Bugg, 2015; Bugg & Hutchison, 2013; Bugg et al., 2011). Specifically, those researchers appear to have conceded that contingency learning, rather than control-based processes, does determine the modulations of the congruency effect that are observed in the item-specific proportion-congruent manipulation originally employed by Jacoby et al. (2003), i.e., the two-item set design. A control-based process would be used only in specific circumstances, for example, when contingency learning is discouraged by including words being associated with no specific response in the task (e.g., in a four-item set design in which mostly-incongruent words appear equally frequently in each of four colors, one congruent and three incongruent), or when the relevant dimension (i.e., the color), rather than the irrelevant dimension (i.e., the word), acts as the potent signal for conflict frequency (Bugg, 2015; Bugg & Hutchison, 2013; Bugg et al., 2011). Notably, the situation examined by Jacoby et al. (2003) would not be one of those circumstances (although see Hutcheon & Spieler, 2014, for evidence in support of a conflict-adaptation explanation of the item-specific proportion-congruent effect in Jacoby et al.'s two-item set design).

The present research

The present research was an attempt to re-examine the conclusion that the item-specific proportion-congruent effect in Jacoby et al.'s (2003) two-item set design is due to contingency learning by using a different approach than the ones used thus far. As noted above, the process of learning word-response associations is typically examined in a color identification task where noncolor words are presented mainly in one specific color (e.g., the word SHOP presented more often in blue than in red; Schmidt et al., 2007; Schmidt, De Houwer, & Besner, 2010). Schmidt et al. (2010) had participants perform this nonconflict color identification task while maintaining a low (e.g., remember 2 digits) or high (e.g., remember 5 digits) working-memory (WM) load. Crucially, they only found a significant contingency-learning effect for the low-load group. For example, in their Experiment 2, Schmidt et al. obtained a 107-ms contingency-learning effect for participants performing the color identification task with a low WM load. In contrast, participants who performed the color identification task with a high WM load were not only overall slower but also showed a smaller and nonsignificant 28-ms contingency-learning effect. Further, an impact of word-response contingencies was not observed when participants were required to carry a high WM load even when those contingencies had been successfully learned in an earlier block in which participants were required to carry a low WM load (Experiment 3). Based on these results, Schmidt et al. (2010) concluded that, even though it might be an implicit process, contingency learning is a resource-dependent process, such that limited-capacity resources are necessary for both learning and using contingencies.

Importantly, because contingency learning is independent from the interference caused by the stimuli being used (Levin & Tzelgov, 2016), the process of learning contingencies should have the same capacity limitations regardless of whether the stimuli are color or noncolor words. Based on the premise that contingency learning is the cause of the item-specific proportion-congruent effect, particularly in Jacoby et al.'s (2003) two-item set design, what Schmidt et al.'s (2010) results imply is that participants performing the Stroop task while carrying no WM load (i.e., the standard situation) or a low WM load should show a regular item-specific proportion-congruent effect, whereas little or no item-specific proportion-congruent effect would be expected for participants who perform the Stroop task while carrying a high WM load similar to the one Schmidt et al. used. In contrast, finding equivalent item-specific proportion-congruent effects in high, low, and no WM load situations would be problematic for the contingency-learning account.

It is worth noting that obtaining an item-specific proportion-congruent effect in a high WM-load condition would also be problematic for theories of cognitive control that assume that successful implementation of any control process is critically dependent on available attentional resources (e.g., Baddeley, Chincotta, & Adlam, 2004; Baddeley & Hitch, 1974). These types of accounts, just like the contingency-learning account, would also seem to predict that increasing WM load should lead to no item-specific proportion-congruent effect (i.e., the congruency effect should be the same for mostly-congruent and mostly-incongruent items). As will be described just below, however, the same would not be true for control accounts such as the DMC account (Braver, 2012; Braver et al., 2007; see also Kane & Engle, 2003) because accounts of this sort appear to predict that a higher WM load would not interfere with use of *reactive* control processes, i.e., the type of processes that would support a mechanism of adaptation to item-specific (as opposed to list-wide) conflict frequency, although it may interfere with *proactive* control processes, i.e., top-down control processes that are based on situational expectancies.

As noted, at least in some circumstances, the item-specific proportion-congruent effect has been claimed to result from the application of reactive control (Bugg, 2015; Bugg et al., 2011; Bugg & Hutchison, 2013), with recognition of a mostly-incongruent word leading to a focus of attention onto the task-relevant (color) dimension and recognition of a mostly-congruent word leading to a relaxation of attention to that dimension. What is important to note is that, according to the DMC account, there is no reason that WM demands would impact this type of reactive control in the same way that they would impact proactive control, as the two control processes appear to be dissociable. For example, in an fMRI memory study, Speer, Jacoby, and Braver (2003) found that an expected low WM load showed an activation pattern consistent with the idea that participants were using a proactive process of maintaining study items in memory in preparation for the upcoming probe. An expected high WM load, in contrast, showed an activation pattern consistent with the use of a reactive process, whereby study items were not actively maintained and the probe was used as a retrieval cue instead. Similar dissociations were obtained in behavioral and neuroimaging research analyzing individual differences in WM resources, typically defined in terms of WM capacity (Burgess & Braver, 2010; Hutchison, 2011; Kane & Engle, 2003).

In general, control accounts which distinguish proactive and reactive control processes appear to suggest that reactive control is, in fact, relatively easily implemented when WM resources are scarce (Braver, 2012; Braver et al., 2007). Importantly, what these ideas then imply concerning the impact of WM load on an item-specific proportion-congruent manipulation would seem to be somewhat different from the predictions made by a contingency-learning account. Specifically, assuming, as control accounts such as the DMC do, that the item-specific proportion-congruent effect is due, in whole or in part, to a reactive control process (i.e., adaptation to item-specific conflict frequency), no reduction in the proportion-congruent effect should be observed with increasing WM load (regardless of WM capacity). The reason is that having fewer available WM resources should make reactive control at least as prominent a process as it is in normal circumstances (i.e., when WM resources are not taxed by a concurrent task), with the result being a good-size proportion-congruent effect. In contrast, as discussed, the contingency-learning account would predict that if available WM resources are low due to a high concurrent WM load, contingency learning cannot take place, leading to a very reduced proportion-congruent effect.

The present research involved a number of experiments investigating the role of WM load in contingency-learning and item-specific proportion-congruent effects. Using vocal responses to the colors, Experiments 1A and 1B sought to replicate Schmidt et al.'s (2010) findings in the nonconflict color identification task and to expand them to the Stroop task using a two-item set design, i.e., the design that presumably favors use of contingency learning instead of conflict-adaptation processes (Bugg, 2014; Bugg & Hutchison, 2003). To preview, we were not able to replicate the original pattern in the nonconflict color identification task (i.e., contingency-learning effects did not diminish as WM load increased). Therefore, Experiments 2A and 2B used manual responses to the colors as well as providing feedback, as in the original article (Schmidt et al., 2010), a situation in which we were able to replicate Schmidt et al.'s (2010) findings for the nonconflict color identification task. However, we did not find a similar reduction in the item-specific proportion-congruent effect in the Stroop task. Finally, Experiments 3A and 3B replicated and expanded the previous results using a within-subject design. In addition, WM capacity for individuals in the no-load group was measured in Experiments 3A and 3B in order to explore the ideas that lower WM resources are associated with either a decrease in contingency-learning effects as proposed by the contingency-learning account or an increased reliance on reactive control, as proposed by the DMC account (Braver, 2012; Braver et al., 2007).

Experiment 1A & 1B (vocal responses)

Would taxing cognitive resources impair contingency learning in the nonconflict, as well as the Stroop, color identification task? To answer this question, in Experiment 1A participants were presented with contingency-biased noncolor words (e.g., the word SHOP presented 75% and 25% of the time in blue and red, respectively), whereas in Experiment 1B participants were presented with both mostly-congruent color words (e.g., the word RED presented 75% and 25% of the time in red and blue, respectively) and mostly-incongruent color words (e.g., the word GREEN presented 75% and 25% of the time in yellow and green, respectively) intermixed in the same list. In both experiments, a two-item set design was used, i.e., each word appeared in two colors only although, overall, four colors and four words were used. As mentioned, this design was used by Jacoby et al. (2003) and is supposed to promote learning of word-response contingencies as the dominant process for performance (Bugg, 2014; Bugg & Hutchison, 2013). In addition, in both experiments, one third of the participants performed the color identification task with no memory load (no-load group). The other two-thirds performed both the color identification task and a concurrent WM task which required holding in memory two digits (the low-load group) or five digits (the high-load group), as in Schmidt et al. (2010).

Colors were responded to vocally and participants received no feedback on their performance, whereas Schmidt et al. (2010) had participants respond to colors via button pressing and provided them with feedback (i.e., participants were warned when an error was made). However, Schmidt et al. provided no indication that response modality or feedback should matter in terms of the impact of WM load on contingency learning: As long as cognitive resources are properly taxed, one should obtain a reduction in contingency-learning effects under load.

Method

Participants

Sixty-one participants took part in Experiment 1A (nonconflict color identification task) and another 60 took part in Experiment 1B (Stroop task). These sample sizes were determined based on Schmidt et al.'s (2010) Experiment 2, in which 60 participants were tested. In Experiment 1A, 1 participant was removed because of an excessive number of errors and null responses (above 25%). In both experiments, the final 60 participants were equally distributed across the no-, low-, and high-load groups in each experiment (20 participants per group in each experiment). Participants were all students at the University of Western Ontario, aged 18–29 years and had normal or corrected-to-normal vision. Their participation was compensated with course credit or $10.

Materials

Four color-unrelated words (SHOP, CULT, BRAG, WIDE) and four color words (RED, BLUE, GREEN, YELLOW) were used as carrier words and four colors (red [R: 255; G: 0; B: 0], blue [R: 0; G: 112; B: 192], green [R: 0; G: 176; B: 80], and yellow [R: 255; G: 255; B: 0], corresponding to “red”, “blue”, “green” and “yellow” in the standard DMDX palette) were used as targets. Participants in Experiment 1A only saw color-unrelated words and participants in Experiment 1B only saw color words. The nature of the word-color combinations used is represented in Tables 1 and 2. Both noncolor and color words were divided into two sets, one set (e.g., SHOP and CULT for Experiment 1A, RED and BLUE for Experiment 1B) was only presented in red and blue ink colors, the other set (e.g., BRAG and WIDE for Experiment 1A, GREEN and YELLOW for Experiment 1B) was only presented in green and yellow ink colors. In Experiment 1A, the frequency of word-color combinations was manipulated so that each word was paired with one of the colors 75% of the time (thus creating a high-contingency item) and with the other color 25% of the time (thus creating a low-contingency item). In Experiment 1B, one set of words (e.g., RED and BLUE) was paired with the congruent color 75% of the time and with the incongruent color 25% of the time (i.e., serving as mostly-congruent items), while the other set of words (e.g., GREEN and YELLOW) was paired with the congruent color 25% of the time and with the incongruent color 75% of the time (i.e., serving as mostly-incongruent items). Assignment of words to the frequent and the infrequent color was counterbalanced across participants. Overall, congruent and incongruent items were equally probable in Experiment 1B. Both Experiment 1A and Experiment 1B included 192 trials.

-Tables 1 & 2 around here-

Procedure

Participants were randomly assigned to the no-load, low-load, or high-load group. Each trial began with a fixation symbol (“+”) displayed for 250 ms in the center of the screen followed by a 250-ms blank screen. For participants in the low- and high-load groups, this blank screen was followed by a set of two random digits (low-load; e.g., 3   2) or five random digits (high load; e.g., 3   2   4   1   7), presented with three spaces between each digit for 2000 ms. In the next display, a colored word appeared in uppercase Courier New font, pt. 14, displayed for 2000 ms or until the participant’s response, which was recorded with a microphone connected to the testing computer. Participants were instructed to name the color of the word as quickly and as accurately as possible while ignoring the word itself. Following an 800-ms blank screen, another set of two digits (for the low-load group) or five digits (for the high-load group) was presented flanked by two arrows on each side (e.g., >> 3   2 <<) for 2000 ms or until the participant’s response. In this probe set of digits, either a randomly selected digit in the memory set was changed to a new random digit or none of the digits were changed. Participants were required to press the right shift key if the probe set of digits was identical to the memory set of digits and the left shift key if the two sets of digits were different. Trials requiring “same” and “different” responses were equally probable, and this manipulation was orthogonal to the manipulations involving colored words (e.g., low- and high-contingency items appeared on trials requiring a “same” response as often as on trials requiring a “different” response, etc.).

Participants in the no-load group were only presented with the colored words, which were presented right after the fixation symbol. Stimuli were presented against a medium grey background (R: 169; G: 169; B: 169). No feedback was provided. The 192 trials were presented in two blocks of 96 trials each with a self-paced pause in the middle. The order of trials within each block was randomized. Prior to starting each block, participants performed a practice session of 16 trials mirroring the frequency of word-color combinations in that block. The experiment was run using DMDX (Forster & Forster, 2003) software. This research was approved by the Research Ethics Board of the University of Western Ontario (protocol # 108956).

Results

The waveforms of responses in the color identification task were manually inspected with CheckVocal (Protopapas, 2007) to determine the accuracy of the response and the correct placement of timing marks. Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than the time limit on either the color identification task or the WM task (accounting for 1.7% and 1.9% of the data points in Experiments 1A and 1B, respectively) were discarded. Trials on which participants responded incorrectly on the WM task (which accounted for 3.6% and 8.1% of the data points in the low- and high-load groups in Experiment 1A, and 4.0% and 9.2% of the data points in the low- and high-load groups in Experiment 1B) were discarded as well.[2] Latency analyses were conducted only on trials in which the response in the color identification task was also correct.[3]

Different analyses were performed for Experiment 1A and Experiment 1B due to the different nature of the stimuli (noncolor vs. color words) and design. For Experiment 1A, a 2 (Contingency: low vs. high, within-subjects) X 3 (WM Load: no vs. low vs. high, between-subjects) ANOVA was conducted. For Experiment 1B, the design of the ANOVA was a 2 (Congruency: congruent vs. incongruent, within-subjects) X 2 (Item Type: mostly congruent vs. mostly incongruent, within-subjects) X 3 (WM Load: no vs. low vs. high, between-subjects).[4] In addition to traditional null-hypothesis significance testing analyses, we also performed Bayes Factor analyses when a theoretically important null effect was obtained in order to quantify the evidence supporting the presence vs. the absence of that effect. These analyses were performed in R version 3.5.1 (R Core Team, 2018) using the BayesFactor package, version 0.9.12-4.2 (Morey & Rouder, 2018) by comparing the model without the effect of interest (interpreted as the null hypothesis H_0) and the model with that effect (interpreted as the alternative hypothesis H_1). One million iterations were used to evaluate each model. The result of this comparison was BF_{01}, with $BF_{01} < 1$ suggesting evidence in support of H_1 (i.e., the presence of the effect), whereas $BF_{01} > 1$ suggesting evidence in support of H_0 (i.e., the absence of the effect) ($BF_{01} = 1$ would suggest equal evidence for the two hypotheses). Jeffreys's (1961) classification scheme (as reported in adjusted form by Lee and

Wagenmakers, 2013) was used to help interpret the size of the Bayes Factor. The mean RTs and error rates are presented in Tables 3 and 4 for Experiments 1A and 1B, respectively. For this and for the following experiments, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (see above for this information for Experiments 1A and 1B; Simmons, Nelson, & Simonsohn, 2012). The raw data and the scripts used for the analyses are also publicly available at https://osf.io/rtnw2/.

-Tables 3 & 4 around here-

Experiment 1A (nonconflict color identification task)

RTs. Both the main effects of Contingency (high-contingency faster than low-contingency), $F(1, 57)$ = 8.40, MSE = 510, p = .005, η_p^2 = .128, and WM Load, $F(2, 57)$ = 13.45, MSE = 32148, p < .001, η_p^2= .321, were significant. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was faster than both the low-load group (p < .001) and the high-load group (p = .001), but the low-load and the high-load groups did not differ from one another (p = .487). Importantly, Contingency and WM Load did not interact (F = .19, MSE = 510, p = .825, η_p^2 = .007), with equivalent contingency learning effects in the no- (12 ms), low- (15 ms), and high-load groups (9 ms). The Bayes Factor for the comparison between the model with the interaction and the model without it was BF_{01} = 6.43 ± .83%, meaning that the data were 6.43 times more likely to occur under the hypothesis of no interaction than under the hypothesis of an interaction. In Jeffreys's (1961) classification scheme, this value would suggest "moderate" evidence for the absence of the interaction.

Error rates. No effect reached significance (all *F*s < 1).

Experiment 1B (Stroop task)

RTs. There were main effects of Congruency (congruent faster than incongruent), *F*(1, 57) = 99.64, *MSE* = 3754, $p < .001$, η_p^2 = .636, and WM Load, *F*(2, 57) = 9.94, *MSE* = 46916, $p < .001$, η_p^2 = .259. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was faster than both the low-load group (*p* = .022) and the high-load group ($p < .001$), but the low-load and the high-load groups did not differ from one another (*p* = .222). The only significant interaction was that between Congruency and Item Type, *F*(1, 57) = 56.18, *MSE* = 1765, $p < .001$, η_p^2 = .496, indicating that a regular item-specific proportion-congruent effect was found, with a larger congruency effect for mostly-congruent items (120 ms) than for mostly-incongruent items (38 ms). There was no three-way interaction between Congruency, Item Type, and WM Load, however, *F*(2, 57) = .230, *MSE* = 1765, *p* = .795, η_p^2 = .008, suggesting that the item-specific proportion-congruent effect was equivalent in all load groups. The Bayes Factor, BF_{01} = 7.14 ± 6.27%, indicated “moderate” evidence for the absence of the three-way interaction.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), *F*(1, 57) = 33.39, *MSE* = .001, $p < .001$, η_p^2 = .369, Item Type (mostly incongruent more accurate than mostly congruent), *F*(1, 57) = 12.71, *MSE* = .001, *p* = .001, η_p^2 = .182, and WM Load, *F*(2, 57) = 6.68, *MSE* = .001, *p* = .002, η_p^2 = .190. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the low-load group was more accurate than the no-load group (*p* = .002), but did not differ significantly from the high-load group (*p* = .065). The no-load and the high-load groups did not significantly differ from one another either (*p* = .389). An overall item-specific proportion-congruent effect was obtained, as shown by the significant interaction between Congruency and Item Type, *F*(1, 57) = 23.36, *MSE* = .001, $p < .001$, η_p^2 = .291. However, Congruency also interacted with WM Load, *F*(2, 57) = 4.82, *MSE* = .001, *p* = .012, η_p^2 = .145, and the three-way interaction was also significant, *F*(2, 57) = 4.08, *MSE* = .001, *p* = .022, η_p^2 = .125.

To explore the interactions involving WM Load, three additional ANOVAs were performed comparing every pair of load groups. Inspection of two-way interactions between Congruency and WM Load revealed smaller congruency effects for the low-load group (0.7%), than for either the no-load group (3.3%), *F*(1, 38) = 9.21, *MSE* = .001, *p* = .004, η_p^2 = .195, or the high-load group (2.4%), *F*(1, 38) = 7.56, *MSE* = .001, *p* = .009, η_p^2 = .166, whereas the no-load and high-load groups did not differ from one another, *F*(1, 38) = .60, *MSE* = .002, *p* = .442, η_p^2 = .016. Similarly, inspection of the three-way interaction between Congruency, Item Type, and WM Load revealed that the Congruency by Item Type interaction for the low-load group differed from those for both the no-load, *F*(1, 38) = 7.08, *MSE* = .001, *p* = .011, η_p^2 = .157, and high-load groups, *F*(1, 38) = 6.89, *MSE* = .000, *p* = .012, η_p^2 = .153, but no difference was found between the no-load and high-load groups, *F*(1, 38) = .39, *MSE* = .001, *p* = .538, η_p^2 = .010. Separate analyses for each load group showed the reason for this interactive pattern was that although significant Congruency by Item Type interactions (with larger congruency effects for mostly-congruent than mostly-incongruent items) were obtained for both the no-load, *F*(1, 19) = 10.90, *MSE* = .001, *p* = .004, η_p^2 = .365, and the high-load group, *F*(1, 19) = 13.56, *MSE* = .001, *p* = .002, η_p^2 = .416, there was no significant interaction for the low-load group, *F*(1, 19) = .82, *MSE* = .000, *p* = . 376, η_p^2 = .041. In general, it appears that low-load group did behave somewhat differently than the no-load and high-load groups. However, the most likely reason for this difference is not that there is a nonmonotonic impact of load on error rates but rather because of the very low number of errors (less than 1%) committed by participants in the low-load group.

Discussion

Experiments 1A and 1B were attempts to replicate Schmidt et al.'s (2010) findings from a nonconflict color identification task and to extend those findings to the Stroop task using vocal responses.

Surprisingly, however, the nonconflict color identification task (Experiment 1A) showed no impact of WM load on the magnitude of contingency effects, thus failing to replicate Schmidt et al. in a task requiring vocal responses (as opposed to manual responses as in Schmidt's original article). Similarly, WM load did not alter the magnitude of item-specific proportion-congruent effects in the Stroop task (Experiment 1B) either with the exception of the accuracy data for the low-load group in Experiment 1B. That group produced no item-specific proportion-congruent effect in the accuracy data but also very few errors in general, suggesting that their accuracy data may reflect a floor effect and, hence, should be interpreted extremely cautiously. Although this pattern of results supports the idea that item-specific proportion-congruent effects in the Stroop task and contingency-learning effects in the nonconflict color identification task follow the same pattern, potentially due to the fact that they are the result of the same process, the fact that increasing WM load had no effect on the size of contingency-learning effects is problematic for the assumption that contingency learning depends on limited-capacity resources, an assumption that is a basic premise of the present research (Schmidt et al., 2010).

In trying to understand the pattern of data in Experiment 1, two observations are in order. First, the WM load manipulation was effective: Latencies in the color identification task were faster for the no-load group than for the other groups (although this difference was compensated for by the drop in error rates for the low-load group in Experiment 1B), and the high-load memory task elicited more errors than the low-load memory task did. Given also that the memory task was identical to the one Schmidt et al. (2010) used, it would appear that the reason for the discrepancy between the present results and Schmidt et al.'s would seem to have little to do with the way we employed the WM-load manipulation.

Second, the contingency-learning effect in Experiment 1A, a nonconflict color identification task requiring vocal responses, was small (12 ms in the no-load condition) compared to what is typically reported in the literature, where manual-responding versions of the task are prevalent (e.g., Schmidt et al., 2007, reported a 60-ms contingency-learning effect with a design similar to the one used here using a manual-responding procedure). If response modality is responsible for this difference, this specific pattern of results is actually somewhat surprising based on findings from the Stroop task suggesting that vocal responding may favor processing of the word (Melara & Mounts, 1993; Virzi & Egeth, 1985). If word processing is enhanced because of the use of the vocal response mode, it would seem that contingencies between words and responses should be learned more effectively, with the likely result being, if anything, *larger* contingency-learning effects with vocal than manual responding.

Reduced contingency learning effects for vocal responding are more easily reconciled with a view that emphasizes the role of compatibility between relevant stimuli and responses in contingency learning, i.e., the degree to which responses map readily onto relevant stimuli (Schmidt, 2018). According to this view, although contingencies may be efficiently learned in both vocal and manual responding situations, contingency learning will have a smaller impact on performance when the requested response is relatively compatible with the stimulus (e.g., a vocal response, an overtrained response for a color) than when the requested response is relatively incompatible with the stimulus (e.g., a keypress response, an undertrained response for a color). The reason is that, because contingency learning operates at the response stage (Schmidt et al., 2007), this process will have a smaller window for influencing behavior when stimuli can be quickly translated into compatible (vocal) responses than when they are more slowly translated into incompatible (manual) responses. As a result, contingency learning will be reduced in a vocal responding situation.

Vocal-responding and manual-responding contingency-learning paradigms, however, typically do not differ only in the type of response that is required but also in whether responding is assisted with feedback, which is often absent with vocal responses but present with keypress responses. Indeed, a more complicated story emerged when we tried to address this concern in a series of nonconflict color identification tasks requiring vocal versus manual responses with or without feedback (Spinelli et al., under review). What we found was that the presence of feedback was crucial in order to observe a larger contingency-learning effect for manual than for vocal responding. When no feedback was given, contingency-learning effects were equivalent across response modalities. Specifically, removing feedback reduced contingency learning in manual responding to the size of the contingency learning effect in vocal responding but had no impact on the (small) contingency learning in vocal responding.

Although the reasons for these results can be complex (for a discussion, see Spinelli et al., under review), the crucial message for the present research is that manual responses with feedback might be the only situation producing a good-size contingency-learning effect. Other situations, including vocal responses without feedback (the situation of Experiments 1A and 1B) might produce such small contingency-learning effects that observing a significant reduction in their size might be challenging (for a similar point, see Kinoshita et al., 2018). Insofar as manual responding plus feedback elicits larger baseline contingency-learning effects, this situation might not only provide a more direct replication of Schmidt et al. (2010), but also be more appropriate for testing the idea that WM load impairs the process of learning contingencies. This hypothesis provides the motivation for Experiments 2A and 2B.[5]

Experiments 2A & 2B (manual responses)

Experiments 2A and 2B were identical to Experiments 1A and 1B, except that manual responding to colors along with feedback on each trial was used. We reasoned that this change would not only allow us to replicate Schmidt et al.'s (2010) original experiment more closely but also increase the size of baseline contingency-learning effects, thus providing a better opportunity to observe modulations of such effects.

Method

Participants

Sixty-three participants took part in Experiment 2A (nonconflict color identification task) and another 63 took part in Experiment 2B (Stroop task). In both Experiment 2A and Experiment 2B, 3 participants were removed because of an excessive number of errors and null responses (above 25%), leaving 60 participants equally distributed across the no-, low-, and high-load groups in each experiment (20 participants per group in each experiment). All were students at the University of Western Ontario, aged 17–21 years and had normal or corrected-to-normal vision. They received course credit for their participation.

Materials

The materials in Experiments 2A and 2B were identical to those in Experiments 1A and 1B, respectively.

Procedure

The procedure was the same as in Experiments 1A and 1B, with some exceptions. Rather than responding vocally, participants performed the color identification task by pressing the “J” key for red, the “K” key for blue, the “L” key for green, and the “;” key for yellow using the four fingers of their right hand. In addition, they performed the memory task by pressing the “Y” key for “same” responses and the “N” key for “different” responses with two fingers of their left hand. Similar to Schmidt et al. (2010), no timeout was used for the memory task, although participants were encouraged to respond as quickly and as accurately as they could. Finally, responses to colors and digits were followed by a feedback message following a 300-ms blank screen. The message was displayed for 500 ms in white Courier New, pt. 14, in the center of the screen, and read “Correct”, “Incorrect” or “No response” for correct, incorrect, or missed responses, respectively. The reason for these changes was to reproduce as closely as possible the conditions under which Schmidt et al. obtained their pattern (reduced contingency-learning effects with increasing WM load). For that same reason, we maintained 16 practice trials as in Experiments 1A and 1B even though, in manual responding, 16 practice trials are likely not enough for participants to effectively learn color-to-key mappings. The implication would be that at least some participants were likely still in the process of learning those mappings in the course of the experiment. However, because we failed to replicate Schmidt et al.'s pattern in Experiment 1A, we deemed it important that the procedure in Experiment 2A not deviate too much from Schmidt et al.'s procedure, one in which there were no practice trials at all.

Results

Prior to the analyses, responses faster than 300 ms on either the color identification task or the WM task and responses slower than the time limit on the color identification task (accounting for 1.1% and 1.4% of the data points in Experiments 2A and 2B, respectively) were discarded. Trials on which participants failed to respond correctly on the WM task (which accounted for 4.4% and 7.7% of the data points in the low- and high-load groups in Experiment 2A, and 4.0% and 7.1% of the data points in the low- and high-load groups in Experiment 2B) were removed as well. Latency analyses were conducted only on trials in which the response to the color identification task was also correct. Experiments 2A and 2B were analyzed in the same way as Experiments 1A and 1B, respectively. The mean RTs and error rates are presented in Tables 5 and 6 for Experiments 2A and 2B, respectively.

-Tables 5 & 6 around here-

Experiment 2A (nonconflict color identification task)

RTs. Both the main effects of Contingency (high-contingency faster than low-contingency), $F(1, 57) = 41.86$, $MSE = 758$, $p < .001$, $\eta_p^2 = .423$, and WM Load, $F(2, 57) = 5.164$, $MSE = 28876$, $p = .009$, $\eta_p^2 = .153$, were significant. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was faster than both the low-load group (p = .024) and the high-load group (p = .016), whereas the low-load and the high-load groups did not differ significantly from one another (p = .987). This time, Contingency and WM Load interacted, F(2, 57) = 6.80, *MSE* = 758, p = .002, η_p^2 = .193. Follow-up ANOVAs comparing every pair of load groups were conducted to explore this interaction. Inspection of Contingency by WM Load interactions showed that the contingency-learning effect in the no-load group (57 ms) was larger than those in the low-load (28 ms), F(1, 38) = 5.54, *MSE* = 703, p = .024, η_p^2 = .127, and high-load groups (12 ms), F(1, 38) = 15.10, *MSE* = 670, p < .001, η_p^2 = .284, whereas the low- and high-load groups did not significantly differ from each other, F(1, 38) = 1.62, *MSE* = 901, p = .211, η_p^2 = .041. Paired *t*-tests conducted for each load group separately, however, revealed significant contingency-learning effects for both the no-load group, t(19) = -8.27, p < .001, and the low-load group, t(19) = -2.99, p = .008, but not for the high-load group, t(19) = -1.27, p = .219.

Error rates. Only the main effect of Contingency (high-contingency more accurate than low-contingency) was significant, F(1, 57) = 5.66, *MSE* = .000, p = .021, η_p^2 = .090.

Experiment 2B (Stroop task)

RTs. There was a significant main effect of Congruency (congruent faster than incongruent), F(1, 57) = 136.36, *MSE* = 4471, p < .001, η_p^2 = .705. Latencies also tended to slow down as load increased, however the WM Load effect was only marginal, F(2, 57) = 2.52, *MSE* = 61061, p = .089, η_p^2 = .081. The only significant interaction was that between Congruency and Item Type, F(1, 57) = 74.61, *MSE* = 2888, p < .001, η_p^2 = .567, indicating a regular item-specific proportion-congruent effect, with larger congruency effects for mostly-congruent items (161 ms) than for mostly-incongruent items (41 ms). Importantly, there was no three-way interaction between Congruency, Item Type, and WM Load, F(2, 57) = 1.01, *MSE*

= 2888, p = .371, η_p^2 = .034, suggesting that the item-specific proportion-congruent effect was equivalent in the three load groups. The Bayes Factor, BF_{01} = 5.45 ± 2.92%, suggested that there was “moderate” evidence for the absence of the three-way interaction.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), $F(1, 57)$ = 16.93, MSE = .001, p < .001, η_p^2 = .229, Item Type (mostly incongruent more accurate than mostly congruent), $F(1, 57)$ = 6.76, MSE = .001, p = .012, η_p^2 = .106, and WM Load, $F(2, 57)$ = 7.73, MSE = .002, p = .001, η_p^2 = .213. Post hoc t-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was more accurate than both the low-load group (p = .002) and the high-load group (p = .006), whereas the low-load and high-load groups did not differ significantly from one another (p = .924). Congruency and Item Type interacted showing a regular item-specific proportion-congruent effect with a larger congruency effect for mostly-congruent items (3.3%) than for mostly-incongruent items (0.7%), $F(1, 57)$ = 9.16, MSE = .001, p = .004, η_p^2 = .138. There was also a tendency for congruency effects to be larger overall in the load groups (especially the low-load one), however, the Congruency by WM Load interaction did not reach significance, $F(2, 57)$ = 2.89, MSE = .001, p = .064, η_p^2 = .092. The three-way interaction between Congruency, Item Type and WM load was not significant, $F(2, 57)$ = .91, MSE = .001, p = .41, η_p^2 = .031, and the Bayes Factor analysis revealed that there was, indeed, “moderate” evidence for the absence of this interaction, BF_{01} = 4.21 ± 2.73%.

Discussion

Experiments 2A and 2B were an investigation of the impact of WM load on contingency-learning and item-specific proportion-congruent effects using manual responses (and feedback) in both the nonconflict color identification task and the Stroop task. With this modification, the baseline contingency-learning effect was much larger in Experiment 2A (57 ms) than it was in Experiment 1A (12 ms), replicating recent findings that feedback-assisted manual responding elicits larger contingency-learning effects than does vocal responding (Forrin & MacLeod, 2017; Spinelli et al., under review). More importantly, this modification returned a pattern of results that is consistent with Schmidt et al.'s pattern, as the 57-ms contingency learning effect in the no-load group was reduced to a nonsignificant 12-ms effect in the high-load group. Thus, it appears that the concurrent WM task not only interfered with overall performance in the color identification task, but also impaired participants' ability to learn stimulus-response associations.

Importantly, according to the contingency-learning account, the pattern found for the nonconflict color identification task in Experiment 2A should have emerged in the Stroop task in Experiment 2B. That is, there should have been a reduction in the item-specific proportion-congruent effect with increasing WM load. The reason is that according to this account, it is contingency learning that is responsible for the faster latencies that are typically observed for mostly-congruent congruent items compared to mostly-incongruent congruent items, and for mostly-incongruent incongruent items compared to mostly-congruent incongruent items. If WM load impairs contingency learning, the above differences in latencies should be attenuated, resulting in smaller item-specific proportion-congruent effects. However, no evidence in support of this prediction was found, with equivalent item-specific proportion-congruent effects in all load groups.

Note that, again, overall performance worsened with increasing WM load (although that pattern was more apparent for the error rates), confirming that the WM load manipulation was effective. However, somewhat surprising is the fact that the high-load group, the group that showed the smallest contingency-learning effect in Experiment 2A, showed the numerically largest item-specific proportion-congruent effect (a 155-ms congruency effect for mostly-congruent items and a 7-ms congruency effect for mostly-incongruent items) in Experiment 2B even though the three-way interaction was not significant. The Experiment 2B pattern is, of course, exactly the opposite of that predicted by the contingency-learning account which successfully predicted the reduced contingency-learning effect in the high-load condition in Experiment 2A. Note, however, that different participants took part in Experiments 2A and 2B. Experiments 3A and 3B were designed to allow us to re-examine this pattern in the high-load groups in a cleaner fashion by having the same participants perform both the nonconflict color identification task and the Stroop task.

Experiments 3A and 3B (manual responses)

Experiments 3A and 3B essentially replicated Experiments 2A and 2B, the only difference being that the same participants were in both experiments (i.e., task was now a within-subject manipulation although WM load was still a between-subject manipulation). The main purpose of these experiments was to seek confirmation that participants who show reduced contingency-learning effects with increasing WM load in the nonconflict color identification task also show equivalent item-specific proportion-congruent effects across all load conditions in the Stroop task. Replicating this pattern would suggest that contingency learning and item-specific proportion-congruent effects are dissociable phenomena. Specifically, it would suggest that contingency learning may not be an important component in the item-specific proportion-congruent effect, with adaptation to item-specific conflict frequency, a reactive control process, playing a crucial role instead. Indeed, the finding obtained in both Experiments 1B and 2B that WM load has no significant impact on the item-specific proportion-congruent effect in the Stroop task is easily accommodated by the DMC account (Braver, 2012; Braver et al., 2007), which proposes that reactive control, a process that generates a proportion-congruent effect, continues to be a useful option when available WM resources are decreased.

A secondary purpose of Experiments 3A and 3B was to explore the question of whether the availability of WM resources as determined by WM capacity (i.e., the amount of information an individual is able to maintain in working memory while performing a distracting task) impacts performance in a similar way as concurrent memory load does. As noted, Schmidt et al. (2010) have argued that a concurrent WM load interferes with the ability to learn word-response contingencies because that ability requires limited-capacity memory resources. Although not examined by Schmidt et al., this idea suggests that an individual-differences comparison between participants with lower and higher WM resources could be informative. Low WM-capacity individuals performing a simple color identification task (i.e., without load), similar to participants in general performing this task with a taxing concurrent task (i.e., with a high WM load), may not have enough WM resources to allocate to the process of learning word-response contingencies. As a result, contingency learning should be reduced for those individuals. The implication is that, first, the contingency-learning effect emerging in simple nonconflict color identification tasks (i.e., without load) should be smaller for individuals with lower WM capacity. Further, based on the assumption made by the contingency-learning account that the item-specific proportion-congruent effect in the Stroop task really is a contingency-learning effect in disguise, the item-specific proportion-congruent effect (i.e., without load) should also be smaller for individuals with lower WM capacity.

The results from the WM-load manipulation implemented in Experiments 2A and 2B, however, suggest that observing the complete data pattern expected from the contingency-learning account is unlikely. On the one hand, we did find, replicating Schmidt et al. (2010), that contingency-learning effects diminished with higher WM load in the nonconflict color identification task (Experiment 2A). Assuming a parallel between WM load and WM capacity (high load = low capacity), this result would certainly suggest that contingency-learning effects should also be smaller for individuals with lower WM capacity in a simple nonconflict color identification task. On the other hand, contradicting the contingency-learning account, we did not find reduced item-specific proportion-congruent effect with higher WM load in the Stroop task (Experiment 2B). Based on this empirical result, it seems unlikely that item-specific proportion-congruent effects would be smaller for individuals with lower WM capacity in a simple Stroop task.

This expectations derived from the control account concerning the item-specific proportion-congruent effect are somewhat more complicated. The DMC account (Braver, 2012; Braver et al., 2007), in particular, attributes an important role to WM capacity in determining the modes of control (proactive vs. reactive) that individuals can and do use. Low WM-capacity individuals would be relatively unable to implement proactive control and would, therefore, mostly rely on reactive control. In contrast, high WM-capacity individuals would typically engage in proactive control but would have access to reactive control as well. The implications for the impact of WM capacity on the item-specific proportion-congruent effect in the Stroop task would be as follows. First, to the extent that the item-specific proportion-congruent effect reflects reactive control, a form of control both low and high WM-capacity individuals have access to, this effect should emerge in all individuals. Second, in the Stroop task, the preference for high WM-capacity individuals for proactive control would induce them to engage in a process of constant goal maintenance. As noted (see footnote 1), this process could not cause an item-specific proportion-congruent effect, however, it could attenuate this effect. The reason is that focusing attention on task-relevant information would reduce the impact of conflict, and therefore the impact of the frequency with which that conflict arises.

In line with this idea, Hutchison, Bugg, Lim, and Olsen (2016) found that using an informative cue before a Stroop trial to prompt proactive control reduced or eliminated the item-specific proportion-congruent effect (i.e., there was little difference between mostly-congruent and mostly-incongruent items when participants were well prepared to deal with conflict on the upcoming trial). What is possible, therefore, is that high WM-capacity individuals would engage a high level of proactive control even in a regular

Stroop task (i.e., without cues) and that this process would reduce not only the overall congruency effects but also the item-specific proportion-congruent effect, particularly in the error data.

The reason for error data being more likely to show this reduction is two-fold. First, application of proactive control would likely reduce error rates to the floor, making it hard to detect effects in those data. Second, errors are considered a more sensitive index of goal maintenance than latencies because committing a word-reading error on an incongruent trial means that the task goal was neglected (whereas an increased latency does not necessarily indicate goal neglect; Kane & Engle, 2003). If, in a Stroop task, high WM-capacity individuals strive to constantly maintain the color-naming goal by applying proactive control, that means that they would be doing so even when dealing with situations for which concurrent application of reactive control suggests that relaxing attention would be appropriate, e.g., with mostly-congruent items. Thus, in high WM-capacity individuals, the relaxation of attention for mostly-congruent items promoted by reactive control might cause an increased latency when those items are incongruent, but it should not result in an error. The same would not be true for low WM-capacity individuals, however, because they would not have the WM resources necessary to support continuous engagement of proactive control. Therefore, in those individuals, the relaxation of attention for mostly-congruent items induced by reactive control may very well cause a word-reading error in addition to an increased latency when those items are incongruent. As a result, low WM-capacity individuals would tend to show a larger item-specific proportion-congruent effect in the errors than would high WM-capacity individuals.

In sum, while the contingency-learning account would seem to predict that the item-specific proportion-congruent effect, as an instance of contingency learning, would be reduced for low WM-capacity individuals, the prediction made by the DMC account would be, if anything,, a smaller item-specific proportion-congruent effect for *high* WM-capacity individuals with this pattern being more likely to emerge in the errors.

Interestingly, an experiment conducted by Hutchison (2011) does allow at least an initial evaluation of these ideas. Hutchison collected WM-capacity scores for participants performing a Stroop task in which list-wide proportion-congruent effects, item-specific proportion-congruent effects, and contingency-learning effects were examined independently from one another. Of relevance for the present research, he found that although both low and high WM-capacity participants did show a significant item-specific proportion-congruent effect in latencies, only low WM-capacity participants showed this effect in error rates. This result is quite consistent with the DMC view that, while all individuals have access to reactive control (as demonstrated by the emergence of the item-specific proportion-congruent effect in the latencies), high WM-capacity individuals would concurrently engage in proactive control, thus preventing word-reading errors and reducing the item-specific proportion-congruent effect in the error data. Also of note is the fact that Hutchison's data appear difficult to reconcile with the contingency-learning account. In addition to the fact that this account appears unable to explain the pattern of results relative to the item-specific proportion-congruent manipulation, there was also no evidence in his contingency-learning manipulation, which was dissociable from the item-specific proportion-congruency manipulation in his experiment, that contingency-learning effects were larger for high than low WM-capacity individuals, the pattern that the contingency-learning account appears to predict.[6]

What must be noted, however, is that Hutchison's experiment was peculiar in that list-wide proportion congruency, item-specific proportion congruency, and contingency learning were all manipulated in the context of that same experiment. Furthermore, his use of verbal responses to colors might have weakened the contingency-learning effect (which was indeed smaller than generally reported; Forrin &

MacLeod, 2017; Spinelli et al., under review), making individual differences related to this effect harder to observe (as appears to have occurred in the present Experiment 1A).

The present Experiments 3A and 3B allowed a re-examination of these issues in the context of more typical versions of the item-specific proportion-congruent and contingency-learning manipulations, i.e., the two-item set design used thus far. WM capacity was assessed for participants in the no-load group with a battery of WM tests administered after Experiments 3A and 3B were completed. A data pattern consistent with the contingency-learning account would be for there being smaller contingency-learning and item-specific proportion-congruency effects for high WM capacity individuals in the no-load condition. A pattern more consistent with a control account would be that in the no-load condition in the Stroop task individuals with lower WM capacity would show a more pronounced pattern of item-specific proportion-congruent effects than individuals with higher WM capacity, with that difference expected to be more prominent in the error rates than in the latencies because errors appear to index cognitive processes (i.e., goal neglect) which better differentiate low and high WM-capacity individuals performing the Stroop task (Kane & Engle, 2003). Note also that manual responses (and feedback) were used in Experiment 3A (and 3B) in an effort to produce larger contingency-learning effects than those observed by Hutchison (2011), thus providing a better way of determining whether and how WM capacity influences the process of learning word-response contingencies.

Method

Participants

Two hundred and thirty-five participants took part in both Experiment 3A (nonconflict color identification task) and Experiment 3B (Stroop task). Of these, 127 were assigned to the no-load group, 51 were assigned to the low-load group, and 57 were assigned to the high-load group. Twenty-seven participants were removed because of an excessive number of errors and null responses (above 25%) in either Experiment 3A or Experiment 3B, leaving 208 participants, of which 126 were in the no-load group, 43 were in the low-load group, and 39 were in the high-load group. Many more participants were tested in the no-load group than in the other groups because WM-capacity scores were recorded for those participants, and individual-differences research requires large sample sizes. Compared to Experiments 1A/1B and 2A/2B, the sample sizes of the low-load and the high-load groups were approximately doubled because there were half of the number of items per cell in Experiments 3A and 3B (see Materials) and, hence, more potential for noise to affect the results. All participants were students at the University of Western Ontario, aged 17–31 years and had normal or corrected-to-normal vision. They received course credit for their participation.

Materials

The materials were identical to those of Experiments 1A and 1B, respectively, except that each experiment only included 96 trials (rather than 192) because participants did only one block for each experiment instead of two.

Procedure

Participants completed Experiment 3A and 3B in a single session. Each experiment included a single block of 96 trials preceded by 8 practice trials. Half of the participants performed Experiment 3A (nonconflict color identification task) first and Experiment 3B (Stroop task) second and the other half performed Experiment 3B first and Experiment 3A second. Other than this difference, the procedure was the same as in Experiments 2A and 2B. Following these experiments, participants in the no-load group completed a battery of complex span tests including one block of the Operation Span task, followed by one block of the Symmetry Span task, followed by one block of the Rotation Span task

(Conway et al., 2005; Kane et al., 2004; Redick et al., 2012; Unsworth, Heitz, Schrock, & Engle, 2005). These tests were shortened versions of complex span tasks aimed to test different constructs in working memory, so as to obtain reliable measures of WM capacity as a whole while minimizing testing duration (Foster et al., 2015). In these complex span tasks, participants were given a sequence of to-be-remembered items (e.g., a sequence of letters) and had to complete a distractor task (e.g., solving a math problem) between the presentations of each of the to-be-remembered items in the sequence. The sequence of to-be-remembered items varied from two to five items (Symmetry Span and Rotation Span tasks) or from three to seven items (Operation Span task). Scores are calculated by summing the number of items correctly recalled in the correct order, a measure known as the partial score (Turner & Engle, 1989). Participants who completed the complex span tasks also completed a questionnaire collecting measures of their monolingual/bilingual status and other variables known to influence executive functioning. The questionnaire data were irrelevant for the present purposes and were not analyzed.

Results

Prior to the analyses, responses faster than 300 ms on either the color identification task or the WM task and responses slower than the time limit on the color identification task (accounting for 0.6% and 1.2% of the data in Experiments 3A and 3B, respectively) were discarded. Trials on which participants failed to respond correctly on the WM task (which accounted for 4.0% and 6.3% of the data in the low- and high-load groups in Experiment 3A, and 4.3% and 6.8% of the data in the low- and high-load groups in Experiment 3B) were removed as well. Latency analyses were conducted only on trials in which the response to the color identification task was also correct. Experiments 3A and 3B were analyzed in the same way as Experiments 1A and 1B, respectively, with the addition of Order (Experiment 3A first vs. Experiment 3B first) as a factor. To preview the results, Order did reveal some effects of practice (e.g., reduced latencies and error rates if the experiment in question was performed second) but did not modify the theoretically important interactions in the WM-load analysis in either experiment (i.e., the Contingency by WM Load interaction in Experiment 3A and the Congruency by Item Type by WM Load interaction in Experiment 3B). Thus, for simplicity, we present the mean RTs and error rates in Tables 7 and 8 for Experiments 3A and 3B, respectively, without splitting the data by Order.

-Tables 7 and 8 around here-

We also explored the relation between WM capacity and performance in the two experiments for participants in the no-load group (the group for which WM capacity was measured). For this analysis, 28 participants were removed because their accuracy on the distractor component of one or more of the complex span tasks was below 75%.[7] For each of the remaining 98 participants, the partial scores obtained in each of the three complex span tasks were standardized (as the Operation Span task returns scores on a different scale than the other two tasks) and then averaged to obtain a single composite score.

The analysis was conducted using mixed-effects modelling, a type of analysis which permits use of both continuous and categorical variables (the fixed effects), while controlling for variance among the participants and the items being used (the random effects; Baayen, 2008; Baayen, Davidson, & Bates, 2008; for similar analyses in the context of the Stroop task, see Meier & Kane, 2013, 2015). Latencies and errors were analyzed using generalized linear mixed-effects models (GLMMs) in R version 3.5.1 (R Core Team, 2018), treating subjects, colors, and words as random effects. For Experiment 3A, the fixed effects were Contingency (high vs. low), Order (Experiment 3A first vs. Experiment 3B first) and Span Score (the composite score from the complex span tasks, a continuous variable). For Experiment 3B, the fixed effects were Congruency (congruent vs. incongruent), Item Type (mostly congruency vs. mostly incongruent), Order (Experiment 3A first vs. Experiment 3B first) and Span Score.

For both experiments, the Span Score was standardized (i.e., centered and scaled) to help model estimation (Bolker, 2019). Prior to running the model, R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., contr.sum) to help interpret lower-order effects in the presence of higher-order interactions (Levy, 2014; Singmann & Kellen, 2018). The lme4 package, version 1.1-18-1 (Bates, Mächler, Bolker, & Walker, 2015) was used to run the GLMMs. The models were fit by maximum likelihood with the Laplace approximation technique. Model estimation was conducted using the BOBYQA optimizer, an optimizer known to generate fewer false-positive convergence failures than other optimizers in the current version of lme4, with a maximum number of 1,000,000 iterations (Bolker, 2019). The emmeans package, version 1.3.1 (Lenth, 2018), was used to conduct follow-up tests. The ggplot2 package, version 3.1.0 (Wickham, 2016), was used to generate graphs. A Gamma distribution was used to fit the raw RTs, with an identity link between the fixed effects and the dependent variable (Lo & Andrews, 2015), whereas a binomial distribution with a logit link between the fixed effects and the dependent variable was used to fit the error data.

In addition to these mixed-effects analyses, traditional ANOVAs were conducted contrasting participants in the top quartile for span scores (the high WM-capacity individuals) with those in the bottom quartile for span scores (the low-WM capacity individuals). Although this extreme-group analysis does not reflect the current tendency in individual-differences research (e.g., Meier & Kane, 2013, 2015), it is reported in the Appendix for the sake of consistency with previous reports in the relevant literature, particularly in Hutchison (2011) who did use this analysis.

Experiment 3A (nonconflict color identification task)

WM load analysis

RTs. There were main effects of Contingency (high-contingency faster than low-contingency), $F(1, 202) = 49.20$, $MSE = 2225$, $p < .001$, $\eta_p^2 = .196$, Order (overall faster latencies for participants who performed Experiment 3A following Experiment 3B than for participants who performed Experiment 3A first), $F(1, 202) = 9.42$, $MSE = 21867$, $p = .002$, $\eta_p^2 = .045$, and WM Load, $F(2, 202) = 40.39$, $MSE = 21867$, $p < .001$, $\eta_p^2 = .286$. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was faster than both the low-load group ($p < .001$) and the high-load group ($p < .001$), and that the low-load group was faster than the high-load group ($p = .046$). Importantly, Contingency and WM Load interacted, $F(2, 202) = 9.10$, $MSE = 2225$, $p < .001$, $\eta_p^2 = .083$.

Follow-up ANOVAs comparing every pair of load groups were conducted to explore this interaction. Inspection of the Contingency by WM Load interactions showed that the contingency-learning effect for the no-load group (63 ms) was larger than those for the low-load (31 ms), $F(1, 165) = 7.97$, $MSE = 2197$, $p = .005$, $\eta_p^2 = .046$, and the high-load groups (17 ms), $F(1, 161) = 14.89$, $MSE = 2161$, $p < .001$, $\eta_p^2 = .085$, whereas low- and high-load groups did not significantly differ, $F(1, 78) = .76$, $MSE = 2416$, $p = .387$, $\eta_p^2 = .010$. Paired *t*-tests conducted for each load group separately, however, revealed significant contingency-learning effects for both the no-load group, $t(125) = -11.07$, $p < .001$, and the low-load group, $t(42) = -2.93$, $p = .006$, but not for the high-load group, $t(38) = -1.53$, $p = .135$. There was also a marginal interaction between Contingency and Order, $F(1, 202) = 2.98$, $MSE = 2225$, $p = .086$, $\eta_p^2 = .015$, indicating a tendency for overall larger contingency-learning effects for participants who did Experiment 3A following Experiment 3B (54 ms) than for participants who did Experiment 3A first (42 ms). Likely, this marginal interaction reflects a practice effect whereby contingencies are more easily learned when progressing in the experiment (e.g., Schmidt & De Houwer, 2016).

Error rates. The only significant effect was that of Contingency (high-contingency more accurate than low-contingency), *F*(1, 202) = 6.08, *MSE* = .001, *p* = .014, η_p^2 = .029. There was also a tendency for contingency-learning effects to decrease with increasing WM load, although the Contingency by WM Load interaction did not reach significance, *F*(2, 202) = 2.77, *MSE* = .001, *p* = .065, η_p^2 = .027, and the Bayes Factor indicated no real preference for either the model with the interaction or the model without it, BF_{01} = 1.35 ± 5.9%.

WM capacity analysis

RTs. There were main effects of Contingency (high-contingency faster than low-contingency), *ß* = -32.42, *SE* = 2.19, *z* = -14.81, *p* < .001 and Span Score (latencies decreased with higher scores), *ß* = -16.82, *SE* = 3.84, *z* = -4.38, *p* < .001. There were also a three-way between Contingency, Order, and Span Score, *ß* = 5.30, *SE* = 1.97, *z* = 2.69, *p* = .007.

Follow-up tests revealed that the source of this interaction was that Contingency and Span Score had a different relationship for participants performing Experiments 3A first compared to participants performing Experiment 3A following Experiment 3B. The former group of participants showed a marginal tendency for diminishing contingency-learning effects with higher Span Score, *ß* = 11.01, *SE* = 5.81, *z* = 1.90, *p* = .058. This pattern is represented in Figure 1 as a scatterplot of participants' mean latencies to low- and high-contingency items as a function of their Span Score. Here, when moving from the left side of the graph (lower span scores) to the right side of the graph (higher span scores), latencies diminish overall, and so does the distance between the solid line (high-contingency items) and the dashed line (low-contingency items), i.e., the contingency-learning effect.

Participants performing Experiment 3A following Experiment 3B showed a marginal tendency in the opposite direction, *ß* = -10.18, *SE* = 6.13, *z* = -1.66, *p* = .097, with *increasing* contingency-learning effects with higher Span Score. This pattern is represented in Figure 2 where the distance between the solid line (high-contingency items) and the dashed line (low-contingency items) increases, rather than decreases, with higher Span Score.

-Figures 1 and 2 around here-

Errors. There were significant main effects of Contingency (high-contingency more accurate than low-contingency), *ß* = .34, *SE* = .07, *z* = 4.62, *p* < .001, and Span Score (accuracy increased with higher scores), *ß* = .32, *SE* = .10, *z* = 3.34, *p* < .001.There was also a three-way interaction between Contingency, Order, and Span Score, *ß* = -.14, *SE* = .07, *z* = -2.03, *p* = .042. Follow-up tests revealed that, similar to what was found in the latencies, this interaction indicated that while contingency-learning effects tended, if anything, to decrease (but not significantly so) with higher scores for participants who did Experiment 3A first, *ß* = -.21, *SE* = .20, *z* = -1.06, *p* = .292, those effects showed a marginal tendency to increase with higher scores for participants who did Experiment 3A following Experiment 3B, *ß* = .33, *SE* = .17, *z* = 1.89, *p* = .059.

Experiment 3B (Stroop task)

WM load analysis

RTs. There were main effects of Congruency (congruent faster than incongruent), *F*(1, 202) = 301.62, *MSE* = 7367, *p* < .001, η_p^2 = .599, Order (overall faster latencies for participants who performed Experiment 3B following Experiment 3A than for participants who performed Experiment 3B first), *F*(1, 202) = 7.08, *MSE* = 62962, *p* = .008, η_p^2 = .034, and WM Load, *F*(2, 202) = 28.63, *MSE* = 62962, *p* < .001, η_p^2 = .221. Post hoc *t*-tests using the Tukey HSD adjustment for multiple comparisons revealed that the no-load group was faster than both the low-load group (*p* < .001) and the high-load group (*p* < .001), while low-load and high-load groups did not differ significantly from one another (p = .513). There was an interaction between Congruency and Order, F(1, 202) = 7.15, MSE = 7367, p < .001, η_p^2 = .034, indicating that, overall, congruency effects were larger for participants who did Experiment 3B following Experiment 3A (134 ms) than for participants who did Experiment 3B first (111 ms). This result seems to indicate that participants were overall less prepared to deal with conflict in the Stroop task after having performed a version of the color identification task in which there was no conflict to deal with. More importantly, there was also an interaction between Congruency and Item Type, F(1, 202) = 98.41, MSE = 5296, p < .001, η_p^2 = .328, indicating a regular item-specific proportion-congruent effect, with larger congruency effects for mostly-congruent items (180 ms) than for mostly-incongruent items (60 ms). Finally, there was again no three-way interaction between Congruency, Item Type, and WM Load, F(2, 202) = .63, MSE = 5296, p = .54, η_p^2 = .006, indicating that the item-specific proportion-congruent effect was equivalent in all load groups. Indeed, Bayesian analyses revealed that there was “strong” evidence in favor of the absence of this three-way interaction, BF_{01} = 10.76 ± 13.63%.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), F(1, 202) = 39.76, MSE = .003, p < .001, η_p^2 = .164, and Order (participants who did Experiment 3B following Experiment 3A were overall more accurate than those who did Experiment 3B first), F(1, 202) = 4.41, MSE = .006, p = .037, η_p^2 = .021. Congruency also interacted with Item Type, F(1, 202) = 8.63, MSE = .003, p = .004, η_p^2 = .041, indicating that congruency effects were larger for mostly-congruent items (4.1%) than mostly-incongruent items (1.3%). This item-specific proportion-congruent effect was not modulated by WM Load, as no three-way interaction between Congruency, Item Type, and WM Load, F(2, 202) = .923, MSE = .003, p = .40, η_p^2 = .009, was found. Once again, in the Bayesian analyses, there was “moderate” evidence in support of the model without the interaction, BF_{01} = 7.04 ± 6.48%.

The item-specific proportion-congruent effect was, however, modulated by Order, i.e., there was a three-way interaction between Congruency, Item Type, and Order, *F*(2, 202) = 5.28, *MSE* = .003, *p* = .023, η_p^2 = .025. To explore this three-way interaction, two separate ANOVAs were conducted for each order. Inspection of the Congruency by Item Type interaction in these ANOVAs revealed a regular item-specific proportion-congruent effect for participants who did Experiment 3B first (congruency effect for mostly-congruent items: 5.1%; congruency effect for mostly-incongruent items: .7%), *F*(1, 102) = 10.64, *MSE* = .004, *p* = .002, η_p^2 = .094. In contrast, the item-specific proportion-congruent effect was not significant for participants who did Experiment 3B after Experiment 3A (congruency effect for mostly-congruent items: 3%; congruency effect for mostly-incongruent items: 1.8%), *F*(1, 100) = .30, *MSE* = .002, *p* = .59, η_p^2 = .003, presumably because errors were reduced in this situation (as in Experiment 1B).

WM capacity analysis

RTs. There were main effects of Congruency (congruent faster than incongruent), *ß* = -61.03, *SE* = 2.25, *z* = -27.14, *p* < .001, Order (overall faster latencies for participants who did Experiment 3B following Experiment 3A than for participants who did Experiment 3B first), *ß* = -32.77, *SE* = 4.20, *z* = -7.80, *p* < .001, and Span Score (latencies decreased with higher scores), *ß* = -20.34, *SE* = 3.94, *z* = -5.16, *p* < .001. Congruency interacted with Item Type, *ß* = -30.59, *SE* = 2.16, *z* = -14.18, *p* < .001, indicating a regular item-specific proportion-congruent effect. There was also a marginal interaction between Item Type and Order, *ß* = -4.55, *SE* = 2.40, *z* = -1.90, *p* = .058, indicating that the advantage for participants who did Experiment 3B following Experiment 3A compared to participants who did Experiment 3B first was overall more pronounced in the mostly-congruent condition than in the mostly-incongruent condition. Order also interacted with Span Score, *ß* = -12.19, *SE* = 3.42, *z* = -3.57, *p* < .001, indicating a stronger reduction in latencies associated with higher Span Score for participants who did Experiment 3B following Experiment 3A than for participants who did Experiment 3B first. Finally, there was a three-way interaction between Congruency, Order, and Span Score, *ß* = 6.02, *SE* = 2.12, *z* = 2.84, *p* = .005. Follow-up tests revealed that the source of this interaction was that, while congruency effects diminished with higher Span Score for participants who did Experiment 3B following Experiment 3A, *ß* = 17.88, *SE* = 5.97, *z* = 2.99, *p* = .003, congruency effects tended, if anything, to increase (but not significantly so) for participants who did Experiment 3B first, *ß* = -6.20, *SE* = 6.02, *z* = -1.03, *p* = .303.

More importantly, there was no three-way interaction between Congruency, Item Type, and Span Score, *ß* = -1.35, *SE* = 2.38, *z* = -.57, *p* = .57, suggesting that the item-specific proportion-congruent effect, overall, did not change across the range of scores, a pattern represented in Figure 3. In this scatterplot, the distance between the solid line (congruent items in the mostly-congruent condition) and the dotted line (incongruent items in the mostly-congruent condition) is larger than the distance between the long-dashed line (congruent items in the mostly-incongruent condition) and the dot-dash patterned line (incongruent items in the mostly-incongruent condition), indicating an item-specific proportion-congruent effect. This pattern remains similar when moving from the left side of the graph (lower span scores) to the right side of the graph (higher span scores) even as latencies diminish for individuals with higher scores. Note also that Order did not modulate this pattern, i.e., there was no four-way interaction between Congruency, Item Type, Span Score, and Order, *ß* = 3.45, *SE* = 2.51, *z* = 1.38, *p* = .169.

Errors. There were main effects of Congruency (congruent more accurate than incongruent), *ß* = .48, *SE* = .08, *z* = 6.34, *p* < .001, and Span Score (fewer errors with higher scores), *ß* = .37, *SE* = .11, *z* = 3.23, *p* = .001. There was also an interaction between Congruency and Item Type, *ß* = .24, *SE* = .07, *z* = 3.35, *p* < .001, reflecting a regular item-specific proportion-congruent effect. There was also a marginal interaction between Item Type and Order, *ß* = .14, *SE* = .07, *z* = 1.85, *p* = .065, indicating a tendency for participants who did Experiment 3B first to be overall more accurate than participants who did Experiment 3B following Experiment 3A in the mostly-incongruent condition, but not in the mostly-congruent condition. Order also interacted with Span Score, *ß* = -.23, *SE* = .11, *z* = -2.01, *p* = .045, indicating a stronger reduction in error rates associated with higher Span Score for participants who did Experiment 3B first than for participants who performed Experiment 3B following Experiment 3A.

More importantly, there was no evidence that the item-specific proportion-congruent effect was modulated by Span Score, i.e., there was no three-way interaction between Congruency, Item Type, and Span Score, *ß* = -.01, *SE* = .06, *z* = -.16, *p* = .88. The relation between Congruency, Item Type, and Span Score is represented in the scatterplot in Figure 4. Even though the item-specific proportion-congruent effect is more noticeable in the left side of the graph, statistically, there was no evidence that this effect was larger for participants scoring lower on the complex span tasks than for those scoring higher.

-Figures 3 and 4 around here-

Discussion

Using a within-subject design, Experiments 3A and 3B replicated the basic data patterns found in Experiments 2A and 2B: Increasing WM load impairs participants' ability to learn word-response associations in the nonconflict color identification task, but it does not affect the ability of the same participants to produce item-specific proportion-congruent effects in the Stroop task. The robustness of the pattern found for Experiments 2A and 2B is thus confirmed.

The exploratory WM-capacity analysis conducted for participants in the no-load condition produced a pattern which did not parallel that of the load manipulation in the nonconflict color identification task. In this task, replicating Hutchison's (2011) contingency-learning manipulation, there was no overall relation between WM capacity and contingency-learning effects. However, the order in which the two tasks (the nonconflict color identification task and the Stroop task) were completed appeared to have a role in modulating that relation. For participants who completed the nonconflict color identification task first, higher WM capacity was associated with smaller contingency-learning effects. For participants who completed the nonconflict color identification task following the Stroop task, the opposite pattern was found, with higher WM capacity leading to larger contingency-learning effects. Thus although there was some evidence in this analysis for a relation between WM capacity and contingency-learning effects, this relation does not appear to be as straightforward as the relation between WM load and contingency learning appears to be.

Additionally, there was no evidence in the Stroop task that WM capacity modulated item-specific proportion-congruent effects in either the latencies or, more centrally, the error rates. This result does parallel that of the load manipulation but it represents a failure to replicate Hutchison's (2011) data pattern, although this failure may depend on the different analysis we used (we did, in fact, replicate Hutchison when using his extreme-groups approach, see Appendix). Overall, inter-individual variability in WM resources does not appear to have a clear impact on either contingency-learning or item-specific proportion-congruent effects.

General Discussion

The item-specific proportion-congruent effect does not have "everything to do with contingency"

The contingency-learning account of proportion-congruent effects has led to the reconsideration of a vast amount of evidence once thought to lend support to the existence of a mechanism of adaptation to conflict frequency (Schmidt, 2013b). That account has been especially compelling in the case of the item-specific proportion-congruent effect (Jacoby et al., 2003), as most researchers have now concluded that learning of word-response contingencies, rather than adaptation to item-specific conflict frequency, is the default process governing performance in item-specific proportion-congruent manipulations using the two-item set design, i.e., a type of design that allows learning of contingencies for all stimuli (Bugg &

Hutchison, 2013; Schmidt, 2013a, 2013b; Schmidt & Besner, 2008). The present research, however, casts doubt on this conclusion.

In this research, nonconflict and Stroop versions of a color identification task were combined with a concurrent WM-load task in order to examine whether increasing WM load affects the contingency-learning effect and the item-specific proportion-congruent effect in the same way. According to Schmidt et al. (2010), contingency learning is a resource-dependent process, as demonstrated by the fact that a high WM load reduces contingency-learning effects in a nonconflict color identification task. However, if the process that produces contingency-learning effects is the same as the process that produces item-specific proportion-congruent effects (Schmidt & Besner, 2008), a similar pattern should emerge for item-specific proportion-congruent effects under load: Increasing demands on WM should reduce the contingency-learning effects that are assumed to cause the characteristic pattern of the item-specific proportion-congruent effect. As a result, item-specific proportion-congruent effects, similar to contingency-learning effects, should be reduced by increasing WM load.

The results from our experiments are not consistent with this prediction, however. Using vocal responding to colors, Experiments 1A and 1B did yield evidence that contingency-learning and item-specific proportion-congruent effects are alike in that they are both unaffected by WM load. The more central message from these results, however, is merely that the vocal responding procedure fails to replicate Schmidt et al.'s (2010) original finding in the nonconflict color identification task, possibly because vocal responding elicits such small baseline contingency learning effects (Forrin & MacLeod, 2017; Spinelli et al., under review) that an observable further reduction is virtually impossible to achieve.[8]

Manual responding to colors (plus the addition of feedback on each trial), however, not only increased baseline contingency-learning effects in the nonconflict color identification task but also successfully replicated the finding that increasing WM load reduces the magnitude of such effects (Experiments 2A). In contrast, no parallel reduction of item-specific proportion-congruent effects in the Stroop task was observed (Experiments 2B). This pattern was obtained even when the same participants were tested in both the nonconflict task and the Stroop task (Experiments 3A and 3B).[9]

One aspect of our WM load manipulation that should be noted is that in no case did WM load have a strong impact on the basic Stroop congruency effect. Stroop effects have been reported to increase when a WM load is concurrently maintained (e.g., Lavie, 2005), potentially because maintaining that load impairs individuals' ability to proactively maintain the task goal (Kalanthroff, Avnit, Henik, Davelaar, & Usher, 2015). In the present experiments, however, the basic congruency effect, if anything, tended to decrease under higher load. An anonymous reviewer on a previous version of this manuscript pointed out that the failure to observe larger congruency effects with a concurrent WM load might indicate that our load manipulation was ineffective. Although other load manipulations might have been possible (see footnote 5), the goal that our manipulation was required to achieve was to impair contingency learning, i.e., the critical process that, according to the contingency-learning account, underlies the item-specific proportion-congruent effect in the Stroop task. As this goal was achieved (as demonstrated by reduced contingency-learning effects under load in the nonconflict color identification task in Experiments 2A and 3A), the fact that our load manipulation spared the basic congruency effect in the Stroop task does not appear to be at all problematic.

In sum, our overall pattern of results poses a challenge to the view that congruency effects in item-specific proportion-congruent paradigms are the result of a contingency-learning process. This view would predict that increasing demands on WM should impair contingency-learning and item-specific proportion-congruent effects in a similar way, a pattern the present experiments failed to obtain.

An explanation that better accommodates the present results is one that assumes a process other than contingency learning drives the item-specific proportion-congruent effect in the Stroop task. Adaptation to item-specific conflict frequency would be such a process. According to this explanation (Blais et al., 2007; Jacoby et al., 2003; Shedden et al., 2013), participants would learn to associate specific words with a specific control process: Mostly-congruent words would lead to relaxed attention (as the irrelevant dimension is typically not conflicting) whereas mostly-incongruent words would lead to focused attention to the relevant dimension (as the irrelevant dimension is typically conflicting). Importantly, what the present results suggest is that WM load has virtually no impact on participants' ability to implement this type of control processes. At first blush, a claim of this sort may appear surprising, as one would expect that a concurrent WM task diverting attentional resources away from the Stroop task should interfere with a process that is itself attentional. However, research within the DMC framework (Braver, 2012; Braver et al., 2007) suggests that increasing demands on WM may only have that sort of effect on proactive processes, that is, effortful processes that involve sustained maintenance of task goals. In other situations, increasing WM load may, instead, bias individuals to use reactive processes, that is, processes that rely on the environment to re-activate task goals (Burgess & Braver, 2010; Speer et al., 2003). As adaptation to item-specific conflict frequency would be one example of that type of process (Gonthier et al., 2016), the claim that WM load does not interfere with its implementation would follow. Indeed, from this point of view, diminished WM resources should make item-specific conflict adaptation an even more convenient option than it is when those resources are intact.

Item-specific conflict frequency, contingency learning, and WM capacity: A not-so-simple story

The WM-capacity analyses of Experiments 3A and 3B are potentially relevant in evaluating the conclusions suggested by our WM-load manipulations. The DMC account, the account our data appear to favor, concerns itself not only with differences in WM resources that are induced experimentally (e.g., by use of a concurrent WM load) but also with differences in WM resources that occur naturally across individuals (i.e., their WM capacity; Braver, 2012; see also Kane & Engle, 2003). Specifically, it suggests that high WM-capacity individuals might be more prone to engage in proactive control than low WM-capacity individuals. If so, in the Stroop task, this tendency could result in less pronounced item-specific proportion-congruent effects in high than low WM-capacity individuals, a pattern that would most clearly emerge in error rates rather than in latencies because an error, but not necessarily an increased latency, would index participants' inability to successfully maintain the task goal (Kane & Engle, 2003; MacLeod, 1991).

A different set of predictions could be derived from Schmidt et al.'s (2010) idea that contingency learning depends on limited-capacity resources. This idea suggests that contingency learning should be relatively impaired in individuals who possess fewer of those resources, i.e., low WM-capacity individuals. As a result, those individuals, compared to individuals with a higher WM capacity, should show smaller contingency-learning effects in a nonconflict color identification task. Additionally, based on the contingency-learning account's idea that item-specific proportion-congruent effects in the Stroop task really are contingency-learning effects in disguise, item-specific proportion-congruent effects should also be smaller in low than high WM-capacity individuals. Notably, this prediction concerning the item-specific proportion-congruent effect is quite the opposite of that derivable from the DMC account.

Although, as noted, a previous study exists (Hutchison, 2011) which allows an examination of these contrasting predictions, Experiments 3A and 3B allowed for a clearer examination within the original contingency-learning and item-specific proportion-congruent paradigms, respectively. Partial support for the DMC account was obtained in the extreme-groups ANOVAs reported in the Appendix, in the form of larger items-specific proportion-congruent effects for low than high WM-capacity individuals in the error rates in the Stroop task and no relation between contingency learning and WM capacity in either latencies or error rates in the nonconflict color identification task – a replication of the relevant results reported by Hutchison, who used the same type of analysis.

However, the results of our main statistical analysis technique, one where the full range of WM capacity sampled is analyzed within a mixed-effects model (Meier & Kane, 2013, 2015), suggested a more complex story. Although this full-sample analysis, like the extreme-groups analysis, did reveal that WM capacity had an impact on performance overall, with faster and more accurate responding (as well as reduced congruency effects) associated with higher WM capacity, there was no evidence that, in the Stroop task, WM capacity had an impact on the item-specific proportion-congruent effect in the latencies or, most importantly, in the error rates. That is, in this full-sample analysis, in contrast to our extreme-groups analysis and Hutchison's (2011) results, increasing WM capacity did reduce errors overall but did not reduce the item-specific proportion-congruent effect. There was also little evidence in either analysis for an overall reduction in congruency effects with higher WM capacity, a result that would have been expected based on the idea that proactive control in high WM-capacity individuals would help them deal with conflict more efficiently. Essentially, although our WM-capacity results are clearly incompatible with the contingency-learning account, they offer no strong support for the DMC account either.

The contingency-learning account also gained little support from the results of the nonconflict color identification task. As noted, Schmidt et al.'s (2010) idea that the amount of limited-capacity resources available determines the magnitude of contingency-learning effects leads to the expectation that, in a nonconflict color identification task, contingency-learning effects should be smaller in low WM-capacity individuals (i.e., individuals with fewer WM resources) than high WM-capacity individuals (i.e., individuals with more WM resources). As revealed by the full-sample analysis (but not the extreme-groups analysis), this pattern (larger contingency-learning effects with higher WM capacity) did occur in a portion of the data, for participants who performed the nonconcolor identification task following the Stroop task. However, the opposite pattern (*smaller* contingency-learning effects for higher WM-capacity individuals) was found in the group of participants who completed the noncolor identification task first.

These results suggest that a complete explanation of the relation between contingency learning and WM capacity is unlikely to be found in the original contingency-learning account (e.g., Schmidt et al., 2010), which would seem to require additional notions in order to explain the pattern of results that emerged in our full-sample analysis. These notions could include, for example, the idea that contingency learning might be modulated by the amount of attention individuals allocate to the process of making sure that stimulus-response mappings are being correctly implemented. That process, a monitoring process, could certainly divert attentional resources away from the process of learning color-word contingencies. Therefore, to the extent that participants must engage in such a process, doing so would reduce their opportunity to learn the relevant contingencies for the stimuli being used in the experiment in a similar way that maintaining a WM load would (Spinelli et al., under review). In an attempt to explain the contrasting patterns obtained for participants who completed the nonconflict color identification task as the first task vs. the second task (within the contingency-learning framework), it could be assumed that the conditions under which individuals may feel a weaker vs. stronger need to engage in this monitoring process could vary depending on the WM capacity of the individual and/or the amount of practice received in the task. For example, high WM-capacity individuals may feel a strong need to engage in the monitoring process initially, leaving little opportunity to learn the contingencies in the task, but not after an entire block of practice, a situation in which they could relax the monitoring process and be better able to pick up on those contingencies.

These hypotheses are, of course, purely speculative at this point. In general, from the present dataset, it would appear incautious to draw strong conclusions about the nature of the relation between WM-capacity and either item-specific proportion-congruent effects or contingency-learning effects. While, overall, the WM-capacity analyses we conducted do show some consistency with the previous studies (Hutchison, 2011; see also Kane & Engle, 2003) when using the same analysis (i.e., an extreme-groups analysis) used in those studies, they certainly depict a less clear situation than the WM-load analyses do. Part of the reason for this lack of clarity might be that our experiments were relatively underpowered for a WM-capacity analysis, both in terms of the number of items used and the size of the sample tested (a concern that is even more serious for the nonconflict color identification task, where the order effect essentially cuts the sample in half). The fact that participants received few practice trials also suggests that, in all likelihood, many participants performed the experiments while they were still in the process of learning the required stimulus-response mappings. As a result, it is not clear whether any differences related to WM capacity in this situation would be due to this learning process or to variables being manipulated. Finally, it should also be noted that the population of university students might offer a restricted range of WM capacity, making it hard to detect a WM-capacity effect, especially when examining a continuous measure of WM capacity as opposed to using an extreme-groups comparison. Better powered and better designed investigations are needed to clarify whether and how WM capacity influences adaptation to item-specific conflict-frequency and contingency learning.

Challenges and conclusions

The essential message of the present results is that there is a dissociation between contingency learning and item-specific proportion-congruent effects. We interpret these data as suggesting that the two effects reflect qualitatively different phenomena, with the item-specific proportion-congruent effect being mainly a manifestation of a reactive control process of adaptation to item-specific conflict frequency rather resulting completely from a contingency-learning process.

Importantly, since the beginning of the debate on conflict adaptation spurred by the contingency-learning account (Schmidt & Besner, 2008), we are among the first to argue for a role of adaptive control processes in the original, two-item set item-specific proportion-congruent manipulation (Jacoby et al., 2003; for other evidence in support of this position, see Hutcheon & Spieler, 2014; Shedden et al., 2013). We are also aware that this position faces the difficulty of reconciling the present results favoring a conflict-adaptation explanation with previous studies supporting a contingency-learning explanation (Hazeltine & Mordkoff, 2014; Schmidt, 2013a). In those studies, responses to mostly-congruent and mostly-incongruent words presented in incongruent colors, colors that the two types of words appeared in equally often, did not differ from one another, in contrast with the conflict-adaptation prediction that mostly-incongruent incongruent words should be responded to faster than mostly-congruent incongruent words due to the fact that a conflict-adaptation process was, presumably, being implemented in the mostly-incongruent condition.

It is important to note, however, that the design of those studies is different from Jacoby et al.'s (2003) paradigm in potentially important ways. In the two-item set used in Jacoby et al.'s item-specific proportion-congruent manipulation (and in the present experiments), mostly-congruent words appeared in colors that are also mostly-congruent colors, and mostly-incongruent words appeared in colors that are also mostly-incongruent colors. For example, in the version illustrated in Table 2, RED and BLUE function as mostly-congruent words and the red and blue colors also appear mainly with congruent words. Similarly, GREEN and YELLOW are mostly-incongruent words and the colors green and yellow appear mainly with incongruent words. This characteristic of the design might be relevant given recent findings by Bugg et al. (2011, Bugg & Hutchison, 2013) that not only the irrelevant dimension (i.e., the word) but also the relevant dimension (i.e., the color) can function as a signal for conflict frequency. Thus, it is possible that participants can use both word-specific and color-specific information to predict conflict frequency and adapt to it (although in Bugg et al.'s view, there are constraints on the use of color-specific information: Bugg et al., 2011; Bugg & Hutchison, 2013).

What is most relevant to note for present purposes is that word-specific and color-specific conflict frequency provide compatible information in Jacoby et al.'s (2003) two-item set paradigm. For example, the item GREEN$_{\text{yellow}}$ represents both a mostly-incongruent word and a mostly-incongruent color, thus providing a strong bias towards word inhibition. In contrast, word-specific and color-specific conflict frequency provide inconsistent information in some of the cells in Schmidt's (2013a) and Hazeltine and Mordkoff's (2014) four-item set designs. For example, in Schmidt's experiment, the critical comparison for probing conflict adaptation involved mostly-congruent incongruent words and mostly-incongruent incongruent words matched in terms of the frequency that they occurred in the presented (incongruent) color. However, Schmidt's analysis is atypical in that it is based on stimuli that combine words that frequently appear in incongruent colors, i.e., mostly-incongruent words, and colors that frequently appear with congruent words, i.e., mostly-congruent colors. For example, RED and YELLOW were words associated with frequent conflict i.e., (mostly-incongruent words), however, in the crucial conditions in that experiment, they appeared in both blue and green, colors that were associated with infrequent conflict (i.e., most-congruent colors). As such, it is impossible to tell whether and how the contrast between color-specific and word-specific information was resolved for those items. Thus, the comparison between mostly-congruent and mostly-incongruent incongruent words in Schmidt's and Hazeltine and Mordkoff's experiments may be one which is not diagnostic for adjudicating between conflict-adaptation and contingency-learning accounts of the item-specific proportion-congruent effect (see also Spinelli & Lupker, 2019).

Another challenge that our position faces is reconciling the present findings with previous results coming from a control perspective (Bugg et al., 2011; Bugg & Hutchison, 2013), results that, while providing support for a role of control in the item-specific proportion-congruent effect in some circumstances, found no support for control in the two-item set design that we used. In this regard, Bugg and Hutchison's (2013) Experiment 3 is of particular interest. In this experiment, Bugg and Hutchison used both a two-item and a four-item set design of the item-specific proportion-congruent manipulation. In the two-item set design, each word appeared in two colors (one congruent and one incongruent), as in Jacoby et al. (2003) and the present experiments; in the four-item set design, each word appeared in four colors (one congruent color and three incongruent colors). The critical difference between these two versions of the item-specific proportion-congruent manipulation is that while a high-contingency (i.e., more frequent) color existed for mostly-incongruent words in the two-item set design, no high-contingency color existed for mostly-incongruent words in the four-item set design because each word appeared equally frequently in each of the four colors (e.g., RED appeared in red 25% of the time and in each of the three incongruent colors 25% of the time; note that, by necessity, a high-contingency color existed for mostly-congruent words in both designs).

In both designs, an item-specific proportion-congruent effect emerged (i.e., as expected, mostly-incongruent words produced a smaller congruency effect than the corresponding mostly-congruent words), a result that, per se, is compatible with both a contingency-learning and a conflict-adaptation mechanism. What was crucial to adjudicating the mechanism underlying the item-specific proportion-congruent effect, however, was the pattern of results emerging in a new manipulation introduced in the final block of the experiment. In this final block, a new set of colors was added that had not been used before in the experiment, and both mostly-congruent and mostly-incongruent words were presented in those incongruent colors. The rationale for this manipulation was that, if participants learn to focus attention to the color when mostly-incongruent words are presented in the first part of the experiment (i.e., if they apply a conflict-adaptation mechanism), those words should produce less interference even when presented in new incongruent colors than when mostly congruent words are presented in the first part of the experiment. In contrast, if participants learn to associate words with their most likely response in the first part of the experiment (i.e., if they apply contingency-learning mechanism), no advantage for mostly-incongruent words should occur when new incongruent colors are introduced because participants have acquired no information that would allow them to manage conflict more effectively with those words.

What Bugg and Hutchison (2013) found was that mostly-incongruent words did produce shorter latencies than mostly-congruent words when presented in the new incongruent colors in the final block, but only in the four-item version of the task (no difference was observed in the two-item set version). For example, the incongruent color brown (a color used only in the final block of the experiment) was named faster if that color appeared in a mostly-incongruent word than if it appeared in a mostly-congruent word, but only for participants who completed the four-item set version of the experiment initially. Based on these results, Bugg and Hutchison (2013) concluded that distinct mechanisms are involved in the two-item and the four-item set design: In the four-item set design, conflict adaptation would be the dominant mechanism, as demonstrated by the fact that, for the new colors in the final block of their experiment, participants imported previously acquired information about item-specific conflict frequency. In contrast, in the two-item set design, contingency learning would be the dominant mechanism, as demonstrated by the fact that no such transfer of information was observed for the new colors in the final block in that situation. Yet, in the present experiments, we found good evidence in support of conflict adaptation playing an important role in the two-item set design. What could be the cause for this inconsistency?

A possible explanation is that in Bugg and Hutchison's (2013) manipulation, the introduction of new colors in the final block may have discouraged individuals from transferring knowledge about item-specific conflict frequency acquired from the set of stimuli appearing in the first part of the experiment. If so, the failure to observe a transfer effect in the final block (i.e., there was not less interference for mostly-incongruent than mostly-congruent words on the new incongruent colors) cannot be used to conclude that no conflict-adaptation process had been used in the first part of the experiment.

It is possible that conflict adaptation was engaged in both the version of the task that produced transfer in Bugg and Hutchison's paradigm (e.g., the four-item set version) and the version of the task that did not produce transfer (e.g., the two-item set version), with the presence of transfer depending on more marginal factors.

One possibility, for example, is that participants in a two-item set design are more likely than participants in a four-item set design to become consciously aware of the item-specific proportion-congruent manipulation because they are exposed to a more limited number of stimuli (8 color-word combinations in a two-item set design vs. 16 color-word combinations in a four-item set design). Upon noticing the new colors in the final block, participants in the two-item set version may deliberately decide to reset their control settings early in that block, thus purging any item-specific conflict information that they had previously acquired, albeit without becoming aware of having done so. The situation might be different in a four-item set design because, in that scenario, item-specific conflict frequency information may be more frequently learned outside the focus of awareness. Because item-specific conflict frequency information is acquired in a more subtle manner, participants may not feel particularly compelled to reset their control settings in the final block, with item-specific conflict frequency maintaining some impact on performance even for the new colors. Although this hypothesis is purely speculative, it would seem to provide a reasonable explanation for the inconsistency between Bugg and Hutchison's (2013) data and ours (see also Schmidt, 2014, 2019, for another explanation of Bugg and Hutchison's data which assumes that the transfer effect observed in the final block of the four-item set version has, in fact, nothing to do with conflict adaptation).

Clearly, further research is needed to examine more closely the contribution of contingency learning and item-specific conflict adaptation to the item-specific proportion-congruent effect. What the present results suggest, however, is that there might be more to adaptation to item-specific conflict frequency than supporters of the contingency-learning and the control accounts currently believe. The reactive use of associations between words (and/or colors) and their appropriate control setting, in addition to, or as an alternative to, the use of associations between words and motor responses, might be an important cognitive tool in managing item-specific conflict frequency.

Footnotes

1. This reasoning does not imply that proactive control could explain the item-specific proportion-congruent effect because, as noted, this effect must depend on a process initiated in response to specific items. If a proactive process of maintaining the color-naming goal were used for both mostly-congruent and mostly-incongruent items, this process would presumably produce a reduced congruency effect in the task in general, but would not cause differential congruency effects for the two types of items. Thus, although proactive control can be used concurrently with reactive control, only reactive control can provide an explanation for the item-specific proportion-congruent effect (Gonthier et al., 2016).
2. We excluded those trials to avoid including trials in the analyses in which participants had failed to maintain the memory load. For these and the following experiments, we also conducted parallel analyses in which the trials on which participants made an error on the WM task were not excluded. The results were virtually identical in all cases.
3. In addition to the regular analyses on raw RTs, for these and the following experiments, we also conducted parallel analyses on z-score transformed RTs (Faust, Balota, Spieler, & Ferraro, 1999) to determine if the WM-load effects of interest would emerge in terms of proportional changes from baseline. Again, the results were virtually identical in all cases.
4. For this and the following Stroop experiments (Experiments 2B and 3B), we conducted another set of analyses using Contingency (high vs. low) as a factor instead of Item Type (mostly congruent vs. mostly incongruent). Mostly-congruent congruent words and mostly-incongruent incongruent words would be the high-contingency items; mostly-incongruent congruent words and mostly-congruent incongruent words would the low-contingency items. This type of analysis offers a direct parallel to the analysis for the nonconflict color identification task because it allows an evaluation of the interaction between Contingency and WM Load in both types of

tasks. To preview the results, in the Stroop task (Experiments 1B, 2B, and 3B), the interaction between Contingency and WM Load (corresponding, statistically, to the three-way interaction between Congruency, Item Type, and WM Load in the analysis with Item Type as a factor) never approached significance.

5. Another potential reason for the failure to observe a significant reduction in the contingency-learning effect with increasing WM load in Experiments 1A and 1B is that the present load procedure might have led to an underestimation of load effects. Because chance performance was 50% in the two-alternative forced choice WM task that we used, on a significant proportion of trials, participants might have simply guessed the correct answer. As a result, color-naming latencies on those trials would have been included in the analyses even though participants were not necessarily maintaining a WM load during those trials. Although using a WM task without a two-alternative forced choice procedure would have been a reasonable way to minimize this problem in the subsequent experiments, the strategy that we pursued instead was to reproduce the conditions under which Schmidt et al. (2010) obtained their pattern (reduced contingency-learning effects with increasing WM load) as closely as possible. Because the two-alternative forced choice WM task used in Experiments 1A and 1B was the same as that used by Schmidt et al. (2010), for consistency's sake, we decided to maintain their load procedure in the following experiments.

6. This result, not reported in the original article, was obtained by re-analyzing Hutchison's data from the low- and high-contingency incongruent items using contingency (low vs. high) and WM-capacity group (low capacity vs. high capacity) as variables in a split-plot ANOVA. The results indicated a main effect of contingency in the RTs, *F*(1, 84) = 14.13, *MSE* = 1828, *p* < .001, η_p^2 = .144, but not in the error rates, *F* < 1. Most importantly, contingency learning did not interact with WM-capacity in either analysis (both *F*s < 1), indicating that the contingency-

learning effects in this experiment were equivalent for low and high WM-capacity individuals in both RTs (23 and 26 ms respectively) and error rates (0.6% and 0.4%, respectively).

7. We used a 75% cut-off (based on performance in the three complex span tasks) because the commonly used 85% cut-off (e.g., Unsworth et al., 2005) resulted in the exclusion of quite a large number of participants (i.e., 58, that is 46% of the initial 126 participants), thus severely limiting the statistical power of the WM-capacity analysis. Indeed, the pattern of results obtained with an 85% cut-off was numerically equivalent to that obtained using a 75% cut-off, but some of the effects did not quite reach statistical significance in the 85% cut-off analyses.
8. Note that because the contingency-learning effect is relatively small in vocal responding (as shown in Experiment 1A) but a robust item-specific proportion-congruent effect is regularly observed when this response modality is used (as shown in Experiment 1B), one might conclude that, in vocal responding, the item-specific proportion-congruent effect might primarily reflect the action of a conflict-adaptation process rather than that of a contingency-learning process. However, the results of a recent item-specific proportion-congruent manipulation in our lab suggest a more cautious conclusion (Spinelli & Lupker, 2019). In that experiment, the design permitted us to dissociate the independent contributions of contingency learning and adaptation to item-specific conflict frequency to the item-specific proportion-congruent effect. Although a vocal response was required, a robust contingency-learning effect emerged in that situation in addition to a (smaller) effect of adaptation to item-specific conflict frequency. Thus, although in the present Experiment 1A, a non-conflict color identification task with vocal responding, we did not obtain a large contingency-learning effect, contingency learning likely has some role in the item-specific proportion-congruent effect in the Stroop task, even when a vocal response is required (see also Hutchison, 2011).

9. One may object that the reason that we failed to find a significant reduction in the item-specific proportion-congruent effect with increasing WM load is that, because WM load was manipulated between subjects, our experiments did not have enough power to detect that interaction. To alleviate that concern, we conducted an additional set of analyses on the combined the data from Experiments 2A and 3A and Experiments 2B and 3B (Experiment (2A vs. 3A; 2B vs. 3B) had no impact in either analysis and was dropped as a factor). Not surprisingly, the combined analysis of Experiments 2A and 3A revealed that increasing WM load significantly reduced contingency-learning effects in the non-conflict color identification task in the latencies, *F*(2, 265) = 14.81, *MSE* = 1887, *p* < .001, η_p^2 = .101, and marginally so in the error rates, *F*(2, 265) = 2.52, *MSE* = .001, *p* = .083, η_p^2 = .019. However, there was no hint in the combined analysis of Experiments 2B and 3B that WM load produced a reduction in the item-specific proportion-congruent effect in the Stroop task, i.e., there was no three-way interaction between Congruency, Item Type, and WM load, *F*(2, 265) = .22, *MSE* = 4747, *p* = .80, η_p^2 = .002 for the latencies, *F*(2, 265) = .93, *MSE* = .002, *p* = .40, η_p^2 = .007 for the error rates. In fact, the Bayes Factors for both the latencies, BF_{01} = 18.29 ± 3.23, and the error rates, BF_{01} = 10.23 ± 4.09%, indicated “strong” evidence for the absence of the three-way interaction.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*, 693-728.

Atalay, N. B., & Misirlisoy, M. (2012). Can contingency learning alone account for item-specific control? Evidence from within-and between-language ISPC effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1578-1590.

Atalay, N. B., & Misirlisoy, M. (2014). ISPC effect is not observed when the word comes too late: A time course analysis. *Frontiers in Psychology*, *5*, 1410.

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Balota, D., Aschenbrenner, A., & Yap, M. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*, 1563-1571.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Blais, C., Robidoux, S., Risko, E. F., & Besner, D. (2007). Item-specific adaptation and the conflict-monitoring hypothesis: A computational model. *Psychological Review*, *114*, 1076-1086.

Bolker, B. (2019). *GLMM FAQ*. Retrieved from https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106-113.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in working memory* (pp. 76-108). Oxford, UK: Oxford University Press.

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 567-587.

Bugg, J. M. (2015). The relative attractiveness of distractors and targets affects the coming and going of item-specific control: Evidence from flanker tasks. *Attention, Perception, & Psychophysics*, *77*, 373-389.

Bugg, J. M., & Chanani, S. (2011). List-wide control is not entirely elusive: Evidence from picture–word Stroop. *Psychonomic Bulletin & Review*, *18*, 930-936.

Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*, 367.

Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 433-449.

Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011). Why it is too early to lose control in accounts of item-specific proportion congruency effects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 844-859.

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*, 1484-1494.

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: effects of interference expectancy and fluid intelligence. *PloS One*, *5*, e12861.

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect*. Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1703-1723.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.

Crump, M. J., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review*, *13*, 316-321.

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*, 1322-1326.

Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*, 777-799.

Forrin, N. D., & MacLeod, C. M. (2017). Relative speed of processing determines color–word contingency learning. *Memory & Cognition*, *45*, 1206-1222.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226-236.

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*, 778-788.

Hazeltine, E., & Mordkoff, J. T. (2014). Resolved but not forgotten: Stroop conflict dredges up the past. *Frontiers in Psychology*, *5*, 1327.

Hutcheon, T. G., & Spieler, D. H. (2014). Contextual influences on the sequential congruency effect. *Psychonomic Bulletin & Review*, *21*, 155-162.

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 851-860.

Hutchison, K. A., Bugg, J. M., Lim, Y. B., & Olsen, M. R. (2016). Congruency precues moderate item-specific proportion congruency effects. *Attention, Perception, & Psychophysics*, *78*, 1087-1103.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*, 638-644.

Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, *22*, 869-875.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-271.

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*, 75-82.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means.

Levin, Y., & Tzelgov, J. (2016). Contingency learning is not affected by conflict experience: Evidence from a task conflict-free, item-specific Stroop paradigm. *Acta Psychologica*, *164*, 39-45.

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher order interactions. *arXiv*:1405.2094.

Lin, O. Y. H., & MacLeod, C. M. (2018). The acquisition of simple associations as observed in color–word contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 99-106.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166-174.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, *109*, 163-203.

Marini, F., Demeter, E., Roberts, K. C., Chelazzi, L., & Woldorff, M. G. (2016). Orchestrating proactive and reactive mechanisms for filtering distracting information: Brain-behavior relationships revealed by a mixed-design fMRI study. *Journal of Neuroscience*, *36*, 988-1000.

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 748-759.

Meier, M. E., & Kane, M. J. (2015). Carving executive control at its joints: Working memory capacity predicts stimulus–stimulus, but not stimulus–response, conflict. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1849-1872.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, *21*, 627-645.

Musen, G., & Squire, L. R. (1993). Implicit learning of color-word associations using a Stroop paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 789-798.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859-862.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*, 164-171.

Schmidt, J. R. (2013a). The Parallel Episodic Processing (PEP) model: Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychologica*, *142*, 119-126.

Schmidt, J. R. (2013b). Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20*, 615-630.

Schmidt, J. R. (2013c). Temporal learning and list-level proportion congruency: conflict adaptation or learning when to respond?. *PLoS One*, *8*, e82320.

Schmidt, J. R. (2014). Contingencies and attentional capture: the importance of matching stimulus informativeness in the item-specific proportion congruent task. *Frontiers in Psychology*, *5*, 540.

Schmidt, J. R. (2016). Time-out for conflict monitoring theory: Preventing rhythmic biases eliminates the list-level proportion congruent effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*, 52-62.

Schmidt, J. R. (2018). Best not to bet on the horserace: A comment on Forrin and MacLeod (2017) and a relevant stimulus-response compatibility view of colour-word contingency learning asymmetries. *Memory & Cognition*, *46*, 326-335.

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, 1-19. Advance online publication.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 514-523.

Schmidt, J. R., Crump, M. J., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*, 421-435.

Schmidt, J. R., & De Houwer, J. (2016). Time course of colour-word contingency learning: Practice curves, pre-exposure benefits, unlearning, and relearning. *Learning and Motivation*, *56*, 15-30.

Schmidt, J. R., De Houwer, J., & Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, *19*, 235-250.

Schmidt, J. R., Notebaert, W., & Bussche, E. V. D. (2015). Is conflict adaptation an illusion?. *Frontiers in Psychology*, *6*, 172.

Shedden, J. M., Milliken, B., Watter, S., & Monteiro, S. (2013). Event-related potentials as brain correlates of item specific proportion congruent effects. *Consciousness and Cognition*, *22*, 1442-1455.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*, 4-7.

Singmann, H., & Kellen, D. (2018). An Introduction to Mixed Models for Experimental Psychology. In D. Spieler, & E. Schumacher (Eds.), *New Methods in Cognitive Psychology*. New York: Routledge.

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 155-167.

Spinelli, G., & Lupker, S. L. (2019). Item-specific control of attention in the Stroop task: Contingency learning is not the whole story in the item-specific proportion-congruent effect. *Memory & Cognition*. Advance online publication.

Spinelli, G., & Lupker, S. L. (2020). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication.

Spinelli, G., Perry, J. R., Lupker, S. L. (under review). The role of stimulus-response compatibility and feedback in color-word contingency learning.

Spinelli, G., Perry, J. R., Lupker, S. L. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture-word interference task. *Journal of Experimental Psychology: Human Perception and Performance, 45,* 995-1014*.*

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*, 127-154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, *13*, 304-319.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. New York: Springer.

Author notes

This research was partially supported by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker and the Ontario Trillium Scholarship to Giacomo Spinelli. We would like to thank Keith Hutchison for sharing his data with us and Ken Paap for sharing his questionnaire with us.

The raw data and the SPSS and R files used for the analyses are publicly available at https://osf.io/rtnw2/.

Correspondence concerning this article may be addressed to: Giacomo Spinelli, Department of Psychology, University of Western Ontario, London, Ontario, N6A 5C2, Canada (e-mail: gspinel@uwo.ca) or Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, N6A 5C2, Canada (e-mail: lupker@uwo.ca).

Table 1

Template for the Frequency of Color-Word Combinations in Experiment 1A

	Word			
Color	SHOP	CULT	BRAG	WIDE
Red	36	12		
Blue	12	36		
Green			36	12
Yellow			12	36

Table 2

Template for the Frequency of Color-Word Combinations in Experiment 1B

	Word			
	Mostly-congruent words		Mostly-incongruent words	
Color	RED	BLUE	GREEN	YELLOW
Red	36	12		
Blue	12	36		
Green			12	36
Yellow			36	12

Table 3

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 1A – Vocal Nonconflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	589 (16)	.010 (.003)
Low	601 (18)	.013 (.005)
Contingency effect	12	.003
Low load		
High	786 (35)	.005 (.002)
Low	801 (37)	.009 (.003)
Contingency effect	15	.004
High load		
High	743 (30)	.005 (.001)
Low	752 (28)	.004 (.002)
Contingency effect	9	-.001

Table 4

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 1B – Vocal Stroop Task

Congruency	RTs		Error rates	
	Mostly-congruent items	Mostly-incongruent items	Mostly-congruent items	Mostly-incongruent items
No load				
Congruent	626 (17)	665 (23)	.001 (.001)	.008 (.006)
Incongruent	751 (25)	715 (18)	.071 (.017)	.024 (.007)
Congruency Effect	125	50	.070	.016
Low load				
Congruent	726 (23)	757 (24)	.001 (.001)	.002 (.002)
Incongruent	855 (30)	794 (22)	.011 (.006)	.007 (.003)
Congruency Effect	129	37	.010	.005
High load				
Congruent	793 (33)	821 (30)	.000 (.000)	.005 (.004)
Incongruent	899 (29)	849 (32)	.053 (.012)	.017 (.007)
Congruency Effect	106	28	.053	.012

Table 5

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 2A – Manual NonConflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	712 (26)	.025 (.005)
Low	769 (25)	.036 (.008)
Contingency effect	57	.011
Low load		
High	829 (28)	.027 (.005)
Low	857 (27)	.036 (.007)
Contingency effect	28	.009
High load		
High	843 (29)	.035 (.006)
Low	855 (27)	.044 (.009)
Contingency effect	12	.009

Table 6

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 2B – Manual Stroop Task

	RTs		Error rates	
Congruency	Mostly-congruent items	Mostly-incongruent items	Mostly-congruent items	Mostly-incongruent items
<u>No load</u>				
Congruent	766 (29)	817 (33)	.008 (.005)	.011 (.005)
Incongruent	940 (41)	885 (36)	.021 (.009)	.013 (.006)
Congruency Effect	174	68	.013	.002
<u>Low load</u>				
Congruent	794 (30)	857 (36)	.021 (.007)	.020 (.007)
Incongruent	947 (33)	904 (35)	.077 (.013)	.036 (.007)
Congruency Effect	153	47	.056	.016
<u>High load</u>				
Congruent	877 (25)	916 (26)	.028 (.007)	.027 (.012)
Incongruent	1032 (28)	923 (24)	.058 (.013)	.030 (.005)
Congruency Effect	155	7	.030	.003

Table 7

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 3A – Manual NonConflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	665 (7)	.021 (.002)
Low	728 (10)	.036 (.004)
Contingency effect	63	.015
Low load		
High	783 (18)	.028 (.004)
Low	814 (18)	.039 (.008)
Contingency effect	31	.011
High load		
High	846 (25)	.030 (.004)
Low	863 (23)	.027 (.007)
Contingency effect	17	-.003

Table 8

Mean RTs and Error Rates (and corresponding Standard Errors) for Experiment 3B – Manual Stroop Task

	RTs		Error rates	
Congruency	Mostly-congruent items	Mostly-incongruent items	Mostly-congruent items	Mostly-incongruent items
No load				
Congruent	699 (10)	771 (14)	.018 (.003)	.029 (.006)
Incongruent	885 (14)	829 (13)	.061 (.008)	.041 (.004)
Congruency Effect	186	58	.043	.012
Low load				
Congruent	825 (24)	878 (24)	.022 (.005)	.033 (.010)
Incongruent	1012 (26)	948 (21)	.074 (.014)	.048 (.007)
Congruency Effect	187	70	.052	.015
High load				
Congruent	869 (26)	920 (28)	.026 (.004)	.023 (.009)
Incongruent	1022 (24)	974 (21)	.048 (.011)	.038 (.007)
Congruency Effect	153	54	.022	.015

Table A1

Mean RTs and Error Rates (and corresponding Standard Errors) for Low and High WM-Capacity groups in Experiment 3A – Manual NonConflict Color Identification Task

Contingency	RTs	Error rates
Low WM capacity		
High	657 (16)	2.9 (.7)
Low	720 (18)	6.3 (.9)
Contingency effect	63	3.4
Medium-low WM capacity		
High	726 (13)	1.7 (.4)
Low	801 (18)	2.5 (.7)
Contingency effect	75	.8
Medium-high WM capacity		
High	635 (14)	1.6 (.4)
Low	696 (20)	1.7 (.6)
Contingency effect	61	.1
High WM capacity		
High	637 (15)	1.3 (.3)
Low	704 (20)	3.4 (.9)
Contingency effect	67	2.1

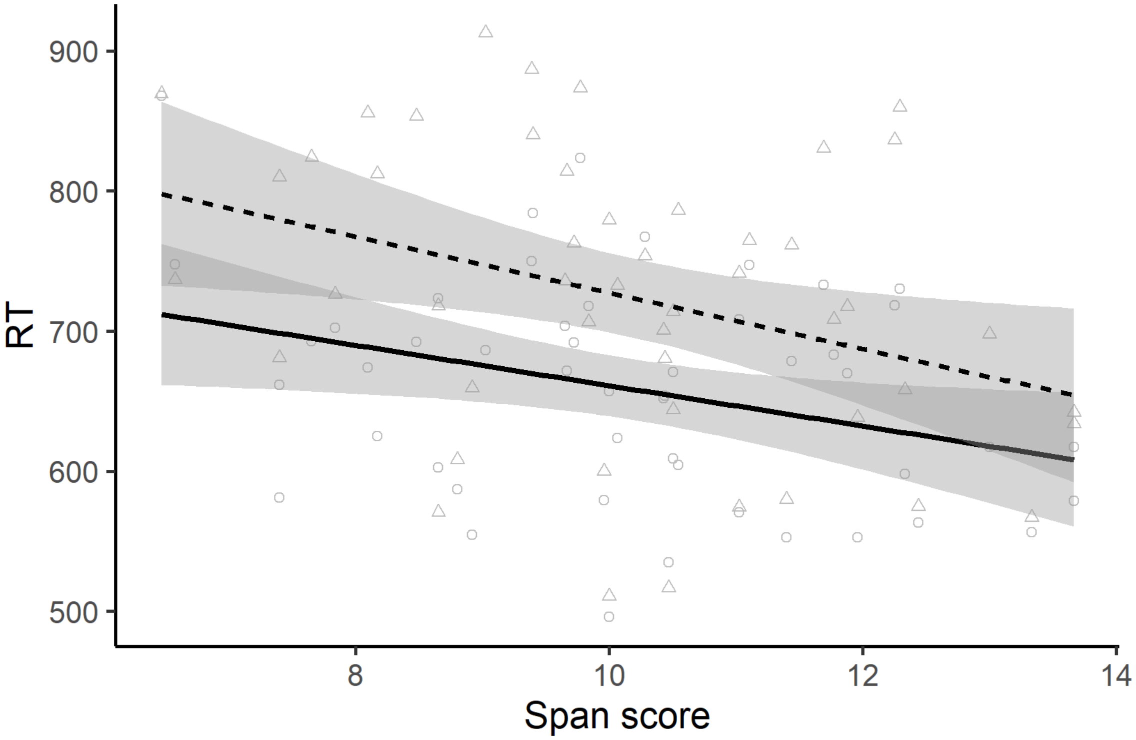
Table A2

Mean RTs and Error Rates (and corresponding Standard Errors) for Participants in the Four WM-Capacity Quartiles in Experiment 3B – Manual Stroop task

	RTs		Error rates	
Congruency	Mostly-congruent items	Mostly-incongruent items	Mostly-congruent items	Mostly-incongruent items
Low WM capacity				
Congruent	702 (25)	761 (32)	.027 (.008)	.039 (.011)
Incongruent	918 (35)	844 (30)	.112 (.024)	.058 (.012)
Congruency Effect	216	83	.085	.019
Medium-low WM capacity				
Congruent	742 (23)	817 (31)	.015 (.006)	.013 (.008)
Incongruent	902 (34)	876 (29)	.048 (.014)	.027 (.008)
Congruency Effect	160	59	.033	.014
Medium-high WM capacity				
Congruent	660 (19)	750 (30)	.011 (.005)	.042 (.021)
Incongruent	834 (27)	798 (25)	.057 (.016)	.034 (.010)
Congruency Effect	174	48	.046	-.008
High WM capacity				
Congruent	666 (21)	721 (27)	.015 (.006)	.007 (.005)
Incongruent	874 (36)	777 (25)	.035 (.012)	.033 (.008)
Congruency Effect	208	56	.020	.026

Figure 1

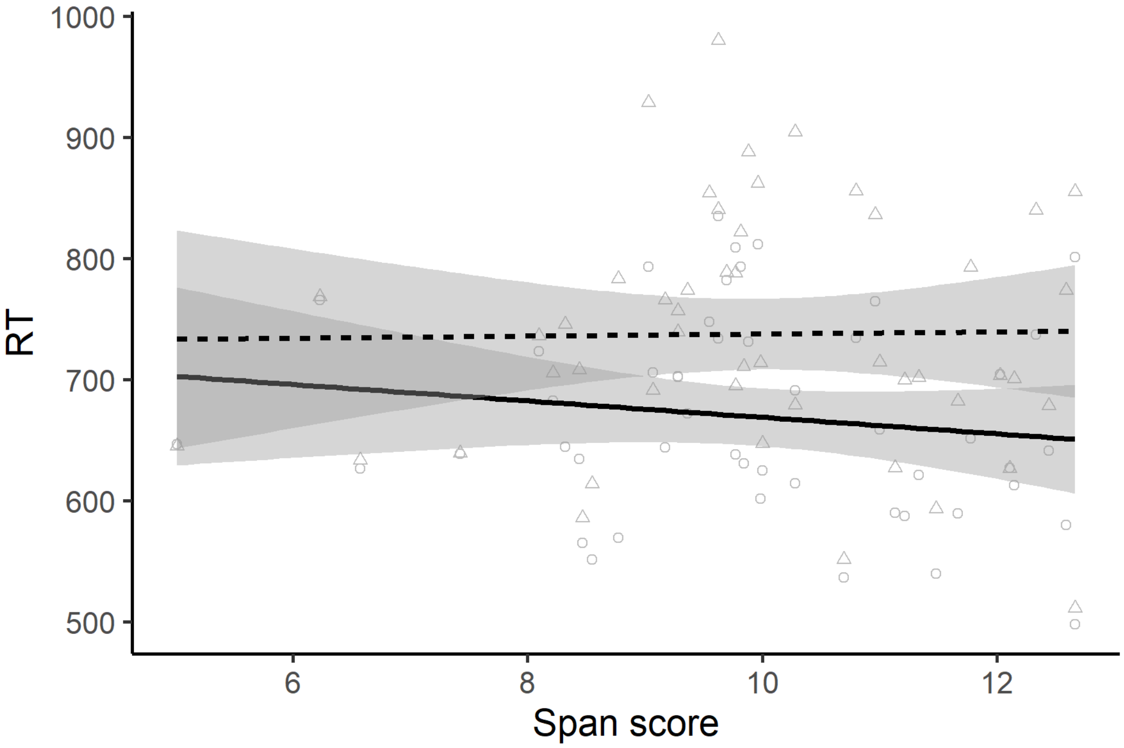
The impact of Span score on the contingency-learning effect in latencies for no-load participants in Experiment 3A who did Experiment 3A first



Note. For each participant, the mean latency for high- and low-contingency items is marked with a circle and a triangle, respectively. Regression slopes (with 95% confidence interval bands) for high- and low-contingency items are marked with a solid line and a dashed line, respectively.

Figure 2

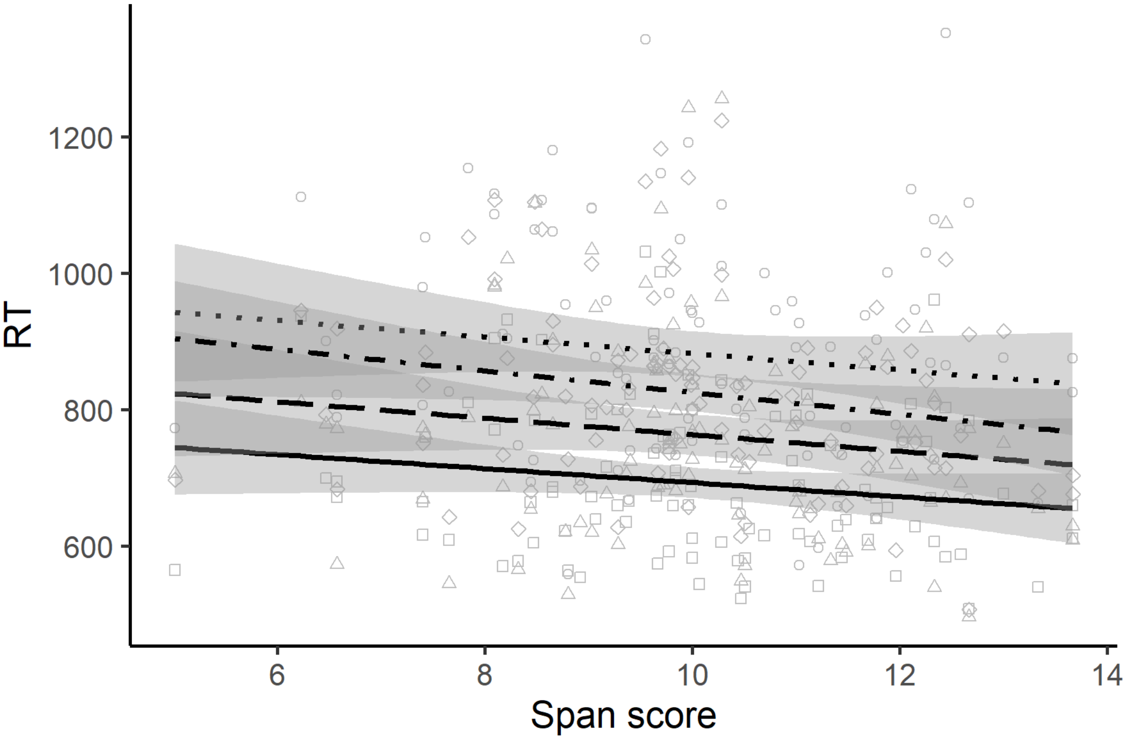
The impact of Span score on the contingency-learning effect in latencies for no-load participants in Experiment 3A who did Experiment 3A following Experiment 3B



Note. For each participant, the mean latency for high- and low-contingency items is marked with a circle and a triangle, respectively. Regression slopes (with 95% confidence interval bands) for high- and low-contingency items are marked with a solid line and a dashed line, respectively.

Figure 3

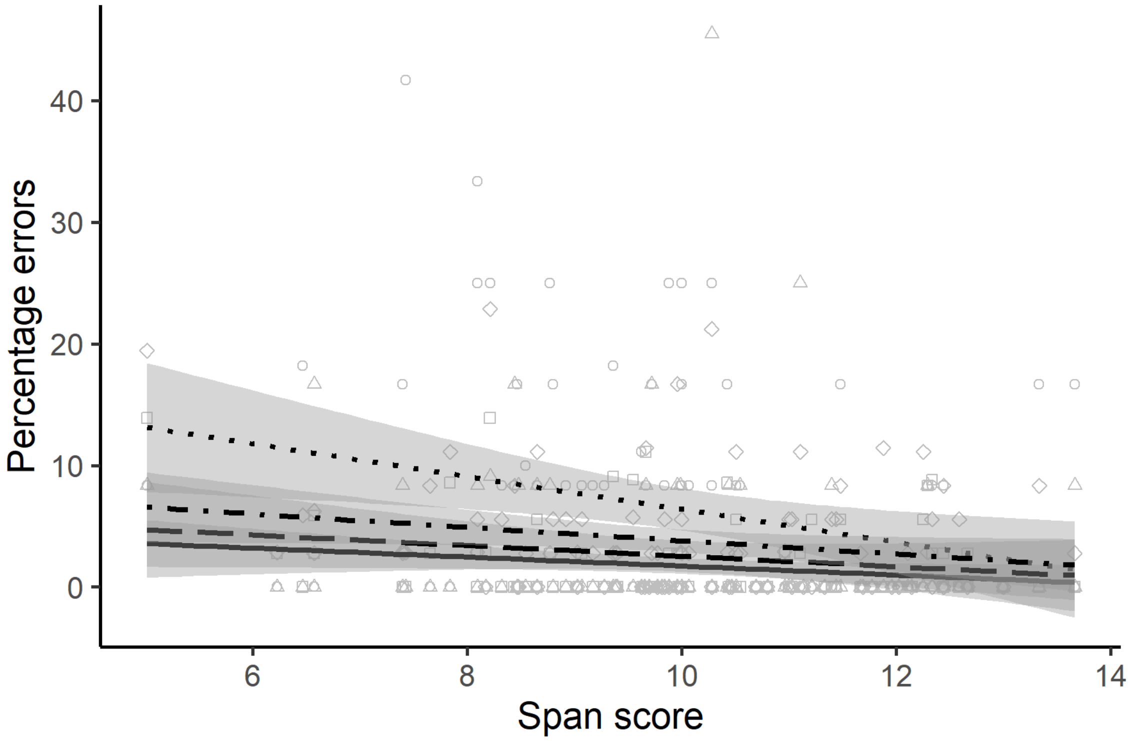
The impact of Span score on the item-specific proportion-congruent effect in latencies for no-load participants in Experiment 3B



Note. For each participant, the mean latency for congruent items in the mostly-congruent condition, incongruent items in the mostly congruent condition, congruent items in the mostly-incongruent condition, and incongruent items in the mostly-incongruent condition, is marked with a square, a circle, a triangle, and a rhombus, respectively. Regression slopes (with 95% confidence interval bands) for congruent items in the mostly-congruent condition, incongruent items in the mostly congruent condition, congruent items in the mostly-incongruent condition, and incongruent items in the mostly-incongruent condition, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Figure 4

The impact of Span score on the item-specific proportion-congruent effect in error rates for no-load participants in Experiment 3B



Note. For each participant, the mean latency for congruent items in the mostly-congruent condition, incongruent items in the mostly congruent condition, congruent items in the mostly-incongruent condition, and incongruent items in the mostly-incongruent condition, is marked with a square, a circle, a triangle, and a rhombus, respectively. Regression slopes (with 95% confidence interval bands) for congruent items in the mostly-congruent condition, incongruent items in the mostly congruent condition, congruent items in the mostly-incongruent condition, and incongruent items in the mostly-incongruent condition, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Appendix

WM-capacity analysis – extreme-groups ANOVA

The data for participants in the no-load condition of Experiments 3A and 3B were analyzed with an extreme-groups ANOVA based on participants' means in each condition in addition to the full-sample, mixed-effects analysis based on unaggregated data reported in the main text. As in Hutchison (2011), the extreme groups were determined with a quartile split based on the composite score of WM capacity obtained from the complex span tasks. The first quartile (composed of twenty-four participants) was classified as the low WM-capacity group and the last quartile (composed of another twenty-four participants) was classified as the high WM-capacity group. The ANOVA was conducted with the same factors as the fixed effects of the mixed-effects analysis, except that Span Score was replaced with WM Capacity. To preview the results, Order had no impact on the most relevant interactions, i.e., the interaction between Contingency and WM Capacity in Experiment 3A and the interaction between Congruency, Item Type and WM Capacity in Experiment 3B. Thus, we present the mean RTs and error rates for the four quartiles (the first and the last quartiles plus the middle two quartiles) in Tables A1 and A2 for Experiments 3A and 3B, respectively, without splitting the data by Order.

-Tables A1 and A2 around here-

Experiment 3A

RTs. Contingency (high-contingency faster than low-contingency) was the only significant effect, $F(1, 44) = 50.36$, $MSE = 1991$, $p < .001$, $\eta_p^2 = .534$. Although individuals with high WM Capacity were numerically faster than individuals with low WM Capacity, WM Capacity was not statistically significant, $F(1, 44) = .52$, $MSE = 12610$, $p = .47$, $\eta_p^2 = .012$. In addition, replicating Hutchison (2011) (see footnote 6), WM

Capacity did not interact with Contingency, *F*(1, 44) = .07, *MSE* = 1991, *p* = .80, η_p^2 = .001, indicating equivalent contingency-learning effects for the low (63 ms) and the high WM Capacity groups (67 ms). Bayesian analyses revealed that there was "moderate" evidence for the absence of the interaction, BF_{01} = 3.31 ± 3.13%.

Error rates. There was a main effect of Contingency (high-contingency more accurate than low-contingency), *F*(1, 44) = 18.52, *MSE* = .001, *p* < .001, η_p^2 = .296, and WM Capacity (high WM-capacity group more accurate than low WM-capacity group), *F*(1, 44) = 7.66, *MSE* = .002, *p* = .008, η_p^2 = .148. There was no interaction between Contingency and WM Capacity, however, *F*(1, 44) = .84, *MSE* = .001, *p* = .37, η_p^2 = .019, indicating that the contingency-learning effects for the low (3.4%) and high WM Capacity groups (2.1%) were equivalent. In the Bayesian analyses, the evidence for the absence of this interaction, however, was only "anecdotal", BF_{01} = 2.18 ± 6.61%.

Experiment 3B

RTs. There was a main effect of Congruency (congruent faster than incongruent), *F*(1, 44) = 179.64, *MSE* = 5237, *p* < .001, η_p^2 = .803, and an interaction between Congruency and Item Type, *F*(1, 44) = 39.57, *MSE* = 6021, *p* < .001, η_p^2 = .473. The interaction indicated an item-specific proportion-congruent effect, with a larger congruency effect for mostly-congruent items (211 ms) than for mostly-incongruent items (70 ms). Although the high WM-capacity group was numerically faster than the low WM-capacity group, WM Capacity did not approach statistical significance, *F*(1, 44) = 1.44, *MSE* = 62296, *p* = .237, η_p^2 = .032. In addition, WM Capacity did not modulate the pattern of item-specific proportion-congruent effects, i.e., there was no three-way interaction between Congruency, Item Type, and WM Capacity, *F*(1, 44) = .33, *MSE* = 6021, *p* = .571, η_p^2 = .007. The Bayes Factor, BF_{01} = 3.01 ± 7.38%, indicated "moderate" evidence for the absence of this three-way interaction. Finally, there was a marginal three-way interaction between Congruency, Order, and WM Capacity, $F(1, 44) = 3.98$, $MSE = 5237$, $p = .052$, $\eta_p^2 = .083$. This interaction indicated a numerical tendency for high WM-capacity participants to show overall smaller congruency effects than low WM-capacity participants, but only for participants who did Experiment 3B following Experiment 3A.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), $F(1, 44) = 21.96$, $MSE = .003$, $p < .001$, $\eta_p^2 = .333$, WM Capacity (high WM-capacity group more accurate than low WM-capacity group), $F(1, 44) = 15.56$, $MSE = .005$, $p < .001$, $\eta_p^2 = .261$, and a marginal effect of Item Type, $F(1, 44) = 3.50$, $MSE = .003$, $p = .068$, $\eta_p^2 = .074$, indicating a tendency for mostly-incongruent items to be more accurate than mostly-congruent items. WM Capacity interacted with Order, $F(1, 44) = 5.51$, $MSE = .005$, $p = .023$, $\eta_p^2 = .111$, indicating that WM Capacity had a larger impact on error rates for participants who did Experiment 3B first (low WM-capacity: 7.2%; high WM-capacity: 2.2%) than for those who did Experiment 3B following Experiment 3A (low WM-capacity: 4.2%; high WM-capacity: 2.9%). The Congruency by Item Type interaction, with larger congruency effects for mostly-congruent items than for mostly-incongruent items, was marginal, $F(1, 44) = 3.73$, $MSE = .003$, $p = .060$, $\eta_p^2 = .078$. Congruency marginally interacted with WM Capacity as well, $F(1, 44) = 3.88$, $MSE = .003$, $p = .055$, $\eta_p^2 = .081$, indicating that congruency effects tended to be smaller for the high WM-capacity group. Most importantly, these two-way interactions were qualified by a three-way interaction between Congruency, Item Type, and WM Capacity, $F(1, 44) = 5.25$, $MSE = .003$, $p = .027$, $\eta_p^2 = .107$.

To explore the three-way interaction, low and high WM-capacity groups were analyzed separately. In the low WM-capacity group, there was a main effect of Congruency, $F(1, 22) = 14.80$, $MSE = .005$, $p = .001$, $\eta_p^2 = .402$, and a Congruency by Item Type interaction, $F(1, 22) = 5.42$, $MSE = .005$, $p = .030$, $\eta_p^2 = .198$. This interaction indicated a regular item-specific proportion-congruent effect, with a larger congruency effect for mostly-congruent items (8.5%) than for mostly-incongruent items (1.9%). In the high WM-capacity group, on the other hand, the only significant effect was that of Congruency, *F*(1, 22) = 7.34, *MSE* = .002, *p* = .013, η_p^2 = .250, with no evidence of an item-specific proportion-congruent effect. Indeed, the congruency effect for mostly-congruent items (2%) was slightly smaller than the congruency effect for mostly-incongruent items (2.6%). This pattern of results is also a replication of Hutchison’s (2011) results, although it is not supported by the results of the full-sample analysis (see main text).