11-1-2019 2:30 PM

Adaptation to Conflict Frequency: Non-Conflict Learning Is Not the Whole Story

Giacomo Spinelli, *The University of Western Ontario*

Supervisor: Lupker, Stephen J., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

Abstract

In the Stroop task, smaller congruency effects (i.e., the color-naming difference between incongruent items, e.g., the word RED in the color blue, and congruent items, e.g., RED in red) are found in conditions in which incongruent items are frequent vs. infrequent. Although the traditional explanation for these "Proportion-Congruent effects" is that attention to task-relevant information is more focused in frequently-conflicting conditions (a process involving adaptation to conflict frequency), Proportion-Congruent paradigms typically have not controlled for the impact of more general learning processes, particularly 1) learning of word-response contingencies (contingency learning), 2) learning about the predictive nature of the stimuli (stimulus informativeness), and 3) learning about response rhythm in the task (temporal learning), processes which could produce the Proportion-Congruent effects obtained in most situations. The present research examined the possibility that those non-conflict learning processes are indeed the whole story in Proportion-Congruent effects. Several different approaches were used. First, the proportion of congruent and incongruent items in a list was manipulated in a variant of the Stroop task in which no individual stimulus was repeated, creating a situation in which neither contingency learning nor stimulus informativeness could have influenced task performance. Second, manipulating the proportion of neutral (i.e., consonant strings) and incongruent items in a list allowed the creation of a parallel situation in the classic color-word Stroop task. A Proportion-Congruent effect and a similar, Proportion-Neutral effect, emerged in both tasks even though contingency learning and stimulus informativeness could have played no role in producing those effects. Further, attempts to examine the influence of temporal learning failed to show any evidence of that process contributing to those effects either. The final set of experiments involved a congruency-proportion manipulation specific to individual words within the same list. Contrary to the idea that the Proportion-Congruent effect obtained in this situation results from contingency learning, a concurrent working memory load impaired contingency learning in a non-conflict color identification task but spared the Proportion-Congruent effect in the Stroop task, favoring a conflict-adaptation interpretation of this effect. Overall, these results support the existence of a process of adaptation to conflict frequency in the human control system.

Keywords: conflict adaptation; Stroop; contingency learning; temporal learning; proportion-congruent effect; proactive control; reactive control; working memory

Summary for Lay Audience

This research was an examination of the processes that individuals use when dealing with distraction created by irrelevant but salient events (e.g., the type of situation created when a smartphone notification occurs while driving). The more specific focus was on the processes that individuals use when they can anticipate that a distracting event will occur. For example, it has normally been assumed that individuals can learn to increase attention to the current goal in situations in which distracting events are frequent. As a result of using this process of adaptation to distraction frequency, distraction becomes less disruptive in those situations than in situations in which distracting events are infrequent. Although many theories of how we deal with distraction assume that humans can and do use this process, recent research has suggested that the crucial experimental finding on which that assumption is based upon (i.e., that distraction is less disruptive in frequently distracting vs. infrequently distracting situations) actually results from more general learning processes for which the distracting vs. non-distracting nature of the event is irrelevant, e.g., people simply learn to produce the response that is typical for a certain event even when that event contains distracting information. In this research, I re-examined this conclusion by creating a series of experimental situations in which general learning processes were prevented from occurring, controlled for in the statistical analyses, or impaired by reducing the cognitive resources necessary to use them. In all of those situations, the crucial finding of reduced distraction in frequently distracting vs. infrequently distracting conditions emerged, suggesting that general learning processes are not the whole story in producing such findings. Instead, what these results suggest is that, consistent with

most theorizing in this research area, humans possess and use the ability to adapt attention so as to deal with distraction more effectively when distracting events occur more frequently.

Co-Authorship Statement

The data presented in this dissertation were obtained in collaboration with Dr. Stephen J. Lupker (all chapters), Dr. Jason R. Perry (Chapters 2 and 4), and Kesheni Krishna (Chapter 4). The written material in this dissertation is my own work. However, Dr. Stephen J. Lupker provided assistance in revision of the content.

Acknowledgments

I am grateful to each and every person who helped me throughout the completion of this thesis. First, I would like to thank my supervisor, Dr. Stephen Lupker, who has always been a reassuring presence during these years. He gave me something I deem invaluable: the opportunity to conduct research in the area of my choice. It is fair to say that this thesis would not have even been conceived without that opportunity. I equally appreciated the fact that he has always made sure of the solidity of this research by offering continuous, rigorous support for all of its aspects, from development to write-up. The thoughtful help I received for many non-academic matters is another reason for my gratitude to him.

I extend my gratitude to the members of my advisory committee, Dr. Bruce Morton and Dr. Derek Mitchell. Their comments challenged me to explore alternative interpretations for my results and helped me clarify my own interpretation. I would also like to thank my lab mates, Zian Chi, Lingling Li, Mark McPhedran, Alex Taikh, and Huilan Yang, for the patience with which they endured my often-contorted lab talks, for their co-operation in data collection, and for the great times spent together during our travels. Many thanks also to Angela Zhao for her zealous help in data collection, and to Jason Perry and Kesheni Krishna for the enthusiastic discussions we had during our collaboration. I have also received very valuable feedback on my research from many faculty members and graduate students in the Psychology Department, especially from Dr. Debra Jared's and Dr. Albert Katz's labs, to whom I also extend my thanks. I am also very thankful for the generous support of the Ontario Trillium Scholarship and the University of Western Ontario, without which this research would not have been possible.

Finally, I would like to thank my wife Cecilia, who has always been at my side during this journey. Not once did she fail to cheer me up when I felt overwhelmed with the challenges that a PhD and a life in a foreign country in general involve. My mother Federica, my father Marco, my sister Chiara, and my grandmother Renza never ceased to show me their love and support either, albeit from overseas. Thank you also to my in-laws, to Luciana and Leonardo, and all of my relatives and friends in Italy and abroad.

Table of Contents

List of Tables

Chapter 4

List of Figures

Chapter 2

Chapter 4

List of Appendices

List of Abbreviations

AATC = Associations as Antagonists to Top-Down Control hypothesis (Bugg, 2014a).

DMC = Dual-Mechanisms-of-Control account (Braver, 2012; Braver, Burgess, & Gray, 2007).

MC = Mostly Congruent condition.

ME = Mostly Easy condition.

MH = Mostly Hard condition.

MI = Mostly Incongruent condition.

MN = Mostly Neutral condition.

PC = Proportion-Congruent manipulation.

PE = Proportion-Easy manipulation.

PEP = Parallel Episodic Processing model (Schmidt, 2013c)

PN = Proportion-Neutral manipulation.

WM = Working Memory.

Chapter 1:

General Introduction

Proportion-Congruent effects: Conflict adaptation at multiple levels of control

A question of fundamental interest in cognitive psychology concerns what role control processes play in goal-oriented behavior. A primary role of these processes is to prevent conflict created by task-irrelevant information from disrupting that behavior. Although such conflict may produce a processing cost, a control mechanism must exist that resolves that conflict and allows for the selection of an appropriate response. For example, in the Stroop (1935) task, participants are required to name the ink color of a word which can be congruent with the word (e.g., the word RED in the color red col), incongruent with the word (e.g., the word BLUE in red), or the word (or letter string) can be color-neutral (e.g., the consonant string XXX in red). The typical result is that congruent items produce slightly faster latencies than neutral items (i.e., there is some facilitation) and incongruent items produce (much) slower latencies than neutral items (i.e., there is (larger) interference). The color-naming difference between incongruent and congruent items is the combination of these two effects and is known as the congruency effect (for a review, see MacLeod, 1991). What is worth noting is that even if incongruent words cause substantial interference, in most cases participants are able to correctly identify the colors that those words are presented in, suggesting that the control system eventually resolves the conflict that those words create.

A question that has received increasing research interest in recent years is whether, in addition to resolving conflict, the control system might have other functions that would help to deal with conflict. A popular model in the cognitive control literature, the conflict-monitoring model (Botvinick, Braver, Barch, Carter, & Cohen, 2001), provides such an example. Botvinick et al. (2001) proposed that humans possess a conflict-monitoring system that monitors for the presence of conflict in a stimulus and adapts attention between task-relevant and task-irrelevant dimensions of the stimulus accordingly. Specifically, when a conflict is detected (e.g., in an incongruent item in the Stroop task), a top-down signal is emitted indicating a need for more focused attention to task-relevant information (i.e., the color). This signal will not be emitted when little or no conflict is detected (e.g., in a congruent item), causing a relaxation of attention in that case. What is worth noting, however, is that, according to this theory, this mechanism does not just help to resolve the conflict experienced on any given trial, it also influences subsequent performance. For example, experiencing conflict on a trial will induce focused attention to task-relevant information on subsequent trials as well. With the system already prepared for conflict, the interference produced by incongruent task-irrelevant information on subsequent trials will be reduced. The implication is that the control system might not be limited to the resolution of conflict but may also have the function of adapting to conflict – a “conflict adaptation” function.

In the past few years, the idea of a conflict adaptation mechanism in cognitive control has spurred a wealth of research that has attempted to characterize this mechanism (see, e.g., Bugg & Crump, 2012). More recently, however, a growing concern has arisen that led some researchers, most notably Schmidt (2013b, 2019), to dispute the existence of such a mechanism and to propose alternative explanations, explanations in which conflict plays little or no role, for the evidence supporting the idea of conflict adaptation. The purpose of the present research was to provide a close examination of the arguments that Schmidt has put forward to make his case that conflict adaptation may not be a mechanism that humans use. The specific focus of this examination was the Proportion-Congruent (PC) paradigm, a paradigm that has had a central role in the debate on the existence of conflict adaptation. Given the widespread use of this paradigm in cognitive research (Bugg & Crump, 2012), an exact understanding of the processes being engaged in this paradigm is of fundamental importance. An exact understanding of the PC paradigm would inform not only the theory of cognitive control function but also the applications of that theory in other domains, for example, clinical and neuroscientific domains. In the following, I offer a brief overview of the PC paradigm, starting from the standard paradigm and following with its more recent variations.

In this initial section, I review the conflict-based explanations that have been offered for the findings typically reported in this paradigm, explanations that maintain that adaptation to conflict frequency occurs at multiple levels. In the following section, I review the non-conflict explanations of those findings, explanations that, unlike conflict-based explanations, assume no role for conflict adaptation. Finally, I introduce the approaches used in the present research to examine those non-conflict explanations.

<u>The standard (list-wide) Proportion-Congruent effect</u>

One of the most important pieces of evidence that has supported the idea of a conflict adaptation function of the control system comes from what is known as the PC paradigm. This paradigm, typically implemented in Stroop and Stroop-like tasks (e.g., the picture-word interference task: Lupker, 1979), consists of manipulating the frequency of congruent and incongruent items in a list so as to create a Mostly-Congruent (MC) list in which congruent items are frequent and incongruent items are infrequent, and a Mostly-Incongruent (MI) list in which incongruent items are frequent and congruent items are infrequent. The typical result is that the congruency effect (i.e., the color-naming difference between incongruent and congruent items) is larger in the MC list than in the MI list, a finding traditionally known as the PC effect and more recently referred to as the *list-wide PC effect* to distinguish it from other types of PC effects discovered later (see below; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979; Logan, Zbrodoff, & Williamson, 1984; Lowe & Mitterer, 1982).

The list-wide PC effect has traditionally been interpreted as the manifestation of a mechanism of adaptation to conflict frequency in the list, an interpretation that is easily accommodated within the conflict-monitoring model (Botvinick et al., 2001). According to this interpretation, frequent experience of conflict (on incongruent trials) in an MI list would induce tightened control, with overall more focused attention to task-relevant information in that list. Interference from irrelevant information will thus be minimized, resulting in a small congruency effect. In contrast, infrequent experience with conflict in an MC list would induce a relaxation of attention in that list, with the result being increased interference from task-irrelevant information on the few trials in that list in which that information produces a conflict (i.e., on the few incongruent items in that list), and hence, a large congruency effect.

Note that the conflict-monitoring model is not the only framework that has been used to explain the list-wide PC effect in the control literature (e.g., Braver, 2012; Braver, Gray, & Burgess, 2007; Kane & Engle, 2003; De Pisapia & Braver, 2006). What is common among these explanations, however, is the idea that the frequency of conflict encountered in a list plays an important role in directing attention appropriately between task-relevant and task-irrelevant dimensions of the stimuli. For example, Kane and Engle (2003) proposed that an MC list would favor goal neglect whereas an MI list would favor goal maintenance. This view is consistent with the idea that in an MC list, attention is relaxed because of the many congruent items for which a response can be made using task-irrelevant information (e.g., the identity of the word in the Stroop task), thus favoring neglect of the task goal (e.g., the color-naming goal). In contrast, in an MI list, the presence of many incongruent items requires attention to be more focused on task-relevant information (e.g., the ink color of the word), the information that is relevant to the task goal. As a result, that goal will be maintained more easily throughout an MI list. In sum, a mechanism of attentional adaptation to the frequency of conflict in the list appears to be central in most explanations of the list-wide PC effect that have been offered from a control perspective.

The item-specific Proportion-Congruent effect

In recent years, research in cognitive control has broadened the scope of conflict adaptation functions, suggesting that the control system might adjust to conflict at multiple levels. For example, Gratton et al. (1992) reported evidence suggesting that attention to task-relevant information might be enhanced not only in situations in which conflict is repeatedly experienced (such as in a MI list) but also in a more transient fashion, after a single experience with conflict. What Gratton et al. found was a larger congruency effect following a congruent trial than following an incongruent trial, a pattern that has become known as the congruency sequence effect (for a review, see Egner, 2007). (note 1) This phenomenon, along with the list-wide PC effect, has typically been explained in the context of the conflict-monitoring model

(Botvinick et al., 2001; but see, e.g., Mayr, Awh, & Laurey, 2003). According to this explanation, experiencing conflict during an incongruent trial would induce more focused attention to the task-relevant dimension and, hence, reduce the impact of conflict on the subsequent trial, with the result being a small congruency effect on such trials. Conversely, experiencing little or no conflict during a congruent trial would induce relaxed attention because there is less reason to tighten control. The result would thus be increased interference from the task-irrelevant dimension on the subsequent trial, and hence, a large congruency effect on such trials. (note 2)

More recently, however, a series of results has been reported that the conflict-monitoring model (Botvinick et al., 2001) appears unable to explain. The first result in this series is the finding, reported by Jacoby, Lindsay, and Hessels (2003), of an *item-specific PC effect*. Jacoby et al. designed a new version of the PC paradigm in which one set of color words (the MC items, e.g., GREEN and YELLOW) were presented mainly in their congruent color (e.g., the word GREEN appearing more often in green than in yellow) and another set of color words (the MI items, e.g., RED and BLUE) were presented mainly in an incongruent color (e.g., the word RED appearing more often in blue than in red). Similar to the list-wide PC effect, an item-specific PC effect emerged, with MC items eliciting a larger congruency effect than MI items.

Following Jacoby et al.'s (2003) article, a number of paradigms were developed in which congruency proportion was manipulated for a particular context feature, for example, the position of the word on the screen (Crump, Gong, & Milliken, 2006; Crump & Milliken, 2009), the font that the word is presented in (Bugg et al., 2008), or the ink color of the word itself (Bugg & Hutchison, 2013). In those cases as well, a PC effect specific to the particular context feature used for the PC manipulation was typically observed (for a review, see Bugg & Crump, 2012).

The reason that the conflict-monitoring model can explain the list-wide PC effect (and the congruency sequence effect) but cannot explain item-specific and context-specific PC effects has to do with the type of process assumed in that model for adaptation to conflict. In the conflict-monitoring model, adaptation to conflict occurs in a preparatory or *proactive* manner. That is, the implementation of a process of focusing vs. relaxing attention to task-relevant information based on experience with conflict is applied *before* any specific item appears. This proactive process, therefore, does not depend on any particular feature of the current stimulus. For example, a participant who encountered several incongruent items while performing an MI list should be more focused on task-relevant information in the current trial regardless of the specific stimulus that appears on that trial (e.g., RED in red vs. GREEN in yellow). The situation is different in item-specific and context-specific PC paradigms, however. Because in those paradigms congruent and incongruent items are equally probable in the list as a whole, whatever process produces the PC effect in those situations needs to be a *reactive* process, i.e., a process that is initiated *after* the item appears and that is based on the MC/MI nature of that item (or the context that the item appears in). For example, in Jacoby et al.'s (2003) paradigm, there is no way of knowing whether the upcoming stimulus is an MC or an MI item. The implementation of a process of relaxed attention (for an MC item) vs. more focused attention to task-relevant information (for an MI item) must be applied after the stimulus has appeared, in *reaction* to the nature of that stimulus.

Because the original conflict-monitoring model did not implement such an item-specific reactive conflict-adaptation process, this model cannot account for item-specific and context-specific PC effects. Indeed, Blais, Robidoux, Risko, and Besner (2007) failed to simulate the item-specific PC effect using Botvinick et al.'s (2001) conflict-monitoring model. On the other hand, they did manage to simulate that effect when the conflict-monitoring model was modified so that conflict adaptation could occur at the item level, with conflict arising for a certain item leading to increased attention to task-relevant information only for that item (as opposed to leading to increased attention to task-relevant information in general). As a result, in this modified model, associations between specific items and conflict frequency could be learned and used to recruit control settings appropriate to those items. For example, the recognition of an MC word, e.g., GREEN, would induce a relaxation of attention, resulting in large interference when the MC word does conflict with the color. In contrast, the recognition of an MI word, e.g., RED, would induce focused attention to the color, resulting in reduced interference for that word. Thus, although both the original conflict-monitoring model (Botvinick et al., 2001) and the modified version (Blais et al., 2007) implement a conflict adaptation mechanism, the level at which this mechanism operates is not the same in the two models: The former involves a mechanism based on the list composition that is applied to all items indiscriminately whereas the latter involves a more local mechanism that is applied to specific items in the list. (note 3)

<u>The list-wide Proportion-Congruent effect revisited: Dissociating proactive and reactive control</u>

As noted, the fact that in the item-specific PC effect the conflict adaptation process leading to relaxed/focused attention is initiated after the word is recognized makes that process a reactive process, as opposed to the proactive process that supposedly underlies the list-wide PC effect. However, recent research has recognized that a list-wide PC effect does not necessarily indicate that a proactive control process is the only process being used (Blais et al., 2007; Braver & De Pisapia, 2006; Bugg et al., 2008; Bugg, 2014a; Hutchison, 2011; Kane & Engle, 2003). The reasoning is that, first, a form of reactive control is likely implemented in list-wide PC paradigms, specifically in the MC list. Because that list favors neglect of the color naming goal (i.e., a relaxation of attention), that goal would need to reactively be retrieved upon detection of conflict when one of the infrequent incongruent items in that list appears, or else a word-reading error would be committed. Thus, the list-wide PC paradigm likely engages both proactive control (i.e., preparatory control that determines the overall focused vs. relaxed state of attention applied in MI vs. MC lists) and a form of reactive control exclusively applied in the MC list to help retrieve the color naming goal to deal with the unexpected conflict created by incongruent items in that list (Braver & De Pisapia, 2006; Kane & Engle, 2003).

More importantly, however, it is also the case that assuming a proactive process may not be necessary at all in order to explain the list-wide PC effect obtained in the typical list-wide paradigm. The reason is that, in the typical list-wide PC paradigm, each word being used appears more often in its congruent color in MC lists and more often in incongruent colors in MI lists. Thus, all items in an MC list are MC items and all items in an MI list are MI items. The implication is that a PC effect observed in this situation could be the result of a proactive conflict adaptation process leading to focused attention in the MI list and relaxed attention in the MC list (potentially combined with a reactive process that would help resolve conflict for incongruent items in the MC list) or it could entirely be the result of a reactive conflict adaptation process whereby the appropriate control setting is applied to individual items (e.g., focus attention to the color for MI items vs. relax attention for MC items), i.e., the same process that produces the item-specific PC effect in item-specific PC paradigms. (note 4)

In an effort to determine the role of proactive vs. reactive control in the list-wide PC effect, Bugg et al. (2008) introduced a new paradigm aimed to dissociate the two processes by constructing list-wide PC manipulations in which item-specific (i.e., reactive) conflict adaptation could be used for one set of items, referred to as the “context items”, but not for another set of items, referred to as the “transfer items” (see also Blais & Bunge, 2010; Bugg, 2014a; Hutchison, 2011). The transfer items (e.g., the words RED and BLUE and the corresponding colors) were 50:50 congruent/incongruent and were intermixed in a list with the context items (e.g., the words GREEN and YELLOW and the corresponding colors) that were either Mostly Congruent (creating an overall MC list) or Mostly Incongruent (creating an overall MI list). The rationale for this manipulation was that, while a PC effect obtained on the context items might result from item-specific, reactive control (leading to focused attention to color for MI context items in the MI list and relaxed attention for MC context items in the MC list), a PC effect on the transfer items, items that are identical in the two lists, could only be explained by a mechanism of list-wide, proactive control based on the frequency of conflict in the list. (note 5)

Indeed, in addition to a PC effect on the context items (an effect that is compatible with either reactive or proactive control), a (smaller) PC effect on the transfer items did emerge in some of those studies (Bugg, 2014a; Hutchison, 2011), although not in others (the magnitude of the congruency effect on the transfer items was the same in the MC list and the MI list in some situations: Blais & Bunge, 2010; Bugg et al., 2008). These results led those researchers to conclude that conflict adaptation may be engaged both reactively and proactively, although the latter form of control may only be applied in certain situations (as will be explained later).

Overall, the idea that conflict adaptation can be used both proactively and reactively is in agreement with a recently developed theory of cognitive control, the Dual Mechanisms of Control (DMC) framework (Braver, 2012; Braver, Gray, & Burgess, 2007; De Pisapia & Braver, 2006; see also Gonthier, Braver, & Bugg, 2016). According to this account, proactive and reactive control represent distinct modes of control: On the one hand, proactive control (similar to Kane and Engle's (2003) notion of goal maintenance) involves the sustained maintenance of task-relevant information and is preferentially engaged in situations that repeatedly reinforce task relevance (e.g., in an MI list in the list-wide PC paradigm). On the other hand, reactive control would be engaged in at least two types of situations: First, it would be engaged to retrieve the task goal upon detection of conflict in situations that rarely reinforce task relevance (e.g., when an incongruent item appears in an MC list in the list-wide PC paradigm), situations that may lead to neglect of that goal. Because this process is based on the congruency of the item but not on the identity of the item (e.g., the word RED vs. the word GREEN), it is a reactive process, but not an item-specific process. Second, a reactive control process would also be engaged in situations in which contextual information can be used to re-activate previously acquired information. These situations would include item-specific PC paradigms in which recognition of an item would lead to the recruitment of the appropriate control setting for that item, as well as any other situation in which MC or MI items appear, for example, standard list-wide PC paradigms (in which, as noted, all items are either MC [in the MC list] or MI [in the MI list]; e.g., Logan & Zbrodoff, 1979), or context items in more recent list-wide PC paradigms (in which the context items are MC items in the MC list and MI items in the MI list, whereas the intermixed transfer items are typically neither MC nor MI items; e.g., Bugg et al., 2008). Because this process is based on the MC/MI nature of the items, it is a reactive and item-specific process.

Although individuals may vary in the extent to which they have access to proactive control, a more effortful and resource-demanding control mode (see, e.g., Burgess & Braver, 2010), all individuals appear to have access to both proactive and reactive control, as demonstrated by the fact that, for example, the same individuals produce list-wide and item-specific PC effects (Gonthier et al., 2016). As the present discussion has indicated, however, these processes, according to the DMC account, take a number of forms and appear in different situations. Because the characterization of list-wide and item-specific PC effects in terms of proactive and reactive control, as presented above, has an important role in the considerations guiding the present research (especially the research presented in Chapter 4), a summary of this characterization is presented in Table 1.

Table 1.

The Processes Involved in List-Wide and Item-Specific PC Paradigms According to the DMC Account

Process	Description of the process	Conditions in which the process is engaged
Proactive	Sustained maintenance of task goal	- MI list (list-wide PC paradigm)
Reactive (non-item-specific)	Retrieval of task goal upon detection of conflict	- MC list (list-wide PC paradigm)
Reactive (item-specific)	Retrieval of the control setting most appropriate for the item (relaxed attention for MC items vs. more focused attention to task-relevant information for MI items)	- All items in standard list-wide PC paradigms (i.e., with no distinction between context and transfer items; e.g., Logan & Zbrodoff, 1979) - Context items in list-wide PC paradigms (e.g., Bugg et al., 2008) - Item-specific PC paradigm

Is conflict adaptation actually an illusion? Non-conflict learning can explain Proportion-Congruent effects

Despite the popularity of PC paradigms as measures of conflict adaptation processes, in the last decade, a growing body of research has cast doubt on conflict adaptation as a valid explanation for PC effects (as well as other effects traditionally thought to index conflict-induced control of attention, such as the congruency sequence effect: e.g., Schmidt, 2013b, 2019; Schmidt & Besner, 2008; Schmidt, Notebaert, & van den Bussche, 2015). The reason for this doubt is the realization that PC paradigms typically contain one or more confounds that allow for alternative interpretations of PC effects. Crucially, because these confounds are related to general learning abilities that are irrelevant to the conflicting vs. non-conflicting nature of the stimuli, the alternative explanations they afford are explanations in which conflict adaptation processes are essentially unnecessary. In the following sections, I review the challenges that non-conflict learning processes pose for control-based interpretations of PC effects, both in traditional paradigms and in more recent paradigms in which attempts were made to control for those confounds. Prior to discussing those proposed processes, a summary of the processes and the situations in PC paradigms in which those processes represent a confound is provided in Table 2.

Table 2.

Non-Conflict Learning Processes in PC paradigms

Process	Description of the process	Conditions in which the process is engaged and represents a confound
Contingency learning	Learning and using associations between a stimulus dimension (e.g., a word) and the response frequently made to that stimulus dimension	- All items in standard list-wide PC paradigms (i.e., with no distinction between context and transfer items; e.g., Logan & Zbrodoff, 1979) - Context items in list-wide PC paradigms (e.g., Bugg et al., 2008) - Two-item set item-specific PC paradigm (Jacoby et al., 2003)
Stimulus informativeness	Increasing attention to a stimulus dimension (e.g., a word) if that stimulus dimension allows contingency learning	- List-wide PC paradigms in which the MC list is more informative than the MI list (e.g., Bugg, 2014a – Expts. 1A and 2B) - Four-item set item-specific PC paradigm in which MC items are more informative than MI items (Bugg & Hutchison, 2013 – Expt. 3)
Temporal learning	Learning and using temporal expectancies for response emission based on the average difficulty of the stimuli in the list	- All list-wide PC paradigms

Contingency learning

One of the main confounds contained in PC paradigms is that there typically are associations, or contingencies, between a stimulus and a motor response (Lin & MacLeod, 2018; Schmidt, Crump, Cheesman, & Besner, 2007; Musen & Squire, 1993). Because the words used in PC paradigms tend to appear in some colors more often than in other colors (and, thus, they tend to require some responses more often than other responses), it is the case that participants in PC paradigms can learn those contingencies. Most importantly, this contingency learning process would produce a PC effect without the need to assume conflict adaptation processes (Schmidt & Besner, 2008).

For example, if in an item-specific PC paradigm the MI word GREEN appears more often in yellow (an incongruent, but high-contingency [i.e., most probable] color for that word) than in green (the congruent, but low-contingency [i.e., less probable] color for that word), participants will be able to use the word GREEN to predict a yellow response, thus speeding up responses to GREEN presented in yellow (causing them to be fast for an incongruent stimulus) while not necessarily affecting responses to GREEN presented in green. The congruency effect for this word will thus be somewhat small. Conversely, individuals will use the MC word RED to predict the (congruent) red response. Thus, latencies will speed up when the word and color are congruent, producing a large congruency effect. Similarly, in the typical list-wide PC paradigm, any word in the MC list would most frequently require a congruent response (e.g., the word RED would typically require the “red” response as it frequently occurs in the [congruent] red color). If participants learn these contingencies, responses to the (high-contingency) congruent colors will speed up whereas responses to (low-contingency) incongruent colors will not. As a result, the congruency effect in the MC list will increase. The same is not true for MI lists, a type of list in which a contingency learning process would, if anything, lead to a reduction of the congruency effect. This reduction would be observed, for example, if the MI list is constructed so that each word appears only in two colors, the infrequent congruent color and a frequent incongruent color, a situation in which use of contingency learning would speed up responses to the incongruent, but high contingency, color but may not affect responses to the congruent, but low contingency, color. (Note that, alternatively, the MI list could be constructed so that no contingencies can be learned, for example, when four words and four colors are used and each word appears equally often in each of the colors, one congruent and three incongruent. In that case, the congruency effect would not be modified by any contingency learning process in that list, but that effect would still be smaller than the effect in an MC list in which contingency learning would inflate the congruency effect). (note 6)

According to Schmidt (2013b, 2019), assuming that contingency learning, rather than adaptation to conflict frequency, is the main factor driving item-specific and list-wide PC effects, would have considerable advantages. Concerning the item-specific PC paradigm, first, it would explain why the item-specific PC effect is typically driven by both congruent and incongruent items rather than being mainly driven by incongruent items, the pattern that a mechanism of adaptation to conflict frequency would seem to imply (see Schmidt & Besner, 2008). That is, in the item-specific PC effect, the speed-up for congruent items in the MC vs. MI condition is often equivalent to the speed-up for incongruent items in the MI vs. MC condition. However, from a control perspective, because the congruency effect in the Stroop task is mainly due to interference rather than facilitation (MacLeod, 1991), one would expect that the PC manipulation would mainly affect the incongruent items (i.e., more focused attention to MI words should create a large speed-up for incongruent items in that condition because word interference is minimized) and would only minimally affect the congruent items (i.e., relaxed attention to MC words should make congruent items in that condition somewhat faster because words would be more easily processed, but not much faster). On the other hand, a contingency learning account would have an easier time explaining a symmetric pattern because it assumes that the same process (i.e., contingency learning) is applied to both congruent and incongruent items. Nonetheless, as Schmidt (2013b) notes, this line of reasoning is actually not decisive because more recent versions of the contingency learning account, versions which assume that contingency learning effects scale with response time, would also predict an asymmetric pattern.

Another, more critical, advantage of the assumption that contingency learning underlies the item-specific PC effect is that it would easily accommodate the results of studies in which item-specific conflict adaptation and word-response contingency learning processes were dissociated (Hazeltine & Mordkoff, 2014; Schmidt, 2013a). For example, Schmidt (2013a) constructed a Stroop task in which MC words, e.g., RED, and MI words, e.g., GREEN, could be compared on what were "contingency-matched" incongruent trials. For example, blue was a low-contingency and equally probable color for RED and GREEN. The existence of a conflict adaptation mechanism would imply that because MC words should induce relaxed attention whereas MI words should induce focused attention to the color, the MC word RED should produce more interference than the MI word GREEN when those words are presented in blue. Performance on MC and MI words, however, was equivalent, suggesting that no conflict adaptation process was in use. In contrast, a robust contingency learning effect that cannot be explained in terms of item-specific conflict frequency emerged in the comparison between high-contingency and low-contingency items (e.g., the MI word GREEN in the high-contingency color yellow was responded to faster than the word GREEN in the low-contingency color blue). Based on these results, Schmidt concluded that the item-specific PC effect may be fully explained by a contingency learning process (although see Bugg & Hutchison, 2013, reviewed in the next section).

A contingency learning account would also help explain part (although not the totality) of the list-wide PC effect. A critical piece of evidence in support of a role of contingency learning in the list-wide PC effect comes from the PC paradigm introduced by Bugg et al. (2008; see also Blais & Bunge, 2010; Bugg, 2014a; Hutchison, 2011). As noted above, Bugg et al. aimed to dissociate proactive and reactive control explanations of the list-wide PC effect by dividing the stimulus set into context items (e.g., GREEN and YELLOW and the corresponding colors) for which a reactive process of adaptation to item-specific conflict frequency could be used, and transfer items (e.g., RED and BLUE and the corresponding colors) for which this process could not be used, such that a PC effect for those items would uniquely reflect adaptation to the frequency of conflict in the list as a whole. Perhaps unwittingly, however, Bugg et al. also effectively controlled for the contingency learning confound that standard list-wide PC paradigms typically contain.

The reason is that, although contingencies could be learned for the context items and learning those contingencies would produce a PC effect for those items (e.g., learning that GREEN predicts the [congruent] green response in the MC list vs. the [incongruent] yellow response in the MI list would lead to a larger and a smaller congruency effect, respectively), the same would not be true for the transfer items because those items were identical in the two lists. That is, unlike for context items, it would be impossible for a contingency learning process to produce a PC effect for transfer items because it could not differentially modify the magnitude of the congruency effect for transfer items in the MC vs. MI list. Consistent with these ideas, in the original study reported by Bugg et al. (2008; see also Blais & Bunge, 2010), a PC effect was obtained on the (contingency-confounded) context items but not on the (contingency-unconfounded) transfer items. On the other hand, subsequent studies using a similar paradigm (e.g., Bugg, 2014a; Hutchison, 2011) did manage to produce a PC effect for transfer items, suggesting, as noted above, that either adaptation to list-wide conflict frequency is real (although it does not emerge in all situations) or, importantly for Schmidt's position, additional confounds exist that can produce PC effects, confounds that will be reviewed in the following sections.

Stimulus informativeness

Certain PC paradigms contain another confound strictly related to contingency learning, a confound that Schmidt (2014a; 2019) has termed "stimulus informativeness". This term refers to the degree to which words allow learning of contingencies. For example, in Jacoby et al.'s (2003) original item-specific PC paradigm, four words and colors were divided into two nonoverlapping sets, each consisting of the combination of two colors and two words. In this two-item set design, contingency learning is possible for both MC and MI words because a high-contingency color exists for both the former and the latter (e.g., the MC word GREEN typically occurs in the [congruent] green color while the MI word RED typically occurs in the [incongruent] blue color). Because both MC and MI words allow contingency learning, both words are informative (they can be used to predict the color response). The same would not be true in other designs, however.

As an example, in an attempt to dissociate contingency learning from adaptation to item-specific conflict frequency, Bugg and Hutchison (2013) developed a four-item set design in which eight words and colors were divided into two nonoverlapping sets, each consisting of the combination of four colors and four words. What this modification affords is that MI words can be constructed so that they do not allow contingency learning (e.g., when the MI word appears equally often in the four colors, one congruent and three incongruent). Note that MC words necessarily allow contingency learning in any design, an inevitable consequence of the fact that the congruent color is always the high-contingency color for an MC word. As a result, in this design, MC words are informative for participants but MI words are not.

Based on their results using both a two-item set design and their four-item set design, Bugg and Hutchison (2013) argued that only when MI words are uninformative would a process of adaptation to item-specific conflict frequency be used. Their crucial experimental manipulation (Experiment 3) was as follows. In the initial phase of the experiment, Bugg and Hutchison manipulated item-specific conflict frequency using both a two-item set design (in which both MC and MI words were informative) in one version of the experiment and a four-item set design (in which MC words were informative but MI words were not) in another version of the experiment. In both versions of the experiment, MI words produced a smaller congruency effect than MC words, i.e., both versions produced an item-specific PC effect. What is important is that, in the four-item set design but not in the two-item set design, the MI words also produced reduced interference in a new manipulation introduced in the final block of that experiment. In this final block, a new set of colors was used that had not been used before in the experiment, and both MI and MC words were presented in those incongruent colors. In the final block of the four-item set design, latencies for naming, for example, the incongruent color brown (a color used only in the final block of the experiment), were faster if that color appeared in an MI word than if it appeared in an MC word. In contrast, in the final block of the two-item set design, no differences were observed in naming the new incongruent colors appearing in MI and MC words.

To explain these results, Bugg and Hutchison (2013) suggested that the shorter latencies for naming the new incongruent colors in MI words in the final block of the four-item set design were due to the fact that, in that version of the experiment, participants had previously learned to focus attention to the color when those words were presented. This learning process, presumably, occurred as a result of having used a conflict adaptation process in dealing with those MI words earlier in the experiment, when item-specific conflict frequency was being manipulated. According to Bugg and Hutchison, the reason that this conflict adaptation process was engaged was that, in this version of the experiment, contingency learning was not a reliable process overall because contingencies could only be learned for half of the words in the experiment (i.e., the MC words). The situation was different in the two-item set design, however. Here, the failure to observe shorter latencies for naming the new incongruent colors in MI words in the final block of this version of the experiment would suggest that participants had *not* previously learned to focus attention to the color when those words were presented. The likely reason is that a contingency learning process, not a process of adaptation to item-specific conflict frequency, was the process that participants had used when dealing with those MI words earlier in the experiment, presumably because contingency learning was a reliable option for all of the words (i.e., both MC and MI words) rather than for just a portion of them (as was the case in the four-item set design).

Although Bugg and Hutchison's (2013) results do seem to provide evidence for a conflict adaptation process (at least in a limited set of circumstances), Schmidt (2014a; 2019) has offered an alternative account of those results based on the fact that in the four-item set design but not in the two-item set design, MC and MI words differed in informativeness. This explanation is based on the observation that informative stimuli attract attention (e.g., Jiang & Chun, 2001). What is possible, then, is that in the Stroop task, words that allow contingency learning receive more attention than do words for which no contingencies can be learned. As a result of receiving more attention, the former words, if incongruent with the color, would cause larger interference than the latter. The implication is that item-specific conflict adaptation would not be the only explanation for why, in the final block of the four-item set design in Bugg and Hutchison's (2013) experiment, MI words produced less interference than MC words when naming the new incongruent colors. Specifically, those MI words might have received less attention because they were relatively uninformative (i.e., they could not be used to predict the response), a mechanism of attention regulation that has nothing to do with conflict itself. Notably, however, this mechanism would not differentially impact the amount of attention that MC and MI words received in the two-item set design in Bugg and Hutchison's experiment because, in this design, both types of words were informative. Because both MC words and MI words, being informative, attracted attention, they would cause similar interference when presented in the new incongruent colors in the final block of Bugg and Hutchison's manipulation, thus explaining the failure to observe any difference between MC and MI words in that block.

A similar argument can be made for any list-wide PC paradigms in which words in the MI list are overall less informative than words in the MC list, paradigms in which a PC effect can be obtained even when contingency learning is controlled for. As noted, Bugg et al. (2008; see also Blais & Bunge, 2010) failed to obtain a PC effect on transfer items in their modified paradigm, suggesting that adaptation to list-wide conflict frequency may not exist. However, in follow-up work, Bugg (2014a) found that a PC effect on the transfer items (the items that are crucial for probing adaptation to list-wide conflict frequency) can be obtained if the nature of the context items (the items that, although not crucial for probing adaptation to list-wide conflict frequency, determine the congruency proportion of the list) is changed. Specifically, a PC effect for the transfer items emerged when the MI list was a list in which no contingencies could be learned for the MI context words because those words appeared in four equally probable colors (one congruent and three incongruent).

The situation created by Bugg (2014a) differed from those examined by Bugg et al. (2008) and Blais and Bunge (2010) in which the MI list was a list in which contingencies could be learned for MI context words because those words did have a more probable (incongruent) color. To explain these divergent patterns (a PC effect on the transfer items when MI context words do not allow contingency learning vs. no PC effect on the transfer items when MI context words allow contingency learning), Bugg (2014a) proposed that adaptation to list-wide conflict frequency will have a primary role only in situations in which contingencies cannot be used for most of the trials in the task, e.g., in an MI list that does not allow for contingency learning. In these situations, the process being used would be a process of adaptation to list-wide conflict frequency, leading to more focused attention to color information. On the other hand, when reliable contingencies exist in the MI list (Blais & Bunge, 2010; Bugg et al., 2008), contingency learning is the only process being engaged in both MC and MI lists (i.e., no conflict adaptation process is engaged in either list). As a result, the transfer items will be unaffected, causing them to produce the same size congruency effect in the two lists.

However, in this case as well, stimulus informativeness can offer an alternative account (Schmidt, 2014a, 2019). Because in Bugg's (2014a) experiments contingencies could be learned for at least some items (the context items) in the MC list but not for the same items in the MI list, the latter list was, overall, less informative than the former. That is, compared to the MC list, the words used in the MI list, overall, were not as reliable in allowing an accurate prediction of a specific color response. Because the MI list was relatively uninformative, participants could have reduced attention to all words in that list. As a result of receiving less attention, all words in that list, including the transfer words, would produce smaller interference, thus resulting in a reduced congruency effect. The implication is that the reduced congruency effect observed in the MI list for the transfer items could reflect the fact that participants reduced attention to words in that list because the words were relatively uninformative rather than because conflict was frequently experienced in that list. Notably, this account would also explain why no PC effect was obtained for the transfer items when contingencies could be learned for both MC and MI context items (Blais & Bunge, 2010; Bugg et al., 2008). The reason is that, in that situation, MC and MI lists were matched in stimulus informativeness (i.e., contingency learning could be applied to both MC and MI context words). Because the MC list and the MI list were relatively informative, attention would be directed to all words (including the transfer words) in both lists. As a result of receiving attention in both lists, the words would produce similar interference across lists, resulting in congruency effects of the same size (i.e., there should be no PC effect). (note 7)

Temporal learning

Up to this point in this discussion, it would appear that a combination of non-conflict learning processes can account for all the PC effects reviewed with no necessary involvement of conflict adaptation processes. However, at least one study exists in which both contingency learning and stimulus informativeness confounds were controlled for and yet a list-wide PC effect emerged. In a somewhat more complicated design than the ones reviewed thus far, Hutchison (2011) mixed a fixed set of transfer items with a variable set of context items so as to create three lists: An (informative) MC list in which all context items were congruent and allowed contingency learning, an informative MI list in which all context items were incongruent and allowed contingency learning (making contingency learning in that list as reliable as in the MC list), and a relatively uninformative MI list in which all context items were incongruent but did not allow contingency learning (making contingency learning less reliable overall than in the other two lists). Crucially, compared to the MC list, Hutchison found overall smaller congruency effects for the transfer items not only in the uninformative MI list but also in the informative MI list. Furthermore, the congruency effects for the transfer items were the same size in the informative MI list and in the relatively uninformative MI list, suggesting that differences in stimulus informativeness across lists played little role in Hutchison's experiment.

To explain Hutchison's (2011) list-wide PC effect, Schmidt (2013c) proposed that a temporal learning process could explain the effect without invoking a process of adaptation to list-wide conflict frequency. Central in this explanation is the fact that MC and MI lists inevitably differ in response rhythm, as this rhythm is, by necessity, faster when most of the trials elicit a fast response (in the MC list, in which most trials are congruent) and slower when most of the trials elicit a slow response (in the MI list, in which most trials are incongruent). The reason that response rhythm is of potential relevance for the PC effect is that participants in speeded tasks are known to use information about their response time to form temporal expectancies for response emission for the upcoming trials (Lupker, Brown, & Colombo, 1997; Lupker, Kinoshita, Coltheart, & Taylor, 2003).

Based on these ideas, Schmidt (2013c) proposed the existence of a temporal learning process that would produce a list-wide PC effect even in situations in which other confounds typically contained in list-wide PC paradigms (i.e., contingency learning and stimulus informativeness) are controlled for. According to this account, a faster temporal expectancy would cause the difference between easy-to-process stimuli (e.g., congruent items) and hard-to-process stimuli (e.g., incongruent items) to increase because easy stimuli, but not hard stimuli, will speed up because they can be processed fast enough to meet a fast temporal expectancy. Thus, in the case of an MC list (i.e., a situation in which the temporal expectancy is relatively fast), the result would be a large congruency effect. Conversely, a slower temporal expectancy would cause the difference between easy and hard stimuli to decrease because hard stimuli, in theory, may also speed up because they can be processed fast enough to meet the slower temporal expectancy in that situation (although in practice, as will be explained in more detail in Chapter 2, hard stimuli appear to be relatively insensitive to temporal expectancies, at least in many situations; Kinoshita & Mozer, 2006). As a result, an MI list in which the temporal expectancy is relatively slow would produce, if anything, a reduced congruency effect. In sum, a list-wide PC effect could be produced by a temporal learning process rather than by adaptation to conflict frequency, even in situations in which contingency learning and stimulus informativeness confounds are controlled for. (note 8)

To demonstrate that temporal expectancies could explain PC effects obtained in confound-minimized situations, Schmidt (2013c) re-analyzed Hutchison's (2011) data using linear mixed-effects modeling, a type of analysis that, unlike traditional mean-based ANOVAs, allows the evaluation of trial-level predictors. Specifically, in his re-analysis, Schmidt included a trial-level predictor functioning as an index of temporal expectancy, the latency on the most recent trial (i.e., RT on trial $n - 1$), in addition to the typical predictors in a PC paradigm (i.e., list type [MC vs. MI] and congruency [congruent vs. incongruent]). The reason for using that trial-level predictor in the re-analysis was that, because congruent items (i.e., relatively easy stimuli) are more likely to benefit from fast temporal expectancies (i.e., following a fast RT) whereas incongruent items (i.e., relatively hard stimuli) are more likely to benefit, if anything, from slower temporal expectancies (i.e., following a slow RT), the congruency effect on a given trial should be larger following faster responses than following slower responses, an interaction effect that Schmidt did obtain. What makes this temporal learning interaction compelling is that, by necessity, faster responses are more common in an MC list than in an MI list. As a result, the fact that congruency effects increase following faster responses would tend to inflate, overall, the congruency effect in the MC list and reduce it in the MI list, resulting in a PC effect.

Recently, however, Cohen-Shikora, Suh, and Bugg (2018) clearly demonstrated that Schmidt's (2013c) results were likely produced by the nonlinear transformation that he applied to the RT data. The reason that transformations of this sort are applied is that they do a decent job of accommodating the assumption made by linear mixed-effects models that the dependent variable be normally distributed (an assumption that the positively skewed distribution of raw RTs typically fails to satisfy). However, nonlinear transformations of the dependent variable have the downside of systematically altering the pattern and size of interaction terms, making analyses of interactions unreliable overall (Balota, Aschenbrenner, & Yap, 2013). Indeed, Cohen-Shikora et al. re-analyzed a number of datasets (including Hutchison's, 2011) and were unable to replicate Schmidt's (2013c) temporal learning interaction when untransformed, rather than transformed, RT data were used in a type of mixed-effects model that tolerates deviations from normality in the dependent variable (a generalized linear mixed-effects model: Lo & Andrews, 2015). Several additional attempts to evaluate the impact of temporal learning by Cohen-Shikora et al. also yielded no convincing evidence that temporal learning contributes to the PC effect to any extent.

On the other hand, another line of research exists that appears to support a potential role of temporal learning in producing effects like the list-wide PC effect without suffering from limitations in the analyses. This line of research is based on the idea that temporal learning should not be specific to interference tasks such as the Stroop task, i.e., tasks where conflict from an irrelevant dimension produces interference. Instead, any task in which the proportion of easy and hard items is manipulated should produce differences in the magnitude of difficulty effects that parallel those observed with congruency effects in the list-wide PC paradigm, i.e., a smaller difficulty effect in a list in which most of the items are hard and a larger difficulty effect in a list in which most of the items are easy. Indeed, Schmidt obtained evidence of this Proportion-Easy effect in a number of studies where no interfering irrelevant information was presented (Schmidt, 2013c; 2014b, 2016). For example, in a letter identification task, Schmidt (2013c) found faster identification for high-contrast letters (easy items) than for low-contrast letters (hard items). More importantly, however, the size of this difficulty effect was modulated by the proportion of easy items in the list, with larger difficulty effects in a list where most of the items were easy than in a list in which most of the items were hard, as predicted by temporal learning. Although, as noted by Schmidt (2013c), the temporal learning process producing the Proportion-Easy effect in this paradigm is not necessarily the same process that produces the PC effect in list-wide PC paradigms, it is possible that a temporal learning process of that sort does contribute to the emergence of list-wide PC effects.

The present research

Taken together, the non-conflict learning confounds noted by Schmidt (2013b, 2019) for PC paradigms suggest that the PC effects those paradigms typically produce do not necessarily represent humans' ability to learn "how" to deal with conflict, i.e., engaging the control strategy that is best suited for frequently conflicting situations (more focused attention to relevant information) vs. infrequently conflicting situations (relaxed attention). Instead, they could represent humans' ability to learn "what" to respond (i.e., selecting the response that is most likely for the stimulus as well as adapting to the informativeness of individual stimuli or the informativeness of the list as a whole), and/or "when" to respond (i.e., determining the point in time at which a response should be emitted). Notably, Schmidt (2013b) also argues that because these abilities, not being tied to conflict, reflect relatively general processes, the assumption that PC effects are fully explained by one or a combination of these non-conflict processes would constitute a more parsimonious explanation than an explanation that concedes a role for conflict adaptation. However, what may be deemed to be the most parsimonious explanation is not necessarily the correct one, as adaptation to conflict frequency might still be observed when the various non-conflict learning confounds affecting PC paradigms are accounted for. In fact, since Schmidt and Besner's (2008) seminal article, a number of researchers have attempted to modify the traditional list-wide and item-specific PC paradigms so as to obtain evidence of adaptation to conflict frequency that would not be contaminated by non-conflict learning processes (Braem et al., in press).

In this vein, the research reported here was an attempt to examine whether list-wide and item-specific PC effects would emerge in situations in which potential non-conflict learning confounds were accounted for. While some attempts undertaken in this direction suggest that such might be the case, at least in some situations (e.g., Bugg, 2014a; Bugg & Hutchison, 2013; Hutchison, 2011), those attempts often failed to consider non-conflict learning confounds in their entirety (Schmidt, 2013c, 2014a). The unique contribution of the present research was that all of the non-conflict learning confounds indicated by Schmidt (i.e., contingency learning, stimulus informativeness, and temporal learning) were taken into account. In addition, rather than pursuing this research from a single perspective/paradigm, a range of approaches was used with the purpose of gathering converging evidence in support of a conflict adaptation function of the control system. Of importance, as noted, this research would not only contribute to the theoretical debate concerning the existence of processes of adaptation to conflict frequency (Schmidt et al., 2015) but it would also inform those who apply theories of cognitive control in another domains. For example, in neuroscience studies, it is often assumed that PC paradigms elicit control adjustments, and, as a result, patterns of brain activity are typically interpreted in those terms (e.g., Braver & De Pisapia, 2006; Marini et al., 2016; but see Grandjean et al., 2013). Similarly, and perhaps more concerning, researchers who administer PC paradigms to aging and clinical populations tend to interpret abnormal results as deficits in cognitive control functions (e.g., Abrahamse et al., 2016; Bonnin, Houeto, Gil, & Bouquet, 2010; Bugg, 2014b). These interpretations, however, would be inaccurate if it turned out that PC paradigms do not actually measure conflict-induced control of attention but, rather, more general non-conflict learning processes. Because in the present research the idea that adaptation to conflict frequency is engaged in PC paradigms was put to a stringent test, this research could either consolidate that idea or suggest a serious reconsideration of it, with either outcome having a considerable impact on theories of control functioning and the applications of those theories.

Two lines of research were pursued. The first line of research was an examination of the list-wide PC effect. Although the list-wide PC effect has had a long tradition of research since it was first reported (Logan & Zbrodoff, 1979), a clear demonstration that this effect reflects a process of adaptation to list-wide conflict frequency, the process that most control accounts assume to explain it (e.g., Botvinick et al., 2001; Braver & De Pisapia, 2006; Kane & Engle, 2003), is still lacking. The reason is that, as noted, several non-conflict learning confounds typically exist in the list-wide PC paradigm (i.e., contingency learning, stimulus informativeness, and temporal learning), which, individually or in combination, could produce a PC effect without the necessary involvement of adaptation to list-wide conflict frequency. Indeed, to my knowledge, no single study currently exists that controls for all the non-conflict learning confounds indicated by Schmidt (2013b, 2019) for the list-wide PC effect. The research reported in Chapters 2 and 3 was an attempt to address this issue.

In Chapter 2, the question of whether a list-wide PC effect might still be observed when all non-conflict learning confounds are controlled for was addressed by using a list-wide PC paradigm in the picture-word interference task, a task in which participants are required to name or categorize a picture while ignoring a word superimposed on it. This task has been argued to be functionally equivalent to the color-word Stroop task (e.g., Lupker, 1979). However, unlike the color-word Stroop task in which only a few color targets and word distractors can be used, the picture-word interference task affords the opportunity of using many picture targets and word distractors. This characteristic of the picture-word interference task was exploited to overcome the problem that, in the list-wide PC paradigm in the color-word Stroop task, contingency learning is inevitable in the MC list (because at least some words in that list will need to appear more often with their congruent color, which will necessarily be the high-contingency color).

To this end, in two experiments, a situation was created in which no individual picture or word was repeated, making it impossible for participants to learn any word-response contingency in either the MC list or the MI list (with the additional result that in both lists, the words were also completely uninformative based on Schmidt's (2014b, 2019) definition of stimulus informativeness). In addition, to control for temporal learning, the data from those experiments were analyzed using a generalized linear mixed-effects model, a model that allows the evaluation of trial-level predictors of temporal expectancies without requiring transformations of the RT data, making interaction terms clearly interpretable (Cohen-Shikora et al., 2018).

To anticipate the results, a list-wide PC effect was obtained in both experiments, with no evidence that a temporal learning process contributed to the emergence of this effect. The idea that temporal learning could produce a list-wide PC effect in a picture naming task was further examined in an additional experiment in which, similar to Schmidt's (2013c, 2014b, 2016) Proportion-Easy experiments, the frequency with which easy-to-name vs. difficult-to-name pictures appeared in a list was manipulated but the naming difficulty did not derive from an irrelevant stimulus dimension (i.e., there were no word distractors). In this case as well, no evidence was found in support of the temporal learning process proposed by Schmidt.

In Chapter 3, the more traditional color-word Stroop task was used. Here, in order to overcome the problem that an MC list would always allow contingency learning without the possibility of extending the stimulus set substantially (as only a small number of easily nameable colors exists), a different approach was used: manipulating the proportion of neutral and incongruent items in a list (e.g., Tzelgov, Henik, & Berger, 1992). This approach maintains a conflict frequency manipulation (the key aspect of PC paradigms), while also having the advantage of potentially avoiding contingency learning and stimulus informativeness confounds because a Mostly Neutral list (i.e., a list in which there are more neutral than incongruent items), like a Mostly Incongruent list but unlike a Mostly Congruent list, can be constructed so that no individual word can be used to predict the color response. Using this characteristic that neutral items afford, a list-wide Proportion-Neutral manipulation was created in which contingency learning was impossible in both a Mostly Neutral list and a Mostly Incongruent list (again, making all words in the task uninformative). Even though contingency learning and stimulus informativeness were controlled for, a list-wide Proportion-Neutral effect, similar to the list-wide PC effect in the standard paradigm, was obtained, with a larger interference effect (i.e., the latency difference between incongruent and neutral items) in the Mostly Neutral list than in the Mostly Incongruent list. Furthermore, an analysis of the data using the procedure employed in Chapter 2 confirmed that temporal learning had no impact on this Proportion-Neutral effect.

Note that the picture-word interference experiments in Chapter 2 and the Stroop experiment in Chapter 3 were constructed in a way that allowed not only an examination of whether adaptation to conflict frequency exists, but also a determination of the proactive/reactive nature of this conflict adaptation process. As noted, control accounts of the list-wide PC effect such as the DMC account (e.g., Braver & De Pisapia, 2006; see also Kane & Engle, 2003) propose that this effect results from a proactive process of task goal maintenance that is continuously engaged in lists in which conflict is frequent (MI lists) and a reactive process of task goal retrieval that is occasionally engaged upon detection of conflict in lists in which conflict is infrequent (e.g., MC and MN lists). However, as demonstrated by Blais et al. (2007), the list-wide PC effect in traditional paradigms could also be entirely explained by a single, item-specific reactive process whereby the presentation of a frequently conflicting item (i.e., any item appearing in traditional MC lists) leads to more focused attention to task-relevant information than does the presentation of an infrequently conflicting item (i.e., any item appearing in traditional MI lists). The situation was different in the conflict-frequency manipulations reported in Chapters 2 and 3, however. The reason is that, in Chapter 2, the use of nonrepeated targets and distractors effectively prevented the use of any item-specific process. In Chapter 3, while an item-specific process (specifically, adaptation based on color-specific congruency proportion: Bugg & Hutchison, 2013) was possible, this was true for a portion of the items (the context items) but not for the other items (the transfer items). The fact that in both Chapters 2 and 3 a Proportion-Congruent/Proportion-Neutral effect was obtained even in situations in which item-specific reactive processes were impossible suggests that, in line with the DMC account of the list-wide PC effect, those effects were caused by a combination of proactive control (in MI lists) and (non-item-specific) reactive control (in MC and MN lists), rather than by a single item-specific reactive mechanism.

In Chapter 4, the focus of the research shifted to that putative item-specific reactive mechanism, a mechanism that, in theory, may underlie the item-specific PC effect (Jacoby et al., 2003). However, although the number of non-conflict learning confounds individuated in the item-specific PC paradigm is lower than in the list-wide paradigm (because, e.g., temporal learning would not be a confound in the item-specific paradigm: Schmidt, 2013b, 2014a), most researchers now agree that the item-specific PC effect is completely produced by those confounds (particularly, contingency learning), at least in Jacoby et al.'s (2003) two-item set paradigm in which a high-contingency color exists for both MC and MI items (Atalay & Misirlisoy, 2012; Bugg & Hutchison, 2013; Hazeltine & Mordkoff, 2014; Schmidt, 2013a).

In Chapter 4, this contingency learning interpretation of the item-specific PC effect in the original two-item set design was re-examined by having participants perform a Working-Memory task simultaneously with Stroop and non-conflict versions of a color identification task. The rationale for using this dual-task manipulation was that a concurrent working memory load is known to interfere with people's ability to learn contingencies in a non-conflict color identification task (a task in which color-unrelated words, rather than color names, are used: Schmidt, De Houwer, & Besner, 2010). If increasing working memory load reduces contingency learning in a non-conflict color identification task, it should also reduce the item-specific PC effect in the Stroop task if contingency learning is indeed the critical process underlying that effect.

Across three experiments, however, no evidence was found that increasing working memory load reduced the item-specific PC effect in the Stroop task even though, replicating previous research (Schmidt et al., 2010), working memory load did reduce contingency learning effects in the non-conflict color identification task. The implication of these results, also supported (albeit only partially) by an individual-differences analysis of participants' working memory capacity in the final experiment, is that the conclusion that contingency learning fully explains the item-specific PC effect in Jacoby et al.'s (2003) two-item set design is likely incorrect, and a role for adaptation to item-specific conflict frequency in this effect should be acknowledged. In fact, the argument is made in Chapter 4 that the overall pattern of results is better explained by the DMC account, a control account that assumes that the item-specific PC effect results from an item-specific reactive control process whereby attention to task-relevant information is increased upon presentation of MI items but relaxed upon presentation of MC items (Gonthier et al., 2016). (note 9)

Interim summaries linking Chapters 2 and 3 (Chapter 2.5) and Chapters 3 and 4 (Chapter 3.5) follow the relevant chapters. Finally in Chapter 5, all of the findings reported in this dissertation are summarized and discussed within the framework of the DMC account (Braver, 2012; Braver et al., 2007), an account that, as noted, has proven useful in interpreting list-wide and item-specific forms of adaptation to conflict frequency. Furthermore, the case is made that although attention control at multiple levels appears to be real, future research would certainly benefit from employing paradigms, like the ones reported here, that effectively prevent non-conflict learning confounds from contaminating putative measures of conflict-induced attention control.

Footnotes

1. In the literature, the congruency sequence effect is also known as the “conflict adaptation” effect. However, in the present discussion, the term “conflict adaptation” will be used in a more general sense. Specifically, “conflict adaptation” will be used to encompass all forms of conflict-induced adaptive control, including not only adaptation to recent conflict (i.e., the congruency sequence effect) but also adaptation to the frequency with which conflict occurs either in a list as whole (i.e., the list-wide PC effect) or in specific contexts (e.g., the item-specific PC effect). For a similar use of this term, see Schmidt (2013b).
2. It is worth noting that, in theory, the conflict adaptation process underlying the list-wide PC effect could be the very same trial-by-trial conflict adaptation process underlying the congruency sequence effect. That is, the finding that an MI list elicits a smaller congruency effect than an MC list might simply result from the fact that in the former list, it is much more common for a given trial to be preceded by an incongruent item (an item that induces focused attention and, thus, a smaller congruency effect on the subsequent trial) than by an congruent item (an item that induces relaxed attention and, thus, would produce a larger congruency effect on the subsequent trial). Thus, the overall small and large congruency effects obtained in MI and MC lists, respectively, would be the product of many micro-adjustments in attentional control occurring on a trial-by-trial fashion, as opposed to being the result of a list-wide process (i.e., focus attention in the MI list vs. relax attention in the MC list). Recently, however, dissociations between list-wide PC effects and congruency sequence effects have been

reported, such that the former are obtained even in situations in which the latter could not possibly contribute to producing a PC effect (e.g., Torres-Quesada, Funes, & Lupiáñez, 2013; Torres-Quesada, Milliken, Lupiáñez, & Funes, 2014). Thus, although sequential modulations of congruency effects likely contribute to the list-wide PC effect in most situations, this effect seems separable from the congruency sequence effect and may represent a distinct type of conflict adaptation.

3. It could be argued, however, that the conflict-monitoring component of Botvinick et al.'s (2001) model actually does imply a reactive process, although not the item-specific reactive process proposed by Blais et al. (2007). This process is the process that allows one to detect conflict upon presentation of a conflicting stimulus, a process that is functioning even when attention is relaxed. Indeed, as explained in the following section, subsequent control accounts do assume a role for reactive control in the list-wide PC paradigm, particularly in lists in which conflict is infrequent and a reactive mechanism must exist to help to deal with that unexpected conflict. What is important to note at this point, however, is that the original conflict-monitoring model did not implement any (reactive) process that would allow it to treat MC items/contexts differently than MI items/contexts.
4. The idea that the reactive process producing the item-specific PC effect may explain the list-wide PC effect in the list-wide PC paradigm may appear somewhat similar to the idea, reviewed in the next sections, that non-conflict learning processes may explain PC effects (Schmidt, 2013b, 2019). However, both an account that assumes that reactive, item-specific control produces the list-wide PC effect (Blais et al., 2007) and the more

traditional account that assumes that proactive, list-wide control produces that effect (e.g., Kane & Engle, 2003), are essentially conflict adaptation accounts. Thus, where those accounts differ is the level at which that adaptation occurs (the item level vs. the list level), not the nature of the cognitive process involved (e.g., a conflict-based process vs. a non-conflict-based process, the distinction that is relevant for the non-conflict learning accounts of PC effects which will be described subsequently).

5. Here and in the following, for simplicity, proactive adaptation to list-wide conflict frequency will be considered as an explanation for list-wide PC effects obtained in the absence of item-specific conflict adaptation processes even though a reactive process of retrieving the task goal upon presentation of incongruent items in the MC list likely complements the proactive process.
6. In the typical contingency learning explanation, it is assumed that the contingency learning process is merely facilitative (there is a benefit for high-contingency items but no cost for low-contingency items), in line with the initial proposal made by Schmidt and Besner (2008; see also Schmidt, 2013a). More recently, however, Lin and MacLeod (2018) demonstrated that in a non-conflict color identification task (a task in which color-unrelated words, rather than color names, are used), both a benefit for high-contingency items and a cost for low-contingency items are observed when those items are compared to a neutral baseline (a word for which no contingencies can be learned). Thus, it is certainly possible that in PC paradigms, high-contingency items are responded to faster than they would normally be (because there is a benefit from correctly predicting the response to those items) but also low-contingency items are responded

to slower than they would normally be (because there is a cost for those items due to the fact that the [incorrect] predicted response needs to be suppressed in favor of the correct one). Note, however, that although the expected pattern of results for PC effects would not be identical based on benefit-only vs. benefit-and-cost versions of the contingency learning account, both versions would predict a larger congruency effect in MC conditions than in MI conditions, i.e., a PC effect.

7. It is worth noting that the role that Schmidt's (2014, 2019) notion of stimulus informativeness would play in list-wide PC paradigms is similar to the role played by color-word correlations in Algom and collaborators' account of the Stroop effect (Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003; Sabri, Melara, & Algom, 2001). What those researchers proposed is that, in the Stroop task, attention to the (task-irrelevant) words is increased when there is a relationship between the words and the (task-relevant) colors. That is, when the words in the Stroop task provide information about the colors that they appear (or do not appear) in, those words will receive more attention than in a situation in which the words and the colors are randomly paired (a zero-correlation situation). For the reasons noted above, the words used in an MC list are inevitably strongly correlated with colors, whereas in an MI list, the color-word correlation will typically be lower if the list does not allow contingency learning. Thus, a comparison between an MC list with a strong color-word correlation and an MI list with a weaker color-word correlation would create the same concerns as voiced by Schmidt (2014b; 2019): Words would receive more attention in the MC list than in that type of MI list, however, the reason that attention is withdrawn from the words in the MI list

would not necessarily be the frequent experience with conflict in that list (as a control-based account would maintain) but, alternatively, the fact that the words in that list provide relatively little information about the colors that they appear in.

8. Note that because temporal learning appears to be relatively insensitive to differences in difficulty between specific items, this process is not considered as a potential explanation for the item-specific PC effect (Schmidt, 2013b, 2014a).
9. Chapters 2, 3, and 4 were submitted to peer-reviewed journals and, at the time that this dissertation was written, they were either accepted for publication (Chapter 2: Spinelli et al., 2019) or had received a revise and resubmit invitation following a second round of revisions (Chapters 3 and 4). The most recent version of each manuscript was reported with no substantial modification.

Chapter 2:

Adaptation to Conflict Frequency Without Contingency and Temporal Learning: Evidence from the Picture-Word Interference Task

Introduction

An established fact in cognitive research is that goal-oriented behavior requires some form of control for selection of appropriate responses in the face of conflict coming from task irrelevant information. What is less established, however, is whether control can be adaptively modulated in response to experience with conflict. With such a conflict adaptation mechanism, the cognitive control system would, presumably, not just resolve conflict, but also monitor conflict and adapt attention to relevant and irrelevant information accordingly (Botvinick, Braver, Barch, Carter, & Cohen, 2001).

Manipulations of conflict frequency in interference tasks such as the Stroop (1935) task typically produce a pattern of results that is consistent with a conflict adaptation explanation. In the classic (color-word) Stroop task, participants are required to name the ink color of a word while ignoring the word itself. A congruency effect typically arises, with faster (and often more accurate) responding to congruent items (e.g., the word RED in red color, RED_{red}) than to incongruent items (e.g., the word RED in blue color, RED_{blue}) (MacLeod, 1991). Of interest here is the fact that the magnitude of this effect varies as a function of experience with conflict. Specifically, situations in which the proportion of congruent items is high (i.e., infrequent conflict) elicit larger congruency effects than do situations in which the proportion of congruent items is low (i.e., frequent conflict) (e.g., Crump, Gong, & Milliken, 2006; Jacoby, Lindsay, &

Hessels, 2003; Logan & Zbrodoff, 1979; for a review, see Bugg & Crump, 2012). These Proportion-Congruent (PC) effects are readily explained by a conflict adaptation process. When conflict is frequent, there is regular demand for the control system to maintain attention focused on the relevant dimension. Interference from the irrelevant dimension will thus be minimized, resulting in a reduced congruency effect. On the other hand, when conflict is infrequent, the benefit of focusing on the color is rarely reinforced. As a result, interference from the irrelevant dimension on the few incongruent items that are present results in a large congruency effect.

Recent years, however, have witnessed a growing concern among researchers about the validity of conflict adaptation as an explanation for PC effects (Schmidt, 2013b; Schmidt, Notebaert, & van den Bussche, 2015). Such concern has its roots in the realization that, in speeded tasks, responding might be influenced by learned associations, or contingencies, between a stimulus and a motor response (Schmidt, Crump, Cheesman, & Besner, 2007; Musen & Squire, 1993), as well as by the formation of temporal expectancies for the emission of a response (Schmidt, 2013c). The reason these issues are relevant is that PC manipulations are typically confounded with contingency learning biases as well as with temporal learning biases (Schmidt, 2013c; Schmidt & Besner, 2008). As such, some combination of these factors when applied to the mechanisms involved in interference tasks appears to be able to explain the PC effects that are observed in those tasks without needing to posit a role for conflict (Kinoshita, Mozer, & Forster, 2011; Levin & Tzelgov, 2016). What is worth noting at this point is that, as will be described subsequently, these alternative accounts are essentially “facilitation” accounts. That is, their explanations for PC effects are based on the idea that some aspect of processing is facilitated as a result of participants gaining relevant information about the nature of the task. Hence, these accounts offer a radically different view of PC effects than that offered by the conflict adaptation account, which is based on an interference-driven mechanism.

Contingency learning

Contingency learning involves acquiring knowledge that two events tend to occur together (e.g., the presentation of the word RED typically requires the response “green”) and using that knowledge to facilitate responding (Beckers, De Houwer, & Matute, 2007). In color-word identification tasks in which the words used are not color names, contingency learning is presumed to explain why 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) (Schmidt et al., 2007; see also Musen & Squire, 1993). Essentially, according to contingency learning accounts, participants implicitly learn contingencies between words and color responses, i.e., that specific words predict specific color responses (e.g., BRAG predicts green; Schmidt et al., 2007; see also Forrin & MacLeod, 2017; Lin & MacLeod, 2018), allowing them to respond more rapidly when the word appears in its most frequent color.

The reason this issue is relevant for PC effects is that manipulating the proportion of congruent items in the Stroop task typically involves altering the frequency of specific word-color pairs as well. For example, PC experiments might involve a Mostly Congruent (MC) list in which the word RED appears in its congruent, red color 75% of the time and in the incongruent, blue color 25% of the time, and a Mostly Incongruent (MI) list in which the word RED appears in the incongruent, blue color 75% of the time and in its congruent, red color 25% of the time. Doing so, however, means that RED_{red} is more frequent than RED_{blue} in the MC list, whereas RED_{blue} is more frequent than RED_{red} in the MI list. If frequent word-color pairs elicit faster responses, participants might thus speed up on RED_{red} in the MC list but not on RED_{blue} in the MI list. Crucially, fast responding to the congruent item RED_{red} in the MC list will lead to a relatively large congruency effect whereas fast responding to the incongruent item RED_{blue} in the MI list will lead to a relatively small congruency effect. In other words, learning of word-color contingencies, rather than adaptation to conflict frequency, might be responsible for the difference in magnitude of congruency effects that is typically found in PC manipulations in the Stroop task (Schmidt & Besner, 2008).

Importantly, the assumption that contingency learning is the only source of PC effects (Schmidt & Besner, 2008) implies that no PC effects should be observed in PC manipulations that control for contingency learning operations. In an effort to address this issue, a number of studies have been conducted that evaluate PC effects on contingency-controlled stimuli, that is, stimuli which are matched in contingency across MC and MI lists (Blais & Bunge, 2010; Bugg, 2014a; Bugg & Chanani, 2011; Bugg, Jacoby, & Toth, 2008; Gonthier, Braver, & Bugg, 2016; Hutchison, 2011). The rationale is that if PC effects are driven by a mechanism of adaptation to list-wide conflict frequency, that mechanism should produce a PC effect for all stimuli, including the contingency-controlled stimuli. Results from those studies do provide at least partial support for this prediction, with PC effects on contingency-controlled stimuli being reported in a number of circumstances (Bugg, 2014a; Bugg & Chanani, 2011; Gonthier et al., 2016; Hutchison, 2011), although not in all circumstances (Blais & Bunge, 2010; Bugg et al., 2008). Therefore, contingency learning by itself does not appear to offer a complete explanation of PC effects, allowing proponents of the conflict adaptation account to argue that these contingency-controlled PC effects, when obtained, likely reflect the action of a mechanism of adaptation to conflict frequency (Bugg, 2014a). In contrast, Schmidt (2013c) has contended that those effects are better explained by a different, non-conflict learning process – temporal learning.

Temporal learning

While contingency learning is about using stimulus information to predict *what* to respond, temporal learning refers to the process of learning *when* to emit a response. Participants in speeded tasks are known to establish something like a time criterion for when to respond (i.e., the point in time at which they expect to respond) depending on the characteristics of the stimuli. For example, relatively easy stimuli are typically responded to faster when presented in a list where all of the stimuli are easy (i.e., a pure list) than when presented intermixed with harder items (i.e., a mixed list), suggesting that participants adapt their temporal expectations for response emission to the average difficulty experienced in the list (Lupker, Brown, & Colombo, 1997; Lupker, Kinoshita, Coltheart, & Taylor, 2003). Recently, Schmidt (2013c) extended this idea to explain PC effects in the Stroop task that have been obtained in the absence of biases created by contingency learning (Bugg, 2014a; Hutchison, 2011).

According to Schmidt's (2013c) temporal learning account, participants will develop a relatively fast temporal expectancy in an MC list (because most of the items in the list elicit relatively fast responses) and a relatively slower temporal expectancy in an MI list (because most of the items in the list elicit relatively slow responses). Participants will then use those temporal expectancies to anticipate when a response should be emitted. Specifically, congruent items, but not incongruent items, will speed up in the MC list because they can be processed rapidly enough to meet the fast temporal expectancy established for that list. As a result, the congruency effect will be relatively large in the MC list. Conversely, in an MI list, participants anticipate responding late and, as a result, there will be less pressure on them to elicit fast responses to congruent items. This situation would cause slower latencies to congruent items in an MI list relative to congruent items in an MC list. In contrast, according to Schmidt's account, a speed-up could potentially be observed for incongruent items in an MI list because they can be processed fast enough to meet the (slower) temporal expectancy established for that list. The result would be a relatively small congruency effect. In practice, however, hard-to-process stimuli appear to be relatively insensitive to temporal expectancies, at least in some situations (Kinoshita & Mozer, 2006; Kinoshita, Mozer, & Forster, 2011), meaning that the slow temporal expectancy developed for the MI list may have little impact on incongruent items. In any case, the core claim here is that learning of temporal expectancies can inflate the congruency effect in the MC list in comparison to the congruency effect in the MI list. Thus, similar to contingency learning, temporal learning can explain differences in the magnitude of congruency effects across MC and MI lists without invoking any type of conflict adaptation mechanism.

A critical piece of evidence in support of the temporal learning account of PC effects comes from statistical analyses of PC manipulations that take into account the role of temporal expectancies that individuals develop on a trial-by-trial basis. The idea for these analyses was first proposed by Kinoshita et al. (2011) within the framework of their Adaptation to the Statistics in the Environment (ASE) model of optimal response initiation and was then extended by Schmidt (2013a) in his Parallel Episodic Processing (PEP) model of color identification (see also Schmidt, De Houwer, & Rothermund, 2016). Although the two models were developed to explain different phenomena (relatedness proportion effects in masked priming in the case of the ASE model, PC effects in regular Stroop paradigms in the case of the PEP model) and place emphasis on different aspects of response emission (adaptation to perceived difficulty in the case of the ASE model, rhythmic responding in the case of the PEP model), the models make similar assumptions. First, performance on the current trial is influenced by the participant's knowledge of response times on the previous trials (i.e., the trial history), specifically the latencies on the most recent trials. Those latencies, especially the latency on the most recent trial (RT on trial $n - 1$), would function as an index of perceived task difficulty (in the ASE model) or as an index of the rhythm of responding (in the PEP model) that could be used to form an expectancy for response initiation latency on trial n. Thus, RT on trial $n - 1$ can function as an index of temporal expectancy for trial n, with a slower RT on trial $n - 1$ leading to a slower RT on trial n (Kiger & Glass, 1981; Taylor & Lupker, 2001). Second, as noted, easier stimuli are more prone to influences from trial history than are harder stimuli (although this pattern is not inevitable: Kinoshita & Mozer, 2006; Kinoshita et al., 2011).

The critical implication of these assumptions is that with easy stimuli strongly affected by RT on trial $n - 1$ (i.e., they will show a large slow-down following a slower RT on trial $n - 1$) and hard stimuli only weakly affected by RT on trial $n - 1$ (i.e., they will not show a large slow-down following a slower RT on trial $n - 1$), difficulty effects (i.e., the time difference between hard and easy stimuli) will decrease as RT on trial $n - 1$ increases.

Evidence for this pattern has been obtained from experimental data that were analyzed using linear mixed-effects models. This class of models, unlike traditional means-based ANOVAs, allows one to evaluate the impact of RT on trial n – 1, a trial-level continuous predictor, on performance on trial n. In several investigations, use of those analyses revealed that difficulty effects caused by visible or even subliminal distractors were modulated by trial history, with there being smaller effects when the RTs were slower on trial n – 1 (Huber-Huber & Ansorge, 2017, 2018; Kinoshita et al., 2011; Schmidt, 2013c; Schmidt & Weissman, 2016). Most importantly for the present discussion, the fact that congruency effects (and difficulty effects in general) are modulated by temporal expectancies is relevant for PC manipulations because fast RTs inevitably occur more frequently in MC lists than in MI lists. As faster RTs on trial n – 1 result in larger congruency effects, MC lists will tend to produce larger congruency effects than MI lists independent of contingency learning biases or a presumed conflict adaptation mechanism.

Support for the idea that temporal learning is at least partially responsible for PC effects comes from Schmidt's (2013c) re-analysis of the data from Hutchison's (2011) contingency-controlled items using the aforementioned linear mixed-effects model analyses. Those analyses not only replicated the finding of a significant PC effect originally reported by Hutchison (2011), they also indicated that congruency effects decreased with increasing RT on trial n – 1. Furthermore, this decreased congruency effect was accompanied by a reduction (although not an elimination) of the value of the beta parameter for the PC effect (i.e., the interaction) in the model, suggesting that PC effects and temporal learning effects explain common variance in the data. Schmidt interpreted this finding as indicating that temporal learning has the potential of generating PC effects on its own, a point he reinforced by showing that his PEP model, in which temporal learning was an implemented mechanism but adaptation to conflict frequency was not, could simulate Hutchison's (2011) results. At the very least, Schmidt's analysis suggests that temporal learning contributes to PC effects in contingency-controlled situations and, therefore, its role needs to be considered when analyses of PC manipulations are conducted.

More recently, however, Cohen-Shikora, Suh, and Bugg (in press) challenged this conclusion. They noted that the critical interaction between congruency and RT on trial $n - 1$ reported by Schmidt (2013c) was obtained when the typical positively skewed RT distribution was normalized with an inverse transformation (invRT = -1000/RT) in order to accommodate the assumption made by linear mixed-effects models that the dependent variable be normally distributed. A somewhat neglected downside of this type of analysis procedure is that nonlinear transformations of the dependent variable systematically alter the pattern and size of interaction terms, casting doubt on the reliability of analyses of interactions (Balota, Aschenbrenner, & Yap, 2013).

A solution to this problem is offered by generalized linear mixed-effects models, models which do not assume a normally distributed dependent variable and require, therefore, no RT transformation (Lo & Andrews, 2015). Using both inverse-transformed RTs in a linear mixed-effects model and untransformed (i.e., raw) RTs in a generalized linear mixed-effects model, Cohen-Shikora et al. (in press) re-analyzed Hutchison's (2011) dataset along with two additional datasets in which a PC manipulation had been implemented while controlling for contingencies (i.e., Bugg, 2014a; Gonthier et al., 2016). They reported that Schmidt's (2013c) finding that congruency effects decrease with increasing RT on trial $n - 1$ was obtained only with transformed data and not with untransformed data, with the latter data even providing evidence for the opposite pattern in some cases (i.e., congruency effects increased, rather than decreased, with increasing RT on trial n – 1). Furthermore, in all the datasets, the PC effect remained significant when temporal learning indices were included in the analyses, even when the value of the beta parameter for that effect was reduced due to the introduction of those indices. Finally, attempts to improve indices of temporal learning (e.g., by using mean RT on the three most recent trials as a predictor in the analyses) also yielded little evidence for the temporal learning account.

In sum, Cohen-Shikora et al.'s (in press) analyses suggest that previously reported evidence in support of the temporal learning explanation of the PC effect (Schmidt, 2013c) might have been biased due to the nonlinear transformation applied to RT data. Therefore, it would be advisable that research aiming to control for temporal learning avoid this bias by using a more appropriate statistical technique, such as generalized linear mixed-effects modelling.

The present research

Although conflict adaptation and non-conflict learning mechanisms are not necessarily mutually exclusive (Abrahamse, Braem, Notebaert, & Verguts, 2016; Egner, 2014), there has been a mounting debate in recent years concerning whether the classic empirical markers of conflict adaptation are, in fact, actually produced by non-attentional learning biases (e.g., contingency learning, temporal learning), biases that are typically found in manipulations designed to investigate what are presumed to be conflict adaptation effects. For some researchers, such debate has culminated in the idea that conflict adaptation might be an illusion (Schmidt et al., 2015). The fact that conflict adaptation tests are routinely used in clinical settings (e.g., Abrahamse et al., 2016; Bonnin, Houeto, Gil, & Bouquet, 2010) hints at the profound consequences borne by this idea. Presumed markers of conflict adaptation have been reported across the lifespan (e.g., Bugg, 2014a, 2014b) and across a variety of tasks (Bugg & Crump, 2012), with increasing reports coming from neuroimaging research (e.g., Braver, 2012; Sheth et al., 2012; West & Alain, 2000; Wilk, Ezekiel, & Morton, 2012), and these markers have been used in a number of diagnostic situations. An exact understanding of what these findings reflect is therefore crucial.

Motivated by these considerations, the present research aimed to re-examine the PC effect while at the same time accounting for potential non-conflict learning confounds. Specifically, we were interested in providing an answer to the following question: Would evidence for adaptation to conflict frequency emerge when non-conflict learning biases are controlled for or removed from the design altogether? Some attempts undertaken in this direction suggest that the answer might be “yes” (Bugg, 2014a; Bugg & Chanani, 2011; Bugg & Hutchison, 2013; Bugg, Jacoby, & Chanani, 2011; Hutchison, 2011; Gonthier et al., 2016). However, much of that research failed to consider non-conflict learning biases in their entirety and/or was based on experiments that deviate considerably from the original PC paradigm (Schmidt, 2013b, 2014a). In addition, very few attempts have been made to control for temporal learning when analyzing PC effects (Cohen-Shikora et al., in press; Schmidt, 2013c).

One primary objective of the present research was to examine adaptation to conflict frequency in a situation in which contingency learning could not contribute to the PC effect. According to Schmidt (2013a), learning of word-response contingencies is a two-step process: First, on each trial, participants encode information about the word, the color, and the response made into episodic memory. Second, any word presented on a subsequent trial will lead to the retrieval of past episodes involving that word, with facilitation occurring if the currently presented word requires the same response as most of its previous occurrences. Note that repetition appears to be an important aspect of this process. Words need to be repeated at least a few times in the experiment in order for responses associated with them to be able to influence subsequent behavior (Lin & MacLeod, 2018). Because this process is based on learning a predictive association between a specific word and a specific response (Schmidt et al., 2007; but see Schmidt, Augustinova, & De Houwer, 2018), learning of word-response contingencies would thus seem to require repeating words in the relevant colors. In other words, contingency learning would be impossible without repeated word distractors.

Based on these considerations, the present experiments examined whether PC effects emerge in a PC manipulation in a Stroop-like task where no contingency learning would be possible due to the fact that word distractors were never repeated (for a similar argument applied to a context-specific PC manipulation, see King, Korb, & Egner, 2012; see also Schneider, 2015, for a similar idea applied to cued task switching). Because only a limited set of words and colors can be used in the color-word Stroop task, a variant, the picture-word interference task, was used instead (note 1). Experiment 1 involved two picture-word interference tasks in which the proportion of congruent trials was manipulated in a list-wide fashion (for a similar manipulation in the picture-word interference task, see Bugg & Chanani, 2011). Experiment 1A required participants to categorize unrepeated target pictures paired with unrepeated word distractors. Participants in Experiment 1B were presented with the same materials but were required to name the pictures instead. To preview the results, regular PC effects were obtained in both tasks.

Another objective of the present experiments was to investigate the role of temporal learning in PC manipulations. To accomplish this goal, the data from Experiments 1A and 1B were analyzed using RT on trial *n* – 1 as an index of temporal expectancy in a mixed-effects model analysis, similar to those of Schmidt (2013c) and Kinoshita et al. (2011). However, similar to Cohen-Shikora et al. (in press), generalized linear mixed-effects models rather than linear mixed-effects models were used in these analyses. The reason is that, as noted, RTs typically violate the assumption made by linear mixed-effects models of a normally distributed dependent variable, a problem many researchers, including Schmidt (2013c), addressed by normalizing RTs with an inverse transformation. However, as shown by Cohen-Shikora et al., this solution is inappropriate when the research interest lies in interaction terms, as those terms are typically altered by nonlinear transformations of the dependent variable. Generalized linear mixed-effects models, on the other hand, provide a better solution in that, making no assumption about the distribution of the dependent variable, they require no RT transformation and permit a clearer interpretation of interactions (Lo & Andrews, 2015). In addition to the analyses reported by Cohen-Shikora et al., the present analyses can thus shed further light on the questions of whether evidence for temporal learning will be maintained when a more appropriate statistical technique is used and, more crucially, how potential PC effects obtained in Experiments 1A and 1B relate to the effects of temporal learning.

To further strengthen the conclusions from those analyses, Experiment 2 was conducted to isolate potential effects of temporal learning from adaptation to conflict frequency in a conflict-free picture naming task. As will be discussed below, trial-level analyses of Experiments 1A and 1B and results from Experiment 2 provide converging evidence that temporal learning does not appear to pose a challenge for conflict adaptation interpretations of PC effects, at least with the materials and tasks used in the present experiments.

Experiments 1A & 1B

If repetitions of word distractors are necessary for learning associations between specific words and specific responses, learning of such associations should be impossible when word distractors are never repeated. Such a situation should thus allow researchers to examine potential effects of adaptation to conflict frequency in the absence of the contingency learning confound that is typically found in classic PC manipulations using the color-word Stroop task (Melara & Algom, 2003; Schmidt & Besner, 2008). As noted, to this end, a picture-word interference task was used. In the picture-word interference task, participants are required to identify a picture while ignoring a word superimposed on it. Similar to the color-word Stroop task, two types of items were used in the task variant employed in the present set of experiments: congruent items, with words specifying the name of the picture itself (e.g., the picture of a dog with the word DOG superimposed on it), and incongruent items, with words unrelated to the picture (i.e., belonging to a different semantic category than the picture's), as well as not appearing as target pictures in the experiment (e.g., the picture of a dog had the unrelated word BED superimposed on it and no picture of a bed appeared in the experiment).

Using a between-subject PC manipulation, participants were assigned to either an MI or an MC list. (note 2) Experiment 1A required participants to identify unrepeated target pictures paired with unrepeated word distractors as members of a semantic category and to respond vocally. Participants in Experiment 1B were presented with the same materials but were required to name the pictures instead. Note that despite the materials being the same, the word distractors used are more relevant to picture naming than they are to categorization. For example, the word DOG should help more with naming the picture of a dog than it should help with categorizing a dog as an animal. Furthermore, unlike in picture naming, in picture categorization not only incongruent words but also congruent words are absent from the response set. Indeed, in a picture categorization task, Lupker and Katz (1981) obtained only a (nonsignificant) 12-ms difference between conditions that are analogous to the congruent and incongruent conditions of the present experiment. In contrast, picture naming was expected to elicit a much larger congruency effect because of the relevance of word distractors to the task (e.g., Underwood, 1976). Nonetheless, both picture categorization and picture naming were used in order to investigate whether the presence of PC effects might depend on the nature of the task and the basic magnitude of the congruency effect.

In response to the suggestions of two reviewers of the initial version of the present paper, we examined not only the PC effect but also the congruency sequence effect, i.e., the finding that in interference tasks, congruency effects are larger following a congruent trial than following an incongruent trial (Gratton, Coles, & Donchin, 1992). Traditionally thought of as a marker of conflict adaptation (e.g., Botvinick et al., 2001), this finding, similar to the PC effect, has recently received several alternative interpretations (e.g., Hommel, Proctor, & Vu, 2004; Mayr, Awh, & Laurey, 2003), including a temporal learning interpretation (Schmidt & Weissman, 2016). This temporal learning interpretation partially relies on the same interaction that is thought to be responsible for the PC effect (i.e., decreasing congruency effects with higher RT on trial $n - 1$), an interaction that, crucially, Schmidt and Weissman (2016) observed when analyzing inverse RTs. As noted, such a situation makes the interpretation of interaction terms dubious. As such, our use of a generalized linear mixed model analysis, an analysis which permits usage of untransformed RTs, provided a valuable opportunity to assess whether the unreliability of the temporal learning interaction reported by Cohen-Shikora et al. (in press) in the context of the PC effect also applies in the context of the congruency sequence effect, another important marker of conflict adaptation. These additional analyses for Experiments 1A and 1B, along with a discussion of the control and the temporal learning account of the congruency sequence effect, can be found in the Appendix.

Method

Participants

An *a priori* power analysis was performed using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to calculate the sample size needed to have a power of .80 for obtaining a PC effect. Based on the effect size reported by Bugg and Chanani (2011) for a PC effect using contingency-controlled items in a picture-word interference task, we determined that a minimum sample size of 32 participants would be needed. Forty-eight participants took part in Experiment 1A (picture categorization) and another 51 took part in Experiment 1B (picture naming). In Experiment 1B, 1 participant was removed due to an equipment failure and 2 more were removed because of an excessive number of errors and null responses (above 25%), leaving 48 participants. Participants were all students at the University of Western Ontario aged 18–23 years (*SD* = 1.03) and had normal or corrected-to-normal vision. All were native English speakers. Their participation was compensated with course credit.

Materials

One hundred and twenty-five line drawings were sourced from the International Picture Naming Project (IPNP) database (Szekely et al., 2005). Nineteen pictures from the internet matching as closely as possible the style of the IPNP pictures were added to the set, for a total of 144 target pictures, 480 x 480 pixels in size. Of these, 36 represented an animal, 36 represented a human being, 36 represented some type of food, and 36 represented a man-made object. IPNP norms and pilot testing ensured that there was high agreement among English-speaking individuals on the semantic category and name of each picture. One-hundred and forty-four English word distractors, different from the modal names of the pictures, were also selected. As with the target pictures, 36 denoted an animal, 36 a human being, 36 some type of food, and 36 a man-made object. The word distractors were matched in length and CELEX frequency (Baayen, Piepenbrock, & van Rijn, 1993) with the pictures' modal names. Each picture was paired with the modal name of the picture (congruent item) and with an unrelated word belonging to one of the other three categories (incongruent item), with each of the incongruent categories being equally represented across items. For example, among the 36 pictures of animals, 12 were paired with an unrelated word denoting a person, 12 with an unrelated word denoting a food, and 12 with an unrelated word denoting an object. Powerpoint software was used to superimpose the word in 32-point Courier New font in the center of the picture. A light white glow around words' letters was added to ensure that the word was clearly visible. A sample of the stimuli used in Experiments 1A and 1B is presented in Figure 1.

Figure 1.

Sample Stimuli Used in Experiments 1A and 1B

Note. Represented are congruent items (panel A), and incongruent items (panel B) for each of the four categories (ANIMAL, PERSON, FOOD, and OBJECT). In this sample, the pictures of the elephant and the shovel come from the International Picture Naming Project (Szekely et al., 2005) whereas the pictures of the astronaut and bacon were sourced from the internet.

Lists were constructed so that for half of the lists, 25% of the items were congruent (MI lists), and for the other half, 75% of the items were congruent (MC lists). Specifically, in the MI lists, 36 pictures were presented with their congruent word, and 108 pictures with their incongruent word. Conversely, in the MC lists, 108 pictures were presented with their congruent word, and 36 pictures with their incongruent word. Each of the semantic categories was equally probable among the congruent and among the incongruent items (e.g., in the MC list, 9 of the 36 incongruent pictures were animals, 9 were human beings, 9 were food items, and 9 were objects, etc.). Similarly, each of the three semantic categories of incongruent word distractors was equally probable among incongruent items (e.g., in the MC list, 3 of the 9 incongruent animal pictures appeared with an unrelated word denoting a person, 3 with an unrelated word denoting a food item, and 3 with an unrelated word denoting an object).

Lists were also counterbalanced so that each picture appeared with its congruent and incongruent word distractor in both MI and MC lists. To this end, the pictures were randomly divided into four sets, A, B, C, and D. In List 1 of the MC lists, pictures in sets A, B, and C would serve as congruent pictures and pictures in set D would serve as incongruent pictures. In List 2 of the MC lists, pictures in sets A, B, and D would serve as congruent and pictures in set C would serve as incongruent, and so on. Construction of the MI lists was done similarly, with List 1 having pictures in sets A, B, and C serving as incongruent and pictures in set D serving as congruent, etc. Pictures in each set included an equal number of pictures of animals, people, foods, and objects. Overall, 4 MI and 4 MC lists were constructed.

Procedure

Participants were tested individually in a quiet room, seated approximately 60 cm away from a monitor upon which the stimuli were presented. Each trial began with a fixation symbol ("+") displayed for 500 ms in the center of the screen, followed by a picture with a word superimposed on it, displayed for 3000 ms or until the participant's response. Responses were recorded with a microphone connected to the testing computer. Participants in Experiment 1A were instructed to categorize the picture using one of four semantic categories (ANIMAL, PERSON, FOOD, OBJECT) as responses. Care was taken to explain the differences between these categories in order to minimize potential ambiguities (e.g., living animals that are typically eaten by humans, such as chicken, being classified as food). Participants in Experiment 1B were instructed to name the picture instead. In both experiments, participants were told to ignore the word superimposed on the picture and to respond as quickly and as accurately as possible. Participants were randomly assigned to one of the eight lists in Experiments 1A and 1B. Thus, each participant performed only one list for a total of 144 trials.

Prior to the experiment, participants performed a practice session involving twelve items, different from the items in the experiment and mirroring the proportion of congruent items in the upcoming list. They received no feedback. Trials were presented in a different random order for each participant. DMDX (Forster & Forster, 2003) software was used to present the stimuli and collect the data.

Results

In these and the following experiments using vocal responses, response waveforms were manually inspected with CheckVocal (Protopapas, 2007) to determine the accuracy of the response and the correct placement of timing marks. RTs were defined as the time interval between stimulus onset and the beginning of the vocal response. Errors were marked using a conservative criterion: Any response that was not the expected response was considered an error, no matter how close it was to the expected response (e.g., “people” instead of “person” for Experiment 1A, or “cop” instead of “policeman” in Experiment 1B). Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than the time limit (accounting for 0.4% and 2% of the data points in Experiments 1A and 1B, respectively) were discarded. For the latency analyses, trials on which an error was made were discarded, as were the trials for which an error or a too-fast response (< 300 ms) or a too-slow response (> 3000 ms) was made on the preceding trial.

The latencies and the error rates were analyzed using generalized linear mixed-effects modeling in R version 3.4.3 (R Development Core Team, 2015), treating subjects and items (i.e., the target pictures) as random effects and treating Congruency (congruent vs. incongruent) and List Type (MI vs. MC) as within-subject and between-subject fixed effects, respectively (Baayen, 2008; Baayen, Davidson, & Bates, 2008). 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 model was fit by maximum likelihood with the Laplace approximation technique. The lme4 package, version 1.1-15 (Bates, Mächler, Bolker, & Walker, 2015), was used to run the generalized linear mixed-effects model and obtain probability values.

In the latency analyses, a generalized linear mixed-effects model was used instead of a linear mixed-effects model because generalized linear models, unlike linear models, do not assume a normally distributed dependent variable. Therefore, these models can accommodate the typically positively skewed distribution of raw RT data with there being no need to use non-linear transformations, known to systematically alter interaction terms (Balota et al., 2013). A Gamma distribution was used to fit the raw RTs, with an identity link between fixed effects and the dependent variable (Lo & Andrews, 2015). Note that convergence tests for generalized linear mixed-effects models in the current version of lme4 tend to generate many false positives (Bolker, 2018). In the following, we report the data from the BOBYQA optimizer, which returned estimates that were equivalent to other optimizers but never issued convergence warnings. Unlike the error analyses, latency analyses included RT on trial $n - 1$ as a fixed effect to control for temporal learning (Schmidt, 2013c; Schmidt & Weissman, 2016). Standardized (i.e., centered and scaled) RTs on trial $n - 1$ were used instead of raw RTs in order to avoid spurious correlations between the intercept and the slope and to help in evaluating and interpreting the model (Bolker, 2018; Kinoshita et al., 2011; Schielzeth, 2010). The statistical model for the latency analysis was: RT = glmer(RT ~ congruency * list_type * SprevRT + (1|subject) + (1|item), family = Gamma (link = “identity”), control = glmerControl(optimizer="bobyqa")). The statistical model for the error rate analysis was: Accuracy = glmer(accuracy ~ congruency * list_type + (1|subject) + (1|item), family = binomial, control = glmerControl(optimizer="bobyqa")). The mean RTs and error rates based on by-subject data for Experiments 1A and 1B are shown in Table 1. Scatterplots visualizing the relation between RT on trial *n* – 1 and the congruency effect on trial *n* are shown in Figure 2 for Experiment 1A and in Figure 3 for Experiment 1B. The data and R scripts used for the analyses are publicly available at https://osf.io/jnzgb/.

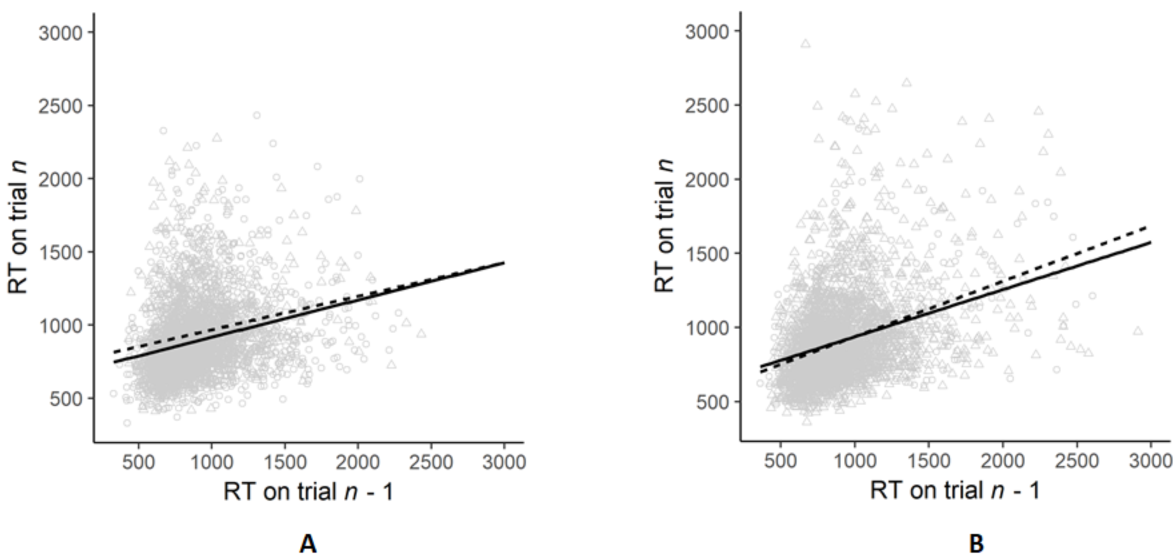
Table 3.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Experiments 1A and 1B

	RTs			Error rates		
Congruency	MC list	MI list	List effect	MC list	MI list	List effect
Experiment 1A						
Congruent	893 (24)	912 (31)	19	1.4 (.3)	2.1 (.6)	.7
Incongruent	947 (24)	910 (37)	-37	2.7 (.6)	2 (.4)	-.7
Congruency Effect	54	-2	-56	1.3	-.1	-1.4
Experiment 1B						
Congruent	764 (26)	797 (21)	33	1.1 (.3)	2 (.9)	.9
Incongruent	1106 (38)	1041 (22)	-65	14 (1.7)	11.2 (.9)	-2.8
Congruency Effect	342	244	-98	12.9	9.2	-3.7

Figure 2.

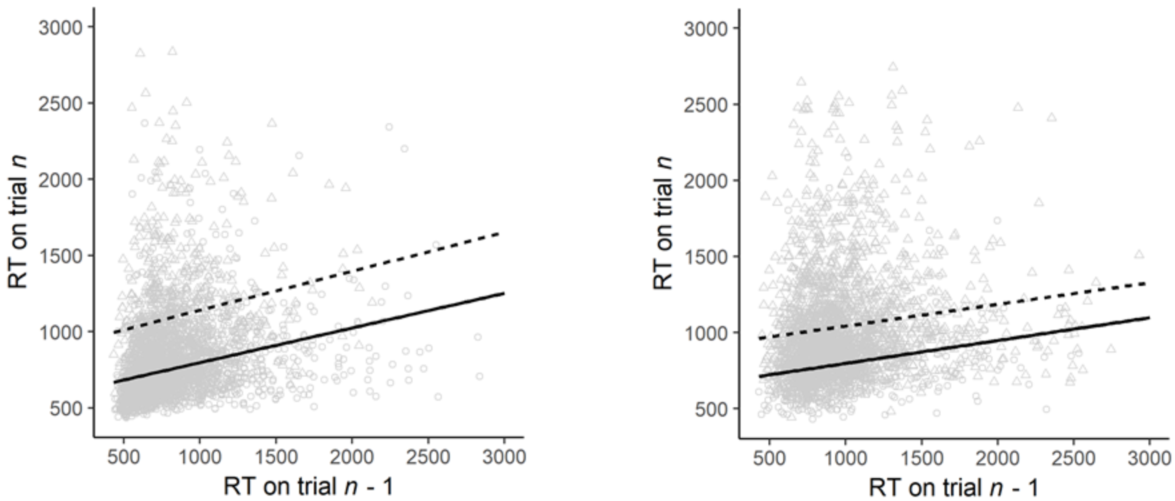
The Impact of RT on Trial n – 1 on Congruency Effects on Trial n in Experiment 1A



Note. The scatterplots represent the relation between RT on trial n – 1 and the congruency effect on trial n in the MC list (panel A), and in the MI list (panel B). Individual observations for congruent and incongruent trials are marked with triangles and circles, respectively. Regression slopes for the congruent condition and for the incongruent condition are marked with solid and dashed lines, respectively.

Figure 3.

The Impact of RT on Trial n – 1 on Congruency Effects on Trial n in Experiment 1B



A **B**

Note. The scatterplots represent the relation between RT on trial *n* – 1 and the congruency effect on trial *n* in the MC list (panel A) and in the MI list (panel B). Individual observations for congruent and incongruent trials are marked with triangles and circles, respectively. Regression slopes for the congruent condition and for the incongruent condition are marked with solid and dashed lines, respectively.

Experiment 1A (picture categorization)

RT. There were significant main effects of Congruency (congruent faster than incongruent), $ß = -10.68$, $SE = 2.36$, $z = -4.53$, $p < .001$, and RT on trial $n - 1$ (faster responses with lower RT on trial $n - 1$), $ß = 23.85$, $SE = 2.82$, $z = 8.44$, $p < .001$ (the main effect of List Type was not significant, $ß = -3.58$, $SE = 4.56$, $z = -.78$, $p = .43$). The interaction between Congruency and List Type was significant as well, $ß = -14.14$, $SE = 2.54$, $z = -5.56$, $p < .001$, indicating that a classic PC effect was obtained, with a larger effect of Congruency in the MC (54 ms) than in the MI (-2 ms) condition. Interestingly, the interaction between Congruency and RT on trial $n - 1$, indexing temporal learning, was not significant, $ß = -1.62$, $SE = 2.61$, $z = -0.62$, $p = .54$, but the three-way interaction between Congruency, List Type, and RT on trial $n - 1$ was, $ß = 6.01$, $SE = 2.60$, $z = 2.31$, $p = .021$.

To explore the three-way interaction, MC and MI lists were analyzed separately. MC lists showed both main effects of Congruency, $ß = -24.89$, $SE = 3.52$, $z = -7.07$, $p < .001$, and RT on trial $n - 1$, $ß = 25.60$, $SE = 3.99$, $z = 6.42$, $p < .001$, but no interaction between the two, $ß = 4.51$, $SE = 3.75$, $z = 1.20$, $p = .23$. In MI lists, on the other hand, RT on trial $n - 1$ was significant, $ß = 22.29$, $SE = 4.58$, $z = 4.86$, $p < .001$, but Congruency was not, $ß = 3.26$, $SE = 3.66$, $z = .89$, $p = .37$. Here, the interaction between Congruency and RT on trial $n - 1$ was significant, $ß = -8.08$, $SE = 3.91$, $z = -2.07$, $p = .039$. Note, however, that the pattern of this interaction is the opposite of that predicted by temporal learning: As illustrated in Figure 2B, the congruency effect on trial n increased, rather than decreased, with higher latencies on trial $n - 1$.

Error rates. Neither Congruency nor List Type was significant. The interaction between the two was marginal, $ß = .17$, $SE = .10$, $z = 1.78$, $p = .075$, indicating a tendency for the Congruency effect to be larger in the MC (1.3%) than in the MI condition (-.1%).

Experiment 1B (picture naming)

RT. There were significant main effects of Congruency (congruent faster than incongruent), $ß = -143.71$, $SE = 2.62$, $z = -54.80$, $p < .001$, List Type (MI faster than MC), $ß = 24.96$, $SE = 4.39$, $z = 5.68$, $p < .001$, and RT on trial $n – 1$ (faster responses with lower RT on trial $n – 1$), $ß = 22.84$, $SE = 3.04$, $z = 7.51$, $p < .001$. The only significant interaction was that between Congruency and List Type, $ß = -23.64$, $SE = 2.89$, $z = -8.19$, $p < .001$, indicating that a classic PC effect was obtained, with a larger effect of Congruency in the MC (342 ms) than in the MI condition (244 ms). Neither the interaction between Congruency and RT on trial $n – 1$, $ß = -3.88$, $SE = 2.66$, $z = -1.46$, $p = .14$, nor the three-way interaction between Congruency, List Type, and RT on trial $n – 1$, $ß = -.18$, $SE = 3.01$, $z = -.06$, $p = .95$, was significant.

Error rates. There was a main effect of Congruency (congruent more accurate than incongruent), $ß = 1.36$, $SE = .09$, $z = 14.71$, $p < .001$. In addition, Congruency interacted with List Type, $ß = .27$, $SE = .09$, $z = 3.00$, $p = .003$, with the congruency effect being larger in the MC (13.0%) than in the MI condition (9.1%).

Discussion

Both Experiment 1A and Experiment 1B produced clear PC effects in a situation where learning of direct associations between words and responses was impossible. Note that, as suggested by previous findings (Lupker & Katz, 1981), the basic congruency effect was much smaller in

Experiment 1A (picture categorization: 26 ms) than in Experiment 1B (picture naming: 293 ms). However, the congruency effect was similarly modulated by conflict frequency across the two tasks, with MI lists in Experiment 1B showing a congruency effect reduced by 98 ms compared to MC lists, and Experiment 1A showing the elimination of the congruency effect in MI lists. As discussed, a contingency learning account would not be able to explain these effects.

Temporal learning also does not seem to offer a reasonable explanation for the present findings. For temporal learning to account for PC effects, one would need to find that congruency effects on trial *n* get smaller as RT on trial *n* – 1 increases, indicating that participants use previous experience in the task to form and adjust to temporal expectancies for responding in the way suggested by Schmidt (2013c). Using generalized linear mixed-effects models to fit raw RTs, robust main effects of RT on trial *n* – 1 were found, with overall slower responses on trial *n* as RT on trial *n* – 1 increased. These sequence effects are routinely reported in speeded tasks (Kinoshita et al., 2011; Taylor & Lupker, 2001). More importantly, no interaction between RT on trial *n* – 1 and the congruency effect on trial *n* was found in Experiment 1B, whereas a complicated pattern emerged in Experiment 1A. Specifically, in Experiment 1A, MI lists (but not MC lists) showed an interaction involving the opposite pattern than was expected from the temporal learning account: Larger congruency effects on trial *n* the higher the RT on trial *n* – 1. While the cause of this result is unclear, it should be noted that Cohen-Shikora et al. (in press) also reported inconsistent temporal learning patterns across the three datasets that they analyzed. (note 3) In general, it is safe to conclude from the overall pattern of results that temporal learning could not have produced, or even contributed to the production of, the PC effects reported here.

In sum, Experiments 1A and 1B showed that PC effects emerge even in the absence of temporal learning and word-response contingencies, a finding that challenges the view that mechanisms of this sort provide a sufficient account of the PC effects that are reported in the literature such that adaptation to conflict frequency may not be a mechanism humans use (Schmidt, 2013b).

To consolidate the idea that temporal learning has little to do with the PC effect obtained in Experiments 1A and 1B, Experiment 2 was conducted to disentangle conflict frequency from potential effects of temporal learning. Note that Schmidt's (2013c) temporal learning account assumes that temporal expectancies for responding are altered as a result of any manipulation that induces appreciable differences in response rhythm. The type of manipulation which can accomplish such an alteration involves changes in the relative frequency of easy and hard stimuli, with the nature of the difficulty elicited by those stimuli playing little or no role. Since difficulty does not need to derive from conflict from an irrelevant dimension, temporal learning should not be specific to the type of task used in Experiments 1A and 1B, i.e., tasks where conflict/interference from an irrelevant dimension produces the difficulty effect. That is, according to the temporal learning account of PC effects, any task in which the proportion of easy and hard items is manipulated should produce differences in the temporal expectancies being formed for responses. As a result, the magnitude of difficulty effects should parallel the pattern observed for congruency effects in the PC effect: Smaller difficulty effects in lists where most of the items are hard and larger difficulty effects in lists where most of the items are easy (Schmidt, 2013c, 2014a, 2016). Experiment 2 tested this prediction for the pictures used in Experiments 1A and 1B, which were presented without the superimposed words and modified in such a way that they were easier or harder to respond to.

Experiment 2

Following Schmidt's procedure (2013c; Schmidt & Weissman, 2016), in Experiments 1A and 1B temporal learning was accounted for in the analyses by using RT on trial *n* – 1 as an index of temporal expectancy, with a lower RT on trial *n* – 1 indicating a faster temporal expectancy for trial *n*. However, the predicted interaction between congruency and *RT* on trial n – 1, with smaller congruency effects the higher the RT on trial *n* – 1, was not found. In fact, Experiment 1A even produced evidence for a reversed interaction in MI lists, with larger congruency effects following higher RTs on trial *n* – 1. These results are in line with recent failures to obtain regular temporal learning effects using untransformed RTs (Cohen-Shikora et al., in press), suggesting that the nonlinear transformations reported in previously published papers (Huber-Huber & Ansorge, 2017, 2018; Kinoshita et al., 2011; Schmidt, 2013c; Schmidt & Weissman, 2016) might have systematically biased the interaction of interest in the direction predicted by temporal learning.

Statistical quirks aside, however, it must be acknowledged that supporters of temporal learning accounts have pointed out that RT on trial *n* – 1 is likely a noisy approximation of temporal expectancies (Kinoshita et al., 2011; Schmidt, 2013c), although several attempts, reported by Cohen-Shikora et al. (in press), to use a less noisy index (e.g., mean RT on the three most recent trials) also failed to produce consistent evidence for the temporal learning account of the PC effect. In fact, one could argue that internally constructed temporal expectancies might deviate considerably from the measured response time on one or more of the preceding trials, an argument that might find support in the observation that time perception is often prone to biases (e.g., Taylor & Lupker, 2006, 2007). The implication is that the analyses performed for

Experiments 1A and 1B might not provide the best means of determining whether temporal learning is a potential contributor to the PC effect observed in those experiments.

A better way to deal with this issue might be found in another approach used by Schmidt (2013c, 2014a, 2016) in his attempts to demonstrate a potential role for temporal learning in the PC effect, an approach that does not require using any index of temporal expectancy in the analyses and thus avoids the potential problems associated with the noisiness of such measures. Relying on the assumption that temporal learning should operate similarly in interference tasks and in tasks in which difficulty does not derive from interference from an irrelevant dimension (e.g., perceptual tasks), this approach involves manipulating the proportion of easy items in a task of the latter type.

Indeed, the existence of a temporal learning mechanism of the sort described by Schmidt (2013c) implies that any task in which the proportion of easy and hard items is manipulated should produce differences in the magnitude of effect sizes in ways that are compatible with the changes observed for congruency effects in the PC effect. Specifically, mostly easy (ME) lists (i.e., lists in which most of the items are relatively easy to process) will favor development of a fast temporal expectancy that can be met by items that allow fast responses (i.e., "easy" items), but not by items that are relatively hard to process (i.e., "hard" items). The result is a speed-up for only the easy items and, hence, a large difficulty effect. Mostly hard (MH) lists (i.e., lists in which most of the items are relatively hard to process), on the other hand, will favor development of a slow temporal expectancy. Because participants anticipate responding relatively late, there will be no reason for them to speed up responses to easy items in this situation, causing them to produce longer latencies. In contrast, as noted, it is possible that latencies for hard items may decrease if they can be processed fast enough to meet the slower temporal expectancy, although, as Schmidt (2013c) has argued, those items tend to be insensitive to temporal expectancies (see also Kinoshita et al., 2011; Schmidt & Weissman, 2016). The end result is that learning of temporal expectancies should produce larger difficulty effects in ME lists than in MH lists.

Schmidt did, in fact, obtain evidence of such a Proportion-Easy (PE) effect in a number of studies where no irrelevant dimension was used (Schmidt, 2013c, 2014a, 2016). For example, in a letter identification task, Schmidt (2014a) found, as would be expected, shorter latencies for high-contrast letters ("easy" items) than for low-contrast letters ("hard" items). Most importantly, the size of this difficulty effect was modulated by the proportion of easy items in the list, with larger difficulty effects in ME lists than in MH lists, similar to the PC effect in the Stroop task. Although this finding is not crucial evidence that the mechanism driving PC effects in the Stroop task and PE effects in non-conflict tasks is the same, it does suggest that temporal learning might play an important role in determining PC effects (Schmidt, 2013c). Specifically, this approach provides a proof of principle that a PC-like effect can be obtained even when little or no conflict is present in the task, suggesting that the mechanism responsible for this PC-like effect might also be operating when conflict is present, e.g., in Stroop paradigms.

The goal of Experiment 2 was to examine a similar, non-conflict situation with the pictures used in Experiments 1A and 1B. Similar to Schmidt's (2013c, 2014a, 2016) use of high-contrast and low-contrast letters, high-resolution and low-resolution pictures were used as easy and hard items, respectively, and participants were assigned to a ME list where most of the pictures had a high resolution, or to a MH list where most of the pictures had a low resolution. Following

Schmidt's (2013c) temporal learning account, it was hypothesized that easy items would be responded to faster in ME than in MH lists, and hard items would be responded to faster (or, at least, no more slowly) in MH than in ME lists. As a result, a PE effect would be obtained, with ME lists showing a larger difficulty effect than MH lists.

It is important to note, however, that a different outcome could be expected from an alternative temporal learning account, specifically, one derived from the literature on blocking effects (Chateau & Lupker, 2003; Lupker et al., 1997, 2003; Kinoshita & Mozer, 2006; Rastle, Kinoshita, Lupker, & Coltheart, 2003; Taylor & Lupker, 2001). Blocking effects refer to the finding that when relatively easy and relatively hard items are mixed in a block (i.e., a "mixed" block, typically with 50% easy and 50% hard stimuli), latencies tend to be more homogeneous compared to latencies for easy versus hard items presented by themselves in "pure" blocks (i.e., blocks where all of the stimuli are either easy or hard). Specifically, there is a mixing cost for easy stimuli (i.e., slower latencies for easy stimuli in mixed blocks than in pure easy blocks) and a mixing benefit for hard stimuli (i.e., faster latencies for hard stimuli in mixed blocks than in pure hard blocks). Lupker and collaborators interpreted this pattern as evidence that participants in speeded tasks establish a time criterion representing the time at which they expect, and will attempt, to initiate a response. Importantly, the placement of the time criterion is dependent upon the characteristics of the stimuli in the block: The criterion will be set early in a pure easy block, late in a pure hard block, and in an intermediate position in a mixed block.

This reasoning can be easily extended to comparisons among mixed lists varying in the proportion of easy items. That is, in ME lists, the criterion will be placed relatively early (although not as early as in a pure easy list), whereas in MH lists it will be placed relatively late (although not as late as in a pure hard list). As a result, both easy and hard items should be responded to faster in ME lists than in MH lists. In other words, under the assumption that adjustments of time criterion are similar for easy and hard items, one might expect main effects of difficulty (easy faster than hard) and list type (ME faster than MH), but not necessarily their interaction, i.e., difficulty effects may be equivalent in ME and MH lists. Of importance, the latter pattern (i.e., similar adjustments of the time criterion for easy and hard items) typically emerges in word naming tasks but not in (button-press) lexical decision tasks (in which only easy items appear to be affected by adjustments of time criterion), even when using the same items in the two tasks (Kinoshita & Mozer, 2006). Since the present experiments used naming, and blocking effects occur for pictures and words alike (Lupker et al., 2003), the expectation would be that the interaction predicted by the temporal learning account would not arise in the task investigated in Experiment 2 (i.e., picture naming).

Method

Participants

An a priori power analysis was performed using G*Power 3.1 (Faul et al., 2009) to calculate the sample size needed to have a power of .80 for obtaining a PE effect. Based on the effect size reported by Schmidt (2013c) for a PE effect in a letter identification task, we determined that a minimum of 68 participants would be needed. One hundred and twelve participants took part in the experiment. Nineteen participants were removed because of an excessive number of errors and null responses (above 25%), leaving 93 participants. They were all students at the

University of Western Ontario aged 18–27 years (*SD* = 1.21) and had normal or corrected-to-normal vision. All were native English speakers. They received either $10 or course credit for their participation.

Materials

The materials were derived from those used in Experiments 1A and 1B. The congruent pictures with their superimposed word removed functioned as high-resolution, easy items. The incongruent pictures with their superimposed word removed were degraded by resizing them to a quarter of their size and then inflating them back to their original size with Bulk Resize Photos (https://bulkresizephotos.com), thus resulting in a lower-resolution image. Those pictures functioned as low-resolution, hard items. Other than high-resolution and low-resolution pictures replacing the congruent and incongruent pictures, respectively, lists and counterbalancing of the items were identical to those in Experiments 1A and 1B, resulting in four ME lists and four MH lists.

Procedure

The procedure was identical to that in Experiment 1B, with the exception that, of course, superimposed words were not mentioned in the instructions, and participants were simply required to name the pictures as quickly and as accurately possible.

Results

Analyses were performed in the same way as was done for Experiments 1A and 1B with the exception that the factor Congruency was replaced with the factor Difficulty (easy vs. hard) and the two levels of the factor List Type were ME and MH instead of MC and MI. In addition, in line with previous PE manipulations (Schmidt, 2013c, 2014a, 2016), RT on trial *n* – 1 was not included as a predictor in the latency analysis because there is no need to control for temporal learning in this context: Any differences between difficulty effects across the two list types should be produced by the learning of temporal expectancies induced by the Difficulty factor itself. (note 4)

Prior to the analyses, invalid trials due to equipment failures and responses faster than 300 ms or slower than the time limit, accounting for 3.3% of the data points, were discarded. Since RT on trial *n* – 1 was not used as a predictor in the latency analysis, only trials where an error was made on the current trial were discarded. The mean RTs and error rates are presented in Table 2. The data and R scripts used for the analyses are publicly available at https://osf.io/jnzgb/.

Table 4.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Experiment 2

	RTs			Error rates		
Difficulty	ME list	MH list	List effect	ME list	MH list	List effect
Easy	908 (17)	948 (20)	40	9.1 (.5)	9.6 (.8)	.5
Hard	972 (19)	1013 (18)	41	15 (1.1)	14.8 (.6)	-.2
Difficulty Effect	64	65	1	5.8	5.1	-.7

RT. There were significant main effects of Difficulty (easy faster than hard), $ß = -31.74$, $SE = 2.19$, $z = -14.53$, $p < .001$, and List Type (faster responses in the ME than the MH condition), $ß = -17.29$, $SE = 3.05$, $z = -5.67$, $p < .001$. However, Difficulty and List Type did not interact, $ß = -.57$, $SE = 1.94$, $z = -.29$, $p = .77$, reflecting equivalent effects of Congruency in the ME (64 ms) and MH lists (65 ms).

Error rates. The only significant effect was that of Difficulty, $ß = .36$, $SE = .04$, $z = 9.99$, $p < .001$.

Discussion

In the present experiment, the difficulty of pictures, instead of word-picture congruency, was manipulated by using high- and low-resolution pictures, similar to the high- and low-contrast letters used by Schmidt (2013c, 2014a, 2016). Unlike Schmidt's results, however, difficulty effects were not any larger in lists where most of the trials were easy than in lists where most of the trials were hard. In fact, the magnitude of difficulty effects was identical in the two conditions, thus failing to replicate the pattern predicted by Schmidt's temporal learning account. Note, however, that the type of list participants were assigned to – ME or MH – did have an effect, with overall faster latencies in ME than MH lists. Thus, this pattern seems more consistent with the time criterion account (Lupker et al., 1997), according to which ME and MH lists should lead to relatively early and late time criteria, respectively, affecting latencies for easy and hard items in a similar way, at least in a naming situation. Most importantly, this pattern is consistent with the analyses performed for Experiments 1A and 1B in indicating that temporal learning may have little or no role in modulating difficulty effects in both interference and non-interference tasks.

General Discussion

Do humans adapt to conflict frequency? Recently, some researchers have cast doubt on this idea by pointing out that PC effects in the Stroop task might be caused by factors other than conflict adaptation, namely, word-response contingency learning and temporal learning (Schmidt, 2013b). The present research addressed this question using a picture-word interference task where contingencies were eliminated and temporal learning was controlled for. Clear, contingency-free PC effects emerged in both picture categorization (Experiment 1A) and picture naming (Experiment 1B) tasks, a finding that challenges the view that contingency learning is a critical factor driving PC effects in the Stroop task. Similarly, the analysis of the impact of trial $n - 1$ latency challenges the view that temporal learning has an important role in producing PC effects. Together, these results clearly demonstrate that it is not the case that PC effects are unobservable when those factors are controlled for (Schmidt, 2013b; Schmidt & Besner, 2008).

It must be noted that although the color-word and picture-word interference tasks are thought to reflect the same underlying processes (see note 1), one important difference between the typical color-word Stroop task used in the literature and the picture-word interference task used here is that the former, but not the latter, elicits response interference. That is, in most implementations of the color-word Stroop task, incongruent words are also used as responses (e.g., the word YELLOW is presented in an experiment in which yellow color targets are also used), whereas incongruent word distractors were not responses in either Experiment 1A nor Experiment 1B (e.g., BED appeared as a word distractor for the picture of a dog but not as a target picture). Using distractors that are not used as responses is known to reduce interference from the irrelevant dimension in both color-word and picture-word interference tasks (Lupker & Katz, 1981; Proctor, 1978), suggesting that response interference, among other factors, contributes to Stroop and Stroop-like effects (La Heij, 1988). As such, what the results of Experiments 1A and 1B provide is evidence for adaptation to conflict frequency even in a situation in which conflict was likely less intense than in a typical Stroop task because response interference was playing little, if any, role.

It is also important to acknowledge, however, that recent findings, published as the present research was in progress, suggest that learning of contingencies might occur at a more abstract level than previously thought. Schmidt et al. (2018) reported two color identification experiments in which words belonging to three different semantic categories were used as distractors, each category being predictive of one color. Similar to the experiments reported here, each individual word was presented only once, thus eliminating individual word-response contingencies. A category-based contingency effect was observed, with faster and more accurate responses when a category item was presented in the color in which most of the other items of that category were presented. Note that although the present experiments were designed to eliminate individual word-response contingencies, they allowed for category-based contingency learning. For example, words denoting animals were mostly associated with pictures of animals in MC lists, whereas they were equally associated with each of the four semantic categories in MI lists. Thus, participants in MC lists potentially could have used the category of the word distractor to predict the response, leading to a speed-up on high-contingency congruent items and therefore, an inflated congruency effect in MC lists.

An account of this sort, however, seems to be unlikely for a couple of reasons: First, the effects reported by Schmidt et al. (2018) (11 ms and a nonsignificant 2 ms in their Experiments 1 and 2, respectively) seem too small to offer a convincing alternative interpretation of the present findings (note that classic word-response contingency learning effects are on the order of 40-60 ms: e.g., Lin & MacLeod, 2018; Schmidt et al., 2007). Second, while the possibility of using the category of the word distractor to predict the response might be tenable for Experiment 1A, where the response was a category name itself, applying this idea to Experiment 1B would imply that a rather complicated mechanism was in place: Participants in MC lists would have had to have used the congruent word distractor to predict the category of the picture, which would then have helped them retrieve the name of the picture (i.e., a name > category > name route). However, since congruent word distractors *are* the names of the pictures, it is unclear why following this name > category > name route would be of any benefit for performance. Finally, it has long been established that pictures are categorized faster than words are (e.g., Lupker & Katz, 1982; Smith & Magee, 1980). Therefore, using the category of word distractors to predict the category of the target pictures would be somewhat counterproductive in a speeded task. As such, adaptation to conflict frequency seems, at present, a much better explanation for the PC effects obtained here.

<u>The present results from the perspective of Bugg's (2014a) AATC hypothesis</u>

When considering the implications of the conclusion that a conflict adaptation strategy is likely responsible for the results that we obtained, one thing that is potentially important to note is that unlike classic PC manipulations in the Stroop task, Experiments 1A and 1B presented participants with a situation where learning of word-response contingencies was not an option at all, as the identity of word distractors could not be used to predict the response. Hence, conflict adaptation may have been essentially the only strategy available for dealing with conflict. Such is not the case, however, when engaging in routine activities in everyday life (e.g., driving to one's workplace). Those situations typically involve attending to a task in the face of stimuli that, re-occurring in time, become predictive of certain events (e.g., the fuel light on the car's dashboard signaling it is time to re-fuel). It is thus critical to understand how control over action is implemented in situations where a contingency learning option is available.

In response to this concern, Bugg (2014a) proposed the Associations as Antagonists to Top-Down Control (AATC) hypothesis to explain how the employment of contingency learning and conflict adaptation mechanisms is regulated. According to this hypothesis, the availability of reliable stimulus–response associations moderates the engagement of top-down mechanisms of conflict adaptation. Specifically, no adaptation to conflict frequency would take place if contingencies can be used to guide responding most of the time. Conflict adaptation would be, in other words, a last resort used by the control system only when learning contingencies – the default mode driving control engagement – is not feasible.

To provide some support for this hypothesis, Bugg (2014a) divided color-word Stroop stimuli into two sets, a "context" set and a "transfer" set, and manipulated conflict frequency and contingency learning for the context set only (transfer words were contingency-unbiased, appearing with congruent and incongruent colors an equal number of times). The transfer items were intermixed in the same list with context items which were either mostly congruent or mostly incongruent, so that transfer stimuli appeared either in a mostly congruent list (when mixed with MC context items) or in a mostly incongruent list (when mixed with MI context items). Crucially, in one version of Bugg's experiments, both MC and MI context items allowed learning of contingencies, making contingency learning a very effective strategy. In contrast, in another version of Bugg's experiments, only MC context items allowed learning of contingencies, as MI context items were constructed such that there were no contingencies to learn (i.e., each of four words was presented equally often in each of four colors). Thus, contingency learning was, overall, not a very useful strategy in this version of her experiments. Consistent with the AATC hypothesis, Bugg obtained PC effects for transfer items (i.e., evidence for a conflict adaptation strategy being applied when responding to those items) in the second version of her experiments, i.e., when learning of contingencies was possible for only a subset of the context items (i.e., the context MC items). When learning of contingencies was possible for context items in both the MC and the MI list (i.e., when such a strategy was much more useful) there was no evidence of adoption of a conflict adaptation strategy for the transfer items (for a counterargument, see Schmidt, 2014b).

Extending the AATC hypothesis to the paradigm used here, the implication for the results of Experiments 1A and 1B is straightforward: Adaptation to conflict frequency occurred not *in spite of* contingency learning being impossible, but *because* it was impossible. That is, contingency-controlled PC effects (i.e., clear markers of conflict adaptation) were obtained because there were no contingency biases at all, and adaptation to conflict frequency was the only remaining good option for maximizing performance in the task.

Implications of the present research for temporal learning accounts

In addition to examining adaptation to conflict frequency, the present research also sheds some light on the mechanism of temporal learning. This general form of learning assumes that participants in speeded tasks form temporal expectancies for emission of a response and, most critically, in Schmidt's (2013c) conceptualization, they adjust to those expectancies by speeding up on the trials in which they can produce a latency that matches the established expectancy. Although some evidence exists in favor of this mechanism (Kinoshita et al., 2011; Schmidt, 2013c, 2014a, 2016; Schmidt & Weismann, 2016), there is virtually no support for it in the present data. That is, the use of generalized linear mixed-effects models, a statistical technique that requires no transformation of the dependent variable (Lo & Andrews, 2015), failed to produce the predicted reduction of congruency effects with increasing RT on trial n – 1 in Experiments 1A and 1B. These results are consistent with those of Cohen-Shikora et al. (in press), who failed to obtain regular temporal learning effects using untransformed RTs in generalized linear mixed-effects models for a number of datasets, including Hutchison's (2011), the dataset which Schmidt first re-analyzed (with transformed RTs as the dependent variable) to make a case for temporal learning.

Because temporal learning is indexed by an interaction (i.e., that between RT on trial n – 1 and congruency on trial n), the present results and Cohen-Shikora et al.'s results raise the suspicion that temporal learning interactions reported in previously published papers (Huber-Huber & Ansorge, 2017, 2018; Kinoshita et al., 2011; Schmidt, 2013c; Schmidt & Weissman, 2016) were created by the use of nonlinear transformations of the dependent variable, an operation that is routinely performed in linear mixed-effects modelling. It is important to again note that, although these transformations do a decent job of accommodating the assumption made by linear mixed-effects models that the dependent variable be normally distributed, they affect the size and the pattern of interactions (Balota et al., 2013). Generalized linear mixed-effects models, requiring no RT transformation, provide researchers with a safer technique to search for interactions, a technique that, moving forward, is well worth considering when interactions represent the main research interest (e.g., Yang, Chen, Spinelli, & Lupker, 2019).

Another example of the present data failing to support Schmidt's (2013c) version of a temporal learning account can be found in the results of Experiment 2. In that experiment, congruent and incongruent items were replaced with easy and hard items, items not requiring the filtering out of irrelevant information as is required by interference stimuli. The results suggested that Schmidt's version of temporal learning was not at work in this situation (i.e., when vocal responding to multiple pictures is required). That is, unlike similar investigations in a button-press letter identification task utilizing low-contrast (i.e., hard) and high-contrast (i.e., easy) letters as stimuli (Schmidt, 2013c, 2014a, 2016), the proportion of easy stimuli in the list did not influence the size of the difficulty effect. As the main point of Experiment 2 was to investigate the potential contribution of temporal learning to the PC effects found in Experiments 1A and 1B, the obvious question raised by the results of Experiment 2 is whether it is possible to reconcile them with Schmidt's (2013c, 2014a, 2016) findings that manipulating frequency of difficulty does alter the magnitude of difficulty effects.

One important difference between Experiment 2 and Schmidt's (2013c, 2014a, 2016) experiments is the nature of the identification that is required (naming of multiple pictures vs. button-press identification of a limited set of letters). As mentioned above, button-press lexical decision and word naming tend to show different patterns of blocking effects, with naming showing equivalent benefits for both easy and hard items in a block containing mainly easy items, whereas button-press lexical decision typically produces an asymmetric pattern, with large benefits for easy items but not for hard items in a block containing mainly easy items (Kinoshita & Mozer, 2006). Extending this idea to proportion-easy manipulations, it is easy to see how a vocal-responding situation where easy and hard items are influenced by the frequency of difficulty in the same way will result in no proportion-easy effect, whereas a manual-responding situation where easy items are influenced by frequency of difficulty, but hard items much less so, will likely result in a proportion-easy effect.

Note that manual and vocal identification do differ in various ways. For example, manual responding generally constrains the number of responses available, whereas vocal responding, as in the present experiments, allows for multiple responses. Furthermore, a button press response requires participants to make a forced choice and commit to it, whereas a vocal response involves a gradual accumulation of evidence (e.g., Perea & Carreiras, 2003). As a result, participants might develop different subjective error estimates in the two situations. That is, their confidence in being able to give the correct response with sufficient time might not be the same.

Indeed, response confidence was the very factor that Kinoshita and Mozer (2006) held responsible for the different patterns of blocking effects observed in word naming and lexical decision tasks. In those tasks, high-frequency and low-frequency words were used as easy and hard items, respectively. Importantly, participants in a word-naming task can be assumed to be relatively confident about their response, even for hard items, but such may not be the case for the same hard items in lexical decision, for which a certain degree of uncertainty might remain even when a response is made (i.e., you know you will eventually name “glabrous” acceptably, but do you know for sure you will correctly classify it as a word or a nonword?).

The story changes, however, if the low-frequency words that are used, despite being harder than the high-frequency words, are familiar enough for participants to confidently classify as words. Using these kinds of low-frequency words, Kinoshita and Mozer (2006) obtained the pattern usually found in naming: equivalent effects for low- and high-frequency words. Kinoshita and Mozer explained these findings in term of their ASE model. Simply put, the ASE model predicts that when an item is so hard that participants may never (i.e., even if they had no time pressure) be completely confident about their response, participants will not wait extra time in pure hard, compared to mixed, blocks, as doing so will not significantly improve accuracy. As a result, they will respond before they are entirely confident in pure hard blocks so that no mixing benefit will be observed for those stimuli. When, however, hard items can still be responded to confidently given more time, it will be worth it to wait the extra time to confidently produce an accurate response, which will result in longer latencies in hard blocks and, hence, a mixing benefit.

One could certainly argue that there might be parallels between the two situations examined by Kinoshita and Mozer (2006) and the two situations created by the present Experiment 2 versus Schmidt’s (2013c, 2014a, 2016) experiments, parallels which might explain the difference between the data patterns in the latter two situations. Although it seems unlikely that participants in Experiment 2 were completely confident about their responses to all low-resolution pictures, it is important to note that participants were presented with stimuli which often had multiple acceptable responses (in fact, several of the responses marked as errors with the conservative criterion adopted here were actually fairly acceptable responses, e.g., “tool” instead of “screwdriver”, “swimming” instead of “swimmer”, etc.). In addition, since participants were not given feedback, as is typical in naming tasks, they were never informed that they were making “errors”. In turn, this inability to know when errors were being made might have led them to assume that their responses were likely acceptable and to conclude that given enough time, they would confidently respond to both easy items (i.e., high-resolution pictures) and hard items (i.e., low-resolution pictures). Therefore, the situation in the present Experiment 2 would be much more like that in a standard naming task, implying that one would expect a speed-up for both easy and hard items in the easy block.

In contrast, participants in Schmidt’s experiments were regularly given feedback, and were presented with stimuli which had only one acceptable response among a limited set of responses. Thus, participants in Schmidt’s experiments had a better idea about how well (or badly) they were performing. Therefore, it is possible that those participants were, in some cases, constantly unsure about the accuracy of their responses to hard items (i.e., low-contrast letters). In turn, this situation could have reduced the impact of frequency of difficulty selectively for hard items, as predicted by the ASE model, thus producing the pattern of blocking effects often found in lexical decision tasks, i.e., the differences in the magnitude of the difficulty effects in ME and MH lists that he observed. An examination of the role of response modality, size of the response set, and feedback in the high/low contrast letter identification paradigm would likely help shed light on the reason why the present results and Schmidt’s differ so remarkably. (note 5)

Conclusion

To conclude, the reported data make a good case for the existence of a conflict adaptation mechanism in humans. Far from being a mere illusion, such a mechanism might be an important resource in coping with tasks that require some degree of distraction suppression. While learning about *what* to respond (contingency learning) and *when* to do it (temporal learning) might be crucial aspects in goal-oriented behavior, learning *how* to respond (i.e., learning the appropriate attentional strategy to achieve the desired goal) is another human ability that should be acknowledged.

Footnotes

1. Note that while there has been some debate in the literature as to whether interference effects in the picture-word interference task reflect the same underlying cognitive processes as interference effects in the color-word Stroop task (Dell'Acqua, Job, Peressotti, & Pascali, 2007), abundant support exists for a functional equivalence of the two paradigms (Lupker, 1979; Schnur & Martin, 2012; Starreveld & La Heji, 2017; van Maanen, van Rijn, & Borst, 2009), suggesting that the picture-word interference task can afford substantially larger target, distractor, and response sets than the color-word Stroop task without otherwise altering the cognitive processes engaged in the original paradigm.
2. Proportion of congruent items was manipulated between subjects because of the limited number of items available.
3. Following a suggestion of one of the reviewers of an earlier version of this paper, we ran an additional analysis in an attempt to determine whether part of what would seem to be noise in Experiment 1A might have resulted from response speed varying across categories. Such variability could potentially have affected the temporal learning process and, consequently, PC effects. Indeed, participants were slower with the animal (930 ms), food (927 ms), and object (949 ms) categories than with the person category (830 ms), the category that also elicited the smallest overall congruency effect (5 ms vs. 44 ms, 27 ms, and 30 ms for animal, food, and object categories, respectively). However, there was no obvious relationship between the overall category latency (and/or the overall congruency effect within a category) and the size of the PC effect (i.e., the RT

difference between the congruency effect in the MC list and the congruency effect in the MI list) that the category elicited (person: 22 ms; animal: -8 ms; food: 147 ms; object: 60 ms). Do note, however, that there was, unavoidably, considerable noise in this analysis, presumably due to the fact that there were very few observations (9 or less) in some of the cells.

4. The model with RT on trial $n - 1$ as an additional predictor did not alter the pattern of results reported (i.e., there were main effects of Difficulty and List Type but no interaction between them). For the interested reader, there was a main effect of RT on trial $n - 1$ (with higher RT on trial $n - 1$ leading to longer latencies on trial n) but no interaction between Difficulty and RT on trial $n - 1$**.** None of the other interactions were significant either.
5. It is important to appreciate the fact that the present discussion rests on the assumption that a temporal learning mechanism is responsible for the pattern reported by Schmidt (2013c, 2014a, 2016) in the letter identification task. However, as recognized by Schmidt (2013c), this assumption may not be correct: If low-contrast letters are thought of as stimuli creating a relatively high level of perceptual conflict, a mechanism of adaptation to the frequency of perceptual conflict could also explain his data (e.g., if participants squint their eyes more in the list containing mostly low-contrast letters, the contrast effect will be reduced). At the same time, the results from Experiment 2 constrain this putative conflict adaptation mechanism in that they suggest that not all forms of stimulus degradation (e.g., the resolution of an image) engender a kind of perceptual conflict that people can adapt to.

Appendix:

Examining the impact of temporal learning on the congruency sequence effect

The congruency sequence effect refers to the finding that, in interference tasks, congruency effects are larger following a congruent trial than following an incongruent trial (Gratton et al., 1992). The traditional, control-based account of this effect (Botvinick et al., 2001) holds that experiencing conflict during an incongruent trial would lead participants to focus attention to the target dimension, thus reducing interference on subsequent trials; conversely, experiencing little or no conflict during a congruent trial would lead to relaxed attention, thus increasing interference on subsequent trials. Like the control account of the PC effect, this explanation has also faced some challenges: For example, in most paradigms, repetitions of stimulus features from one trial to the next seem to contribute to the congruency sequence effect (e.g., Hommel et al., 2004; Mayr et al., 2003), although a congruency sequence effect is still observed when this confound and others are removed (e.g., Schmidt & Weissman, 2014; Weissman, Jiang, & Egner, 2014).

Recently, however, Schmidt and Weissman (2016) proposed that the congruency sequence effect observed when potential confounds are accounted for is best interpreted as being the result of a temporal learning mechanism rather than the result of a conflict adaptation mechanism. This temporal learning explanation is similar to the one proposed for PC effects. Following a trial in which a fast response was emitted (typically, a congruent trial), participants will develop a relatively fast temporal expectancy which will speed up responding to a subsequent item that could be processed rapidly enough to meet that fast temporal expectancy (a situation typically occurring on a congruent trial). Because this speed-up will typically benefit congruent items but not incongruent items, the result will be an inflated congruency effect following a congruent trial, consistent with the pattern of the congruency sequence effect. Conversely, following a trial in which a slow response was emitted (typically, an incongruent trial), participants will develop a relatively slow temporal expectancy. This temporal expectancy could potentially speed up responding to a subsequent slow item that could be processed fast enough to meet that slower temporal expectancy (a situation typically occurring on an incongruent trial), although this result may not be observed in practice because, as noted, temporal expectancies often have little impact on hard-to-process stimuli (Kinoshita & Mozer, 2006; Kinoshita et al., 2011). In any case, the point is that following an incongruent stimulus, there is a potential speed-up for incongruent items. However, in comparison to what happens when the preceding response was fast, there would be no pressure to produce faster responses for fast (i.e., congruent) items. The result would be a congruency effect which should be, if anything, relatively small – again, consistent with the pattern of the congruency sequence effect. In sum, a temporal learning mechanism of this sort would seem capable of creating a pattern of results that mimics the congruency sequence effect with no need to assume a conflict adaptation mechanism.

To support their temporal learning interpretation of the congruency sequence effect, Schmidt and Weissman (2016) re-analyzed Schmidt and Weissman's (2014) data, a confound-minimized study of the congruency sequence effect in the prime-probe task, using RT on trial $n - 1$ as an index of temporal expectancy for trial n in a linear mixed-effects model analysis. They reasoned that the finding of an interaction between RT on trial $n - 1$ and congruency on trial n whereby congruency effects diminish with higher RT on trial $n - 1$ would be evidence that a temporal learning mechanism is being used. Indeed, they obtained not only such an interaction but also a reduction (although not an elimination) of the value of the beta parameter for the congruency sequence effect (i.e., the interaction between congruency on trial *n* and congruency on trial *n* – 1) in the model. Using reasoning similar to that used by Schmidt (2013c) for the PC effect, Schmidt and Weissman (2016) interpreted these results as indicating that temporal learning can generate a congruency sequence effect on its own, an idea they reinforced by successfully simulating the experimental data with an upgraded version of Schmidt's (2013c) PEP model in which temporal learning was an implemented mechanism but trial-to-trial conflict adaptation was not.

However, a fundamental problem with Schmidt and Weissman's (2016) results is that, similar to what was done by Schmidt (2013c) in the context of the PC effect, RTs were inverse-transformed to accommodate the assumption of a normally distributed dependent variable made by linear mixed-effects models. As noted, such a transformation can substantially alter the pattern of interactions and thus casts serious doubt on the interpretation of interactions, including the critical interaction between RT on trial *n* – 1 and congruency, the interaction that indexes temporal learning. In the following, we present additional analyses of Experiments 1A and 1B to examine whether the problems that emerged for the temporal learning explanation of the PC effect when a more appropriate analysis is used (i.e., generalized linear mixed-effects models with untransformed RTs; Cohen-Shikora et al., in press) also emerge when considering the congruency sequence effect.

Results

The analyses were based on the same data as those used for the analyses reported in the main text of the article, with the exception that trials for which an error was made on the preceding trial were removed from both the latency and the error analyses, as is standard for analyses of congruency sequence effects. Furthermore, in order to minimize the impact of feature and response repetitions (Hommel et al., 2004), for Experiment 1A (picture categorization) we removed the trials in which the category of the picture (and hence, the correct response) on trial *n* matched the category of the picture (and correct response) on trial *n* – 1 (e.g., the picture of a dog preceded by the picture of a cat, with both pictures requiring the response ANIMAL). The statistical models were also the same as those used for the analyses reported in the main text of the article, with the exception that Congruency on trial *n* – 1 was included as an additional fixed effect. The mean RTs and error rates for by-subject data for these analyses of Experiments 1A and 1B are presented in Tables 3 and 4, respectively.

Table 5.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for the Congruency-Sequence-Effect Analysis of Experiment 1A

	RTs			Error rates		
Congruency	MC list	MI list	List effect	MC list	MI list	List effect
Previous Congruent						
Congruent	920 (25)	931 (36)	11	1.3 (.3)	.8 (.8)	-.5
Incongruent	970 (25)	913 (37)	-57	2.7 (.7)	2 (.8)	-.7
Congruency Effect	50	-18	-68	1.4	1.2	-.2
Previous Incongruent						
Congruent	923 (24)	935 (32)	12	3.9 (.8)	2.7 (.7)	-1.2
Incongruent	928 (32)	938 (39)	10	4.3 (1.7)	2.2 (.5)	-2.1
Congruency Effect	5	3	-2	.4	-.5	-.9

Experiment 1A (picture categorization)

RT. There was a main effect of RT on trial $n - 1$ (faster responses with lower RT on trial $n - 1$), $ß = 24.01$, $SE = 3.42$, $z = 7.01$, $p < .001$, but not Congruency, $ß = -3.86$, $SE = 3.31$, $z = -1.16$, $p = .25$. The interaction between Congruency and List Type, i.e., the PC effect, was significant, $ß = -8.73$, $SE = 3.49$, $z = -2.51$, $p = .012$. Congruency on trial $n - 1$ marginally interacted with List Type, $ß = 6.75$, $SE = 3.52$, $z = 1.92$, $p = .055$, indicating that in MC lists responses tended to be overall faster when trial $n - 1$ was incongruent, a pattern that was reversed in MI lists. Most importantly, Congruency on trial $n - 1$ and Congruency did not interact, $ß = -2.63$, $SE = 3.44$, $z = -.76$, $p = .47$, although there was a marginal three-way interaction between Congruency on trial n – 1, Congruency, and List Type, $ß = -5.52$, $SE = 3.30$, $z = -1.68$, $p = .094$. As in the analysis presented in the main text of the article, there was also a three-way interaction between Congruency, List Type, and RT on trial $n - 1$, $ß = 7.29$, $SE = 3.29$, $z = 2.21$, $p = .027$.

The three-way interactions were explored by analyzing MC and MI lists separately. In MC lists, the main effects of Congruency on trial $n - 1$ (faster responses when trial $n - 1$ was incongruent), $ß = 10.46$, $SE = 5.08$, $z = 2.06$, $p = .040$, Congruency (congruent faster than incongruent), $ß = -11.96$, $SE = 5.03$, $z = -2.38$, $p = .017$, and RT on trial $n - 1$, $ß = 24.86$, $SE = 5.54$, $z = 4.49$, $p < .001$, were all significant. In addition, there was a marginal interaction between Congruency on trial $n - 1$ and Congruency, $ß = -9.36$, $SE = 5.19$, $z = -1.80$, $p = .071$. This interaction indicates a regular congruency sequence effect, with a tendency for the congruency effect to be reduced following an incongruent trial (5 ms) compared to the congruency effect following a congruent trial (50 ms). Note that this reduction occurred because responses to incongruent trials were faster when following another incongruent trial (928 ms) than when following a congruent trial (970 ms), *ß* = 39.70, *SE* = 17.71, *z* = 2.24, *p* = .025, whereas Congruency on trial *n* – 1 had no impact on congruent trials, *ß* = 2.15, *SE* = 10.38, *z* = .21, *p* = .84. Note that, as in the analysis reported in the main text of the article, Congruency and RT on trial *n* – 1 did not interact, *ß* = 5.90, *SE* = 5.49, *z* = 1.07, *p* = .28, suggesting that there was no temporal learning mechanism being used.

In MI lists, the only significant effect was RT on trial *n* – 1, *ß* = 22.29, *SE* = 4.58, *z* = 4.86, *p* < .001. In particular, there was neither an interaction between Congruency on trial *n* – 1 and Congruency, *ß* = 2.43, *SE* = 5.48, *z* = .51, *p* = .61, nor a numerical tendency for a congruency sequence effect. Similar to the analysis reported in the main text of the article, there was a tendency for congruency effects to increase with higher RT on trial *n* – 1, which is the reverse of the pattern predicted by the temporal learning account (i.e., *de*creasing congruency effects with higher RT on trial *n* – 1). However, the interaction between Congruency and RT on trial n – 1 was not significant in this analysis, *ß* = -7.17, *SE* = 5.48, *z* = -1.31, *p* = .19.

Error rates. Both Congruency, *ß* = .26, *SE* = .16, *z* = 1.67, *p* = .096, and List Type, *ß* = -.31, *SE* = .18, *z* = -1.69, *p* = .091, were marginally significant, with congruent items showing a tendency to elicit fewer errors than incongruent items and MC lists showing a tendency to elicit more errors than MI lists. Congruency on trial *n* – 1 was significant, *ß* = .42, *SE* = .16, *z* = 2.64, *p* = .008, indicating that participants were more accurate following a congruent trial (1.6%) than an incongruent trial (3.2%). No other effect reached significance.

Table 6.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for the Congruency-Sequence-Effect Analysis of Experiment 1B

	RTs			Error rates		
Congruency	MC list	MI list	List effect	MC list	MI list	List effect
Previous Congruent						
Congruent	762 (27)	791 (26)	29	1 (.3)	2.6 (2.1)	1.6
Incongruent	1111 (40)	1054 (26)	-57	15.3 (2.1)	11.4 (1.5)	-3.9
Congruency Effect	349	263	-86	14.3	8.8	-5.5
Previous Incongruent						
Congruent	770 (24)	802 (22)	32	1.3 (.5)	2 (.8)	.7
Incongruent	1083 (42)	1037 (23)	-46	8.5 (2)	11 (1.1)	2.5
Congruency Effect	313	235	-78	7.2	9	1.8

Experiment 1B (picture naming)

RT. The main effects of Congruency (congruent faster than incongruent), $\beta = -143.42$, $SE = 3.44$, $z = -41.74$, $p < .001$, Congruency on trial $n - 1$ (faster responses following an incongruent trial), $\beta = 15.27$, $SE = 3.56$, $z = 4.29$, $p < .001$, List Type (MI faster than MC), $\beta = 22.68$, $SE = 7.63$, $z = 2.97$, $p = .003$, and RT on trial $n - 1$ (faster responses with lower RT on trial $n - 1$), $\beta = 30.80$, $SE = 4.32$, $z = 7.13$, $p < .001$, were all significant. Congruency and List Type interacted, $\beta = -20.67$, $SE = 4.12$, $z = -5.02$, $p < .001$, indicating a regular PC effect. Congruency also interacted with Congruency on trial $n - 1$, $\beta = -11.79$, $SE = 3.62$, $z = -3.26$, $p = .001$. This interaction indicates a regular congruency sequence effect, with a reduced congruency effect following an incongruent trial (274 ms) than following a congruent trial (307 ms). Again, the main reason for this reduction was that responses to incongruent trials were faster when following another incongruent trial (1060 ms) than when following a congruent trial (1083 ms), $\beta = 54.11$, $SE = 11.71$, $z = 4.62$, $p < .001$. In contrast, Congruency on trial $n - 1$ had no impact on congruent trials, $\beta = 6.96$, $SE = 8.32$, $z = .84$, $p = .40$.

There was also an interaction between Congruency on trial $n - 1$ and RT on trial $n - 1$, $\beta = 11.89$, $SE = 4.64$, $z = -2.56$, $p = .010$, with lower RT on trial $n - 1$ producing a larger speed-up for responses on trial n if trial $n - 1$ was congruent than if it was incongruent, and a marginal interaction between Congruency and RT on trial $n - 1$, $\beta = -7.44$, $SE = 3.97$, $z = -1.87$, $p = .061$, with a tendency for congruency effects to increase with higher RT on trial $n - 1$. The former interaction seems consistent with the idea that fast temporal expectancies produced by easy-to-process stimuli (i.e., congruent) have a larger impact on performance than do slower temporal expectancies produced by hard-to-process stimuli (i.e., incongruent; Schmidt &

Weissman, 2016). On the other hand, the finding that congruency effects increased with higher RT on trial n – 1 reflects, once again, the reverse of the pattern predicted by the temporal learning account, according to which higher RT on trial n – 1 should *reduce* congruency effects.

Error rates. Congruency (congruent more accurate than incongruent) was the only significant effect, $ß = 1.28$, $SE = .11$, $z = 11.28$, $p < .001$.

Conclusion

Similar to what was found for the PC effect, temporal learning does not seem to provide a convincing explanation for the congruency sequence effect in the present dataset. According to the temporal learning account, a congruency sequence effect should emerge as a consequence of a mechanism whereby congruency effects decrease with higher RT on trial n – 1. However, in an analysis in which untransformed RTs were used (thus avoiding potential problems associated with nonlinear transformations of the dependent variable), we found, if anything, marginal evidence for the opposite pattern (i.e., congruency effects increasing with higher RT on trial n – 1) in Experiment 1B and a numerical tendency in the same direction in the MI list of Experiment 1A. While this situation suggests that no temporal learning mechanism was being used, a regular congruency sequence effect emerged nonetheless.

It is worth noting that in the present analyses, not only temporal learning was controlled for but also feature and response repetitions were either removed (Experiment 1A) or minimal to begin with (i.e., there were no response repetitions in Experiment 1B because each trial required a different response). Therefore, we are inclined to interpret the congruency sequence effect that was obtained as resulting from a trial-to-trial conflict adaptation mechanism

(Botvinick et al., 2001), with recent experience with conflict leading to focused attention, thereby decreasing interference, and recent experience with little or no conflict leading to relaxed attention, thereby increasing interference. This explanation would be consistent with the finding that the congruency sequence effect was mainly caused by facilitation for incongruent trials following another incongruent trial, a pattern that would reflect reduced interference when interference has recently been dealt with. This explanation would also seem to accommodate the fact that in the MI list in Experiment 1A, no congruency sequence effect was obtained. The high number of incongruent trials produced a complete elimination of the congruency effect in that list, suggesting that there was little conflict to adapt to. Indeed, it seems reasonable to assume that some amount of conflict is necessary for a trial-to-trial conflict adaptation mechanism to be operable. The core claim, in any case, is that not only the PC effect but also the congruency sequence effect, another important marker of conflict adaptation, emerges when potential confounds are accounted for, and most importantly, temporal learning does not seem to offer a convincing alternative to a control-based interpretation of this effect.

Chapter 2.5: Interim Summary

The experiments reported in Chapter 2 produced a clear result: Contrary to the idea that PC effects are fully explained by non-conflict learning confounds that PC paradigms typically contain (Schmidt 2013b, 2019), a list-wide PC effect was observed even in situations in which those confounds could not have produced that effect. To create this type of situation, a picture-word interference task was used in which picture targets and word distractors were never repeated, making it impossible for participants to learn word-response contingencies in either the MC list or the MI list (and, by implication, making the stimuli in both lists equally uninformative as no word in the experiment could be used to predict the response). In both a task in which the pictures required a naming response and a task in which the pictures required a categorization response, a list-wide PC effect emerged despite the magnitude of the basic congruency effect in the two tasks varying considerably (larger in the picture naming task than in the picture categorization task). The implication of these results is that adaptation to list-wide conflict frequency may exist independently from non-conflict learning processes related to contingency learning and stimulus informativeness.

An examination of the potential impact of another non-conflict learning process, temporal learning, did not alter these conclusions. First, including an index of temporal expectancies in the analyses revealed that, in fact, no temporal learning process of the sort proposed by Schmidt (2013c) was engaged in either of the two picture-word interference experiments. Evidence in favor of a temporal learning process also failed to emerge in an additional experiment in which the difficulty of naming pictures did not derive from the conflict caused by distracting information but rather from the perceptual quality of the pictures. According to

Schmidt's temporal learning account, this situation should have produced a larger difficulty effect in the list in which easy-to-name pictures were more frequent than in the list in which hard-to-name pictures were more frequent. However, difficulty effects were the same size in the two lists, although latencies were overall higher in the list with a majority of hard-to-name pictures, a result potentially consistent with other temporal learning accounts (accounts that, unlike Schmidt's, do not pose a challenge for control-based interpretations of PC effects: Lupker et al., 1997, 2003). In sum, both methods of examining temporal learning provided no evidence that a temporal learning process could produce, or even contribute to the production of, the obtained PC effects in the picture-word interference tasks. Those results would thus seem to make a strong case that non-conflict learning processes are not the whole story in the list-wide PC effect, an effect that would instead appear to reflect adaptation to list-wide conflict frequency, at least in the situations examined.

Indeed, because those results were obtained in the context of a picture-word interference task, a valid concern that could be raised is whether those results would directly translate back to the color-word Stroop task from which the debate concerning the nature of PC effects originated. Although picture-word and color-word versions of the Stroop task have been demonstrated to be functionally equivalent in a number of studies, the picture-word interference task used in Chapter 2 is not easily replicable using colors as targets because typical participants can spontaneously name only a limited set of colors. This concern was addressed in Chapter 3 by using a different approach in an attempt to replicate the results from Chapter 2 in the color-word Stroop task. Rather than preventing participants from learning contingencies by removing repetitions from the materials, a design was adopted that allowed a manipulation of conflict frequency using repeated color targets and word distractors but without contingency learning and stimulus informativeness confounds.

As noted in Chapter 1, a typical problem arising when constructing a list-wide PC manipulation is that contingency learning is always possible in an MC list as this list will inevitably contain at least some words for which the congruent color is the high-contingency color. As a result, this list will also typically be relatively informative. Because in the color-word Stroop task, as noted, the limited number of nameable colors means that contingency learning and stimulus informativeness confounds cannot be eliminated by having nonrepeated stimuli in the experiment, negating an impact of these confounds requires a different strategy. The strategy used in Chapter 3 to accomplish this goal involved manipulating the proportion of neutral and incongruent, as opposed to congruent and incongruent, items – a list-wide Proportion-Neutral (PN) paradigm. Notably, as in the standard list-wide PC paradigm, it is still the case in this paradigm that conflict is more frequent in one list (the Mostly Incongruent [MI] list in which incongruent items are frequent and neutral items are infrequent) than in the other (the Mostly Neutral [MN] list in which neutral items are frequent and incongruent items are infrequent). As such, use of a process of adaptation to list-wide conflict frequency would imply that Stroop interferences effects (i.e., the color-naming difference between incongruent and neutral trials) should be larger when conflict is infrequent (in the MN list) than when conflict is frequent (in the MI list) because participants' attention to task-relevant information should be relaxed in the former situation and more focused in the latter. That is, a Proportion-Neutral (PN) effect, similar to the list-wide PC effect in the standard paradigm, should be obtained.

The reason that neutral items are especially useful in this context is that they not only permit the manipulation of conflict frequency, as in the standard list-wide PC paradigm, but they also permit the complete elimination of word-response contingencies in the task as a whole. In Chapter 3, both an MI list (with frequent incongruent items and infrequent neutral items) and an MN list (with frequent neutral items and infrequent incongruent items) were constructed in which contingency learning was impossible because each word appeared in two or more colors equally frequently (as can be the case for MI lists in the standard list-wide PC paradigm but can never be the case in an MC list in the standard list-wide PC paradigm). The implication of this design, in terms of stimulus informativeness as defined by Schmidt (2014b, 2019), is that no word was informative (i.e., no word could be used to predict the color response) in either the MI list or the MN list, thus removing the stimulus informativeness confound present in some list-wide PC manipulations (Bugg, 2014a). As explained in more detail in Chapter 3, this manipulation effectively avoided another non-conflict learning confound that arises in list-wide PC paradigms such as Bugg's, a confound relative to the strength of the correlation between colors and words in MC vs. MI lists that is closely associated, although not identical, with that of stimulus informativeness (Melara & Algom, 2003; see Chapter 1, footnote 6).

In addition, the experiment in Chapter 3 was constructed so that the potential impact of adaptation to the frequency of conflict relative to a specific stimulus feature (specifically, adaptation to color-specific conflict frequency: Bugg & Hutchison, 2013) was dissociated from the potential impact of adaptation to the frequency of conflict in the list as a whole. To achieve this goal, similar to recent research in the list-wide PC paradigm (Blais & Bunge, 2010; Bugg et al., 2008; Bugg, 2014a; Hutchison, 2011), the experiment included both context items for which both list-wide and color-specific conflict adaptation could occur, and transfer items for which only list-wide conflict adaptation could occur, making those items crucial for the question of whether adaptation to list-wide conflict frequency exists in the Stroop task.

Finally, note that in Chapter 3 compared to Chapter 2, relatively less importance was given to temporal learning as a potential explanation for the list-wide PN effect obtained because both the results of Chapter 2 and the recent re-analyses reported by Cohen-Shikora et al. (2018) make a strong case that the temporal learning process proposed by Schmidt (2013c) is quite unlikely to be a convincing explanation for the list-wide PC effect after all. Nonetheless, an analysis of the RT data was conducted using the same technique used in Chapter 2 to control for temporal learning (a generalized linear mixed model with an index for temporal expectancies included as a predictor).

Chapter 3:

Proactive Control in the Stroop Task: A Conflict frequency Manipulation Free of Item-Specific, Contingency Learning, and Color-Word Correlation Confounds

Introduction

The list-wide Proportion-Congruent effect: A marker of proactive control?

A question that has received increasing research interest in the last decade is whether the expectation of conflict between task-irrelevant and task-relevant information can induce individuals to adjust attention between those sources of information (Schmidt, 2013b). This putative conflict adaptation mechanism would bias attention toward task-relevant information when conflict is expected, but not when conflict is not expected (Botvinick, Braver, Barch, Carter, & Cohen, 2001). A typical example of those situations is represented by manipulations of conflict frequency in the Stroop (1935) task.

In the Stroop task, participants name the ink color of a letter string, typically a color word, which can be congruent with the color (e.g., the word RED in red color), incongruent with the color (e.g., the word BLUE in red), or neutral (e.g., the consonant string XXX in red). The typical result is a congruency effect, with faster and/or more accurate responses to congruent than to incongruent items. This effect usually results from interference from incongruent items (typically producing much slower latencies than neutral items) combined with some facilitation from congruent items (typically producing slightly faster latencies than neutral items) (note, however, that the relative magnitudes of interference and facilitation depend on a host of factors, e.g., the nature of the neutral items used: MacLeod, 1991; see also Sabri, Melara, & Algom, 2001).

What is important for the present discussion is that the magnitude of the congruency effect varies as a function of conflict frequency, a situation typically examined by using Proportion-Congruent (PC) manipulations. In the standard (list-wide) PC manipulation, performance in a list in which conflict is infrequent (when congruent items are more frequent than incongruent items, i.e., a Mostly Congruent [MC] list) is compared with performance in a list in which conflict is frequent (when incongruent items are more frequent than congruent items, i.e., a Mostly Incongruent [MI] list). The typical result is that the congruency effect is larger in the MC list than in the MI list (e.g., Logan & Zbrodoff, 1979).

This finding, known as the PC effect, has traditionally been interpreted as evidence for the use of a process of adaptation to list-wide conflict frequency. According to this explanation, a control mechanism exists that monitors conflict and adapts attention accordingly (Botvinick et al., 2001). Specifically, a situation in which task-irrelevant information (i.e., the word) is frequently conflicting (i.e., an MI list) will cause the emission of a signal indicating a need for more focused attention to task-relevant information (i.e., the color). Interference from irrelevant information will thus be minimized, producing a small congruency effect. In contrast, when conflict from task-irrelevant information is infrequent (i.e., in an MC list), attention will be relaxed because there is less reason to increase control. Thus, interference from irrelevant information on the rare incongruent items in an MC list will be especially problematic to overcome, producing a large congruency effect.

According to one of the control models often used to interpret the PC effect, the conflict-monitoring model (Botvinick et al., 2001), this effect is the result of transient control, the type of control that is also observed in trial-to-trial modulations of congruency effects (i.e., reduced congruency effects following incongruent than congruent items: Gratton, Coles, & Donchin, 1992). Because in an MI list, but not in an MC list, conflict often accumulates over the course of the experiment, this transient control would typically lead to a tightening of attention to task-relevant information, resulting in a reduced congruency effect (see also Jiménez & Méndez, 2013; 2014).

More recent research, however, has suggested that the PC effect is dissociable from trial-to-trial control adjustments (e.g., Torres-Quesada, Funes, & Lupiáñez, 2013; Torres-Quesada, Milliken, Lupiáñez, & Funes, 2014) and that control, in general, appears to operate at both short and longer time scales (e.g., Braver, 2012; Braver, Gray, & Burgess, 2007; De Pisapia & Braver, 2006; Kane & Engle, 2003; Jiang, Heller, & Egner, 2014). For example, in the Dual-Mechanism of Control framework proposed by Braver (2012; Braver et al., 2007), an MI list would favor a proactive mode of control that minimizes interference from the word by maintaining the color-naming goal to the extent possible. An MC list, on the other hand, would favor a more transient, reactive mode of control whereby the color-naming goal is often neglected and is only retrieved upon presentation of the infrequent incongruent stimuli (De Pisapia & Braver, 2006; see also Kane & Engle, 2003). Thus, although the distinction between proactive and reactive control is likely blurrier than this explanation suggests (see, e.g., Aben, Verguts, & Van Den Bussche, 2017), the PC effect (more precisely, the reduced congruency effect in an MI list) would result from a proactive form of control, in the sense that this control is applied *before* any specific item appears (e.g., Bugg, 2014a; Gonthier, Braver, & Bugg, 2016; see also Verguts and Notebaert's, 2008, notion of "nonspecific" conflict adaptation). (note 1)

In recent years, however, the idea that the PC effect results from the implementation of a proactive process in high-conflict situations has received considerable criticism (Blais & Bunge, 2010; Blais, Robidoux, Risko, & Besner, 2007; Schmidt, 2013b, 2019). This criticism stems from the realization that a proactive control process is not necessary for generating a PC effect, as this effect could also result from alternative processes that PC paradigms typically allow. These processes, described below, include learning to associate items to responses and/or control states, and learning to adjust attention based on how informative (rather than how conflicting) items in the task are. (note 2) In the present research, we aimed to demonstrate that when all of these alternative processes are accounted for, PC effects can still be observed, providing evidence for the existence of proactive adaptation to conflict frequency.

Reactive accounts of the list-wide Proportion-Congruent effect

The first challenge to an account of the PC effect based on proactive control came from Jacoby, Lindsay, and Hessels's (2003) report of an "item-specific" PC effect. Jacoby et al. designed a new version of the PC paradigm in which half of the words were mainly presented in their congruent color (Mostly Congruent [MC] items) and the other half were mainly presented in an incongruent color (Mostly Incongruent [MI] items), with all words intermixed in a single list. Similar to the list-wide PC effect, an item-specific PC effect emerged, with MC items eliciting a larger congruency effect than MI items.

Because in Jacoby et al.'s (2003) paradigm congruent and incongruent items were equally probable in the list as a whole, the PC effect that they obtained could not have been produced by a proactive process based on list-wide conflict frequency (i.e., a process that is applied *before* any specific item appears). Instead, it must have been produced by a reactive process that is initiated *after* an item is presented in response to the nature of that specific item. This type of reactive process could take two basic forms. First, it might be a conflict adaptation process in which recognition of specific stimuli regulates the recruitment of appropriate control processes (e.g., Bugg & Hutchison, 2013; Gonthier et al., 2016). Specifically, the recognition of an MI word, e.g., RED, would favor focused attention to the color, producing reduced interference. On the other hand, the recognition of an MC word, e.g., GREEN, would favor relaxation of attention, a process that would encourage processing of the word, thus producing a large interference effect when the MC word does conflict with the color. Alternatively, the process producing the item-specific PC effect might be one whereby a contingency is learned between a word and the response typically made to that word (Schmidt & Besner, 2008). For example, if the MI word RED appears most often in blue, individuals will use that word to predict a blue response, with the result being that blue responses will be produced relatively rapidly even though the word itself is RED, leading to a reduced congruency effect. Conversely, individuals will use the MC word GREEN to predict the (congruent) green response. Hence, latencies will speed up for these congruent stimuli, producing an increased congruency effect.

Whether a conflict adaptation or a contingency learning process is responsible for the item-specific PC effect (for discussions, see, e.g., Bugg & Hutchison, 2013; Schmidt, 2013b, 2019), the question this effect raises is whether either of these reactive strategies might also explain the list-wide PC effect. The reason this question is relevant is that, in the list-wide PC manipulation, all items appearing in a MC list are MC items (i.e., all words appear most often in a congruent color) and all items appearing in a MI list are MI items (i.e., all words appear most often in one or more incongruent colors). Thus, it is possible that the list-wide PC effect is not produced by a proactive process dependent on list-wide conflict frequency, but by whatever reactive process produces the difference between MC and MI items in the item-specific PC paradigm (for a demonstration of this possibility within the framework of the conflict-monitoring model, see, e.g., Blais et al., 2007).

To address this question, Bugg (2014a; see also Blais & Bunge, 2010; Bugg, Jacoby, & Toth, 2008) developed a new list-wide PC manipulation which allows for a dissociation of a proactive control process from reactive processes. Bugg divided the items into two sets, referred to as the “context” set and the “transfer” set, and manipulated congruency proportion for the context set only. The transfer items were 50:50 congruent/incongruent and were intermixed in a list with either mostly congruent context items (creating an overall MC list) or mostly incongruent context items (creating an overall MI list). Note that in this type of manipulation, from the participants’ perspective, there is no obvious separation between the two sets of stimuli used in the task, nor are the participants informed about their existence. However, using context and transfer stimulus sets provides the researcher with a meaningful tool to infer the processes normally involved in list-wide PC manipulations. Specifically, the rationale is that while a PC effect obtained with the context items might result from any of multiple processes, the only possible explanation for a PC effect on the transfer items would be adaptation to list-wide conflict frequency.

Indeed, in addition to the expected PC effect for context items in all situations (which, as just noted, is compatible with multiple explanations), a PC effect for transfer items did emerge in one of the situations that Bugg (2014a) examined. Specifically, a PC effect with transfer items emerged when the MI list was a list in which contingency learning was impossible for the MI context items due to the fact that the context words appeared in four equally probable colors (one congruent and three incongruent) so that no contingencies for those words could be learned. (Note that contingency learning is always possible for MC context items, as the congruent color is, unavoidably, also the more probable color). That is, only in this circumstance did transfer items show a smaller congruency effect in the MI list than in the associated MC list. In contrast, when contingency learning was possible for context items in the MI list (i.e., when each of the context words appeared more frequently in one specific incongruent color), no PC effect was obtained on the transfer items (i.e., the congruency effects on the transfer items were the same size in the MC and MI lists).

To explain these results, Bugg (2014a) suggested that adaptation to list-wide conflict frequency is possible, but its usage “will primarily be evident when one cannot rely on use of [word-response contingencies] to guide responding on most trials in an effort to achieve task goals (i.e., minimization of Stroop interference)” (p. 568). In such situations, e.g., in the MI list that did not allow for contingency learning, a conflict adaptation process, that is, a process involving an explicit focus of attention on the color, is the process being used. The result is a congruency effect for the items in that list (including the transfer items) that is smaller than that in the MC list (in which a contingency learning process potentially is being used), producing a PC effect. On the other hand, when reliable contingencies exist in the MI list, no conflict adaptation process is engaged in either list. Rather, contingency learning is the only process being engaged in both MC and MI lists. As a result, the transfer items will be unaffected, causing them to produce the same size congruency effect in the two lists.

The role of stimulus informativeness and color-word correlations

Bugg's (2014) results would seem to make a clear case for the existence of a process of adaptation to list-wide conflict frequency, at least in certain circumstances. Recently, however, Schmidt (2019) has argued that an alternative interpretation for Bugg's results is possible, one that does not involve a role for adaptation to list-wide conflict frequency. According to this explanation, the PC effect on transfer items reported by Bugg when the context MI items did not have a more probable incongruent color resulted from a process in which attention to task-relevant and task-irrelevant information *is* adapted, however, this adaptation is based on what he termed "stimulus informativeness" rather than conflict frequency.

According to Schmidt (2019), this term refers to the degree to which words in a list allow learning of word-response contingencies. In an MC list, words would be relatively informative because contingencies can be learned for at least some of the words (i.e., the context words). Thus, attention to words (including transfer words) would be enhanced in that list. Because they are attended, transfer words will produce more interference, leading to a large congruency effect. In contrast, because words are relatively uninformative in an MI list if no contingencies exist for the context items, attention to words (including transfer words) will be reduced. Transfer words will thus produce less interference in this situation, leading to a reduced congruency effect. Note that this account would also explain why no PC effect is obtained for transfer items when contingencies can be learned for both MC and MI context items. The reason is that context words would be informative in both situations. Thus, attention would be directed to (all) words in both MC and MI lists.

Schmidt’s (2019) stimulus informativeness account echoes an idea proposed previously by Algom and collaborators (Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003; Sabri et al., 2001), that attention to the task-irrelevant dimension is increased when there is a relationship between the values on the task-relevant dimension (e.g., the colors) and the values on the task-irrelevant dimension (e.g., the words). That is, when the words in the Stroop task provide information about the colors that they appear in, the word dimension will receive more attention than in a situation in which the words and the colors are randomly paired (a zero-correlation situation). The strength of the relationship between words and colors (a measure of stimulus informativeness related to, but distinct from, that of Schmidt) can be expressed as a chi-squared based correlation (C), which takes on positive values when the conditional probability of congruent stimuli is relatively large, and negative values when the conditional probability of incongruent stimuli is relatively large (a value of zero would correspond to no correlation; Melara & Algom, 2003).

As Bugg (2014a) noted, in the situation in which she obtained a PC effect on the transfer items (the situation in which contingencies could not be learned for the context words in the MI list), words and colors were always more strongly associated in the MC list than in the MI list (the absolute value of C was higher for the MC list than for the MI list). In contrast, in the situation in which Bugg failed to obtain a PC effect on the transfer items (the situation in which contingencies could be learned for the context words in the MI list), the strength of the relationship between words and colors in the two lists was the same (the absolute value of C for the MC list was the same as that for the MI list). As such, similar to Schmidt's (2019) stimulus informativeness argument, the hypothesis could be entertained that the real reason that Bugg obtained a PC effect in the former situation has to do with the fact that the two lists in that situation differed in the strength of the color-word correlation (whereas they did not in the situation in which no PC effect on the transfer items was obtained). Specifically, the stronger color-word correlation in the MC list would have drawn attention to the word dimension to a larger extent than the (weaker) correlation in the MI list. As a result, a larger congruency effect would have been obtained in the MC list than in the MI list even if no process of adaptation to list-wide conflict frequency was in place.

In discussing her results, Bugg (2014a) considered this account implausible because 1) in all of the situations she examined, the color-word correlation was high ($|C| > .76$), and 2)in the situations in which the strengths of the color-word correlations for the MC list and the MI list did differ, they did not differ greatly (the difference between $|C|$ for the MC list and $|C|$ for the MI list was less than .1). Thus, the claim that differences in the strength of color-word correlations between MC lists and MI lists determined the PC effect obtained in those situations would have to be based on the assumption that participants are sensitive to very small differences of that sort even in the presence of overall high correlations, an idea that appeared unlikely. Even so, the fact remains that Bugg obtained a PC effect on the transfer items when list-wide conflict frequency and strength of color-word correlations were confounded, but did not obtain one when the strength of those correlations was matched across the MC and MI lists. Therefore, in spite of Bugg's claim that a PC effect on transfer items was produced by adaptation to list-wide conflict frequency, alternative accounts related to the general informativeness of stimuli in the list (either defined as the possibility of learning word-response contingencies or as the absolute strength of the color-word correlation) exist that could potentially explain that effect.

<u>The present research</u>

To address the issues described above that hinder the interpretation of Bugg's (2014) results, in the present research we further modified the paradigm developed by Bugg so that a modulation of Stroop interference for transfer items as a result of list-wide conflict frequency could be explained by neither reactive processes (e.g., contingency learning) nor adaptation to stimulus informativeness and/or the strength of color-word correlations, with the only remaining explanation being proactive control.

We achieved this goal by using a Proportion-Neutral (PN), rather than a Proportion-Congruent, manipulation, that is, a paradigm in which the proportion of incongruent and neutral items (e.g., XXX, instead of congruent items, which were not used) in the context set is manipulated to create Mostly Neutral (MN) and Mostly Incongruent (MI) lists. We then evaluated whether interference effects (the color-naming difference between incongruent and neutral items, i.e., consonant strings) on transfer items would be affected by the PN manipulation.

Note that, from the perspective of Botvinick et al.'s (2001) model, this change should be uninfluential, as what is critical for proactive control engagement is the frequency of conflict elicited by incongruent items, items that are more frequent in the MI list than in the MN list. Furthermore, Tzelgov, Henik, and Berger (1992) already demonstrated that, similar to the PC effect obtained in the typical PC paradigm, increasing the proportion of neutral items in a list leads to an increase in interference but not in facilitation (i.e., the latency difference between neutral and congruent items), a pattern Botvinick et al. (2001) were able to simulate in their model.

The reason that neutral items are especially useful in the present circumstances is that negating any impact of stimulus informativeness (Schmidt, 2019) is impossible when the proportion of *congruent* (instead of neutral) and incongruent items is manipulated (i.e., in the standard PC paradigm), because, as noted, the MC list (but not necessarily the MI list) inevitably contains informative words, i.e., words for which contingencies can be learned. Therefore, contingencies can always be learned for context MC words in a design like Bugg's (2014). In contrast, neutral items allowed a manipulation of list-wide conflict frequency in a situation in which, similar to Bugg's experiment that produced the PC effect, contingencies cannot be learned in either the MN list (i.e., the list in which conflict is infrequent, similar to the MC list in the standard paradigm) or the MI list. In this situation, the words appearing in those lists are equally uninformative (in the sense conveyed by Schmidt's definition of informativeness), thus eliminating the stimulus informativeness confound present in Bugg's experiments.

In addition, in our experiment, the absolute strength of the color-word correlation was the same in the MN list and the MI list. Note that, because congruent items were not used and, for the reasons described below, the stimuli were divided into two non-overlapping sets (leaving most possible color-word combinations unused), this correlation was inevitably strong (as it was in Bugg's (2014) experiments). Specifically, all that participants could learn was that each word appeared in a specific set of colors (even though the colors that the word appeared in were equally probable) but not in other colors. Because words could be used to anticipate the colors that they would appear in, this situation might have induced participants in both lists to attend to the task-irrelevant dimension considerably more than they would have done in a zero-correlation situation, i.e., a situation in which colors and words are paired randomly. As a result of receiving attention, incongruent words would elicit much more interference in this type of situation than in the zero-correlation situation (Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003; Sabri et al., 2001). Note, however, that although the basic Stroop interference effect in the present experiment could have been inflated because of the strong word-color correlation that we introduced, what was crucial for the present purposes was that this correlation had the same strength in the MN list and the MI list (i.e., that C had the same absolute value in the two lists). The reason is that what we were interested in was the modulation of interference based on list-wide conflict frequency, not the interference effect itself. Because words and colors had an equivalent strength of association in the MN list and the MI list, attention to the word dimension induced by this correlation could not explain any difference in the magnitude of interference across the two lists.

Although removing contingencies from the design can solve the problems of stimulus informativeness and of color-word correlation strength, it does not prevent use of a different type of reactive process, i.e., adaptation to color-specific conflict frequency. Bugg and Hutchison (2013) showed that in addition to learning associations between words and conflict frequency, individuals can also learn associations between *colors* and conflict frequency, as demonstrated by the fact that MC colors (e.g., the color red appearing often with the word RED) elicit larger congruency effects than MI colors (e.g., the color green appearing often with incongruent words). To prevent this potential color-specific effect in the transfer items, the colors used in the context versus transfer item sets were nonoverlapping. Of necessity, context items appeared only in MN colors in the MN list and only in MI colors in the MI list. In contrast, transfer items appeared in a different set of colors and those colors appeared on neutral versus incongruent transfer trials equally often in the two lists. Thus, while a PN effect (i.e., larger interference in the MN list than in the MI list) on the context items would be compatible with either proactive adaptation to list-wide conflict frequency or reactive adaptation to color-specific conflict frequency, a PN effect on the transfer items would only be compatible with the former process (adaptation to list-wide conflict frequency).

Finally, it is worth mentioning one study in the literature that may cast doubt on the effectiveness of a conflict frequency manipulation comparing MN (instead of MC) and MI lists. Bugg, McDaniel, Scullin and Braver (2011) examined performance on neutral items in a manipulation somewhat similar to ours and did not find any difference between performance on those items in an MI list versus an MN list. In Bugg et al.'s experiment, participants completed an MN list in addition to an MC list and an MI list. All list types contained a fixed set of neutral items (in their case, color-unrelated words, e.g., RABBIT) which functioned as the transfer items. That is, in Bugg et al.'s experiment, transfer items only included neutral items (as opposed to neutral and incongruent items, as in the present experiment). Bugg et al. reasoned that naming the color of those neutral items should be slower when those items appear in an MC list than when they appear in either an MN list or an MI list because inhibiting the automatic tendency for word reading should be harder in a situation that favors word reading (i.e., in MC lists in which reading the word would often result in the correct response) than in situations that do not favor word reading (i.e., in MI and MN lists, lists in which reading the word would often result in an incorrect response). Consistent with these hypotheses, latencies on the transfer neutral items were slowest in the MC list but there were no differences between the MI list and the MN list.

Although these results do suggest that MI and MN lists may be similar in that they do not encourage word reading (unlike MC lists), they do not allow the conclusion that those lists are dealt with using exactly the same process(es). Specifically, it is possible that MI and MN lists would differ in the processes engaged to deal with Stroop interference (interference that is typically caused by incongruent color words: MacLeod, 1991), an aspect that Bugg et al. (2011) could not evaluate because their transfer items did not include incongruent items. Thus, although an MN list (and an MI list) would not encourage word reading to the same extent that an MC list does, an MN list may lead individuals to relax their attention somewhat because words in that list rarely cause substantial interference with color naming. The same would not be true for an MI list because, in that list, dealing with frequent incongruent items would induce more focused attention to the color dimension. Thus, if such a mechanism of adaptation to list-wide conflict frequency were used, the latency difference between incongruent and neutral items should be reduced in an MI list compared to an MN list. In other words, a PN effect, similar to the PC effect in the standard PC paradigm, would have been expected if Bugg et al.'s transfer set had contained incongruent items.

Method

Participants

Eighty students at the University of Western Ontario (age 17–29 years) participated for course credit or $10. We did not conduct a power analysis to determine this sample size. Instead, we determined the sample size based on a pilot experiment conducted in our lab examining a PN effect. Because the PN effect in that experiment could have been due to either or both of two processes (adaptation to list-wide conflict frequency and/or adaptation to color-specific conflict frequency), we decided to double the sample size tested in that experiment (N = 40) for the present experiment, an experiment in which the process of adaptation to list-wide conflict frequency was isolated (for transfer items). All participants were native English speakers and had normal or corrected-to-normal vision.

Materials

Six color names (RED, YELLOW, BLACK, BLUE, GREEN, WHITE) and six neutral "words" of matching lengths (XXX, ZZZZZZ, KKKKK, QQQQ, JJJJJ, HHHHH) were used as distractors, and six colors (red [R: 255; G: 0; B: 0], yellow [R: 255; G: 255; B: 0], black [R: 0; G: 0; B: 0], blue [R: 0; G: 112; B: 192], green [R: 0; G: 176; B: 80], and white [R: 255; G: 255; B: 255], corresponding to "red", "yellow", "black", "blue", "green" and "white" in the standard DMDX palette) were used as targets. For the neutral words, we used consonant strings instead of color-unrelated words because it is known that any readable stimulus can create some degree of interference in the Stroop task (e.g., Dalrymple-Alford, 1972; Klein, 1964). As a result, color-unrelated words may not create the strongest contrast with incongruent words in terms of interference. For example, in Bugg et al.'s (2011) study in which color-unrelated words were used as neutral items, Stroop interference (incongruent – neutral) was, if anything, smaller than Stroop facilitation (neutral – congruent). In contrast, much larger Stroop interference than facilitation is routinely observed when neutral items are consonant strings (MacLeod, 1991). Thus, we reasoned that manipulating the frequency of the conflict elicited by incongruent items may be more effective in a situation in which the incongruent items are compared to neutral items, items that produce little, if any, conflict (i.e., consonant strings), than when the incongruent items are compared to neutral items that produce some conflict (i.e., color-unrelated words). (note 3)

The frequency of word-color combinations is represented in Tables 1 and 2 for the MN and the MI list, respectively. The "words" were divided into two sets, one set (RED, YELLOW, BLACK, XXX, ZZZZZZ, KKKKK) appeared only in red, yellow, and black whereas the other set (BLUE, GREEN, WHITE, QQQQ, JJJJJ, HHHHH) appeared only in blue, green, and white. Each word in each set appeared equally often in two of the three colors (for the color words, neither of these colors was the congruent one). One set (e.g., the words appearing in red, yellow, and black) served as the context set and the other set (e.g., the words appearing in blue, green, and white) served as the transfer set. In the MN list, the colors in the context set appeared 84 times with a neutral word and 12 times with an incongruent word (the color-specific Proportion of Neutral Items [PNI] was thus 87.5%). The reverse mapping was used in the MI list (such that color-specific PNI = 12.5%).

Table 1.

Template for the Frequency of Color-Word Combinations in the MN List

	Word												
Set	Color	RED	YELLOW	BLACK	BLUE	GREEN	WHITE	XXX	ZZZZZZ	HHHHH	QQQQ	JJJJJ	KKKKK
Context	Red		2	2					14	14			
	Yellow	2		2				14		14			
	Black	2	2					14	14				
Transfer	Blue					8	8					8	8
	Green				8		8				8		8
	White				8	8					8	8	

Table 2.

Template for the Frequency of Color-Word Combinations in the MI List

	Word												
Set	Color	RED	YELLOW	BLACK	BLUE	GREEN	WHITE	XXX	ZZZZZZ	HHHHH	QQQQ	JJJJJ	KKKKK
Context	Red		14	14					2	2			
	Yellow	14		14				2		2			
	Black	14	14					2	2				
Transfer	Blue					8	8					8	8
	Green				8		8				8		8
	White				8	8					8	8	

The colors in the transfer set appeared 48 times with a neutral word and 48 times with an incongruent word (color-specific PNI = 50%) in both lists. In both lists, the context set and the transfer set were randomly intermixed. Overall, there were 132 neutral items and 60 incongruent items in the MN list (list-wide PNI = 68.75%), with those numbers reversing in the MI list (list-wide PNI = 31.25%), for a total of 192 items in each list. The assignment of the two sets to context and transfer items was counterbalanced across participants, as was the order with which the MN and MI lists were presented. Finally, for each list, we calculated the contingency coefficient measuring the strength of the color-word correlation, C, using Melara and Algom's (2003) formula (with the exception that C was allowed to take on positive values when the conditional probability of neutral, rather than congruent, stimuli was relatively large, i.e., in the MN list). C was .82 for the MN list and -.82 for the MI list (the absolute strength of color-word correlations was thus the same across lists).

Procedure

Each trial began with a fixation symbol ("+") displayed for 250 ms in the center of the screen followed by a colored word 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. Stimuli were presented in uppercase Courier New font, pt. 14, against a medium grey background (R: 169; G: 169; B: 169). No feedback was provided. There was a self-paced pause between the two lists. The order of trials within each list was randomized. Initially, participants performed a practice session including 6 neutral and 6 incongruent trials. 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 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 (accounting for 2.0% of the data points) were discarded. A 2 (Item Type: Neutral vs. Incongruent) X 2 (List Type: Mostly neutral vs. Mostly incongruent) ANOVA was conducted on both latencies and errors for context items and transfer items separately, paralleling the analyses in previous research using this paradigm (Bugg, 2014a; Bugg et al., 2008). The mean RTs and error rates are presented in Table 3. The raw data and SPSS script used for the analyses are publicly available at https://osf.io/yk57z/.

Table 3.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Context and Transfer Items

	RTs			Error rates		
Item type	MN list	MI list	List effect	MN list	MI list	List effect
Context items						
Neutral	724 (13)	742 (12)	18	1.4 (.2)	2.5 (.5)	1.1
Incongruent	864 (17)	827 (15)	-37	4.9 (.9)	2.8 (.3)	-2.1
Interference Effect	140	85	-55	3.5	.3	-3.2
Transfer items						
Neutral	732 (13)	744 (13)	12	1.2 (.2)	1.2 (.3)	0
Incongruent	836 (15)	824 (14)	-12	3.7 (.5)	3.2 (.4)	-.5
Interference Effect	104	80	-24	2.5	2.0	-.5

Context items

A main effect of Item Type was found both in latencies, $F(1,79) = 211.79$, $MSE = 1010378$, $p < .001$, $\eta_p^2 = .728$, and in error rates, $F(1,79) = 15.60$, $MSE = .030$, $p < .001$, $\eta_p^2 = .165$, indicating faster and more accurate performance on neutral than on incongruent items. An interaction between Item Type and List Type also emerged in the latencies, $F(1, 79) = 28.12$, $MSE = 62129$, $p < .001$, $\eta_p^2 = .263$, and in the error rates, $F(1,79) = 11.58$, $MSE = 020$, $p = .001$, $\eta_p^2 = .128$, reflecting larger interference in the MN list (latencies: 140 ms; error rates: 3.5%) than in the MI list (latencies: 85 ms; error rates: -.3%).

Transfer items

For transfer items, we also found a main effect of Item Type both in latencies, $F(1,79) = 277.68$, $MSE = 678593$, $p < .001$, $\eta_p^2 = .779$, and in error rates, $F(1,79) = 37.19$, $MSE = .039$, $p < .001$, $\eta_p^2 = .320$, with faster and more accurate performance on neutral than on incongruent items. In the latencies, this main effect was qualified by an interaction with List Type, $F(1,79) = 14.08$, $MSE = 10683$, $p < .001$, $\eta_p^2 = .151$, indicating a larger interference effect in the MN list (104 ms) than in the MI list (80 ms) (no interaction was found in the error rates, $F < 1$).

Discussion

A popular control-based explanation for the Proportion-Congruent (PC) effect in the Stroop task (i.e., the finding that the congruency effect increases as the proportion of congruent items in the list increases) assumes that attention to task-relevant information is proactively (i.e., before the appearance of any specific item) increased in a situation in which conflict is frequent (e.g., an MI list) compared to a situation in which conflict is infrequent (e.g., an MC list; De Pisapia &

Braver, 2006; Kane & Engle, 2003). However, multiple alternative explanations have been advanced recently that could explain that effect without invoking this form of proactive control (Schmidt, 2013b, 2019). By replacing congruent items with neutral items in the PC paradigm, we created a situation in which performance in a list in which conflict was infrequent could be compared to that in a list in which conflict was frequent while controlling for information other than list-wide conflict frequency, information that individuals might use to modulate word interference.

A Proportion-Neutral (PN) effect, similar to the PC effect, emerged, with more interference in the MN list (the list in which conflict was infrequent) than in the MI list (the list in which conflict was frequent), for both context and transfer items. For context items interference could have been modulated based on either list-wide conflict frequency information or color-specific conflict frequency information. Such is not the case for transfer items, items for which the only viable mechanism for producing this effect would appear to be a proactive mechanism of adaptation to list-wide conflict frequency, as conceived of by, e.g., De Pisapia and Braver (2006; see also Kane & Engle, 2003; Gonthier et al., 2016). That is, this effect was obtained in a situation in which, similar to that examined by Bugg (2014a), no item-specific conflict frequency information and no word-response contingencies existed that could have produced it (see Spinelli, Perry, & Lupker, 2019, for a PC effect obtained in a similar situation in the picture-word interference task). In addition, unlike the crucial situations examined by Bugg, the present situation was one in which no difference existed in the extent to which words were informative in the two lists and the strength with which the words were correlated with the colors, differences that, in principle, could also lead to participants adjusting attention in a manner compatible with a PC effect (Bugg, 2014a; Schmidt, 2019).

The present results may seem to contrast with those of Bugg et al. (2011), who found no difference between an MN list and an MI list on neutral transfer items, although a difference on those items did emerge when the MN and the MI lists were compared with an MC list (slower latencies in the MC list than in the other two lists). As noted in the Introduction, however, that pattern of results may simply imply that MC lists promote a word-reading process (because reading the word would often result in the correct response) that neither MN nor MI lists promote (because reading the word would often result in an incorrect response). However, an MN list may still induce a relaxation of attention that, while not promoting word reading, would make dealing with incongruent items harder in that list than in an MI list. Our finding of a PN effect, with larger Stroop interference on the transfer items in the MN list than in the MI list, is consistent with this idea.

Although we obtained a PN effect in the absence of item-specific, contingency learning, and stimulus informativeness confounds, another potential confound could have existed in our experiment that might explain that effect without invoking a process of adaptation to conflict frequency. This additional confound refers to the fact that temporal expectancies for the emission of a response are inevitably slower in a list in which most trials elicit a slow response (e.g., an MI list) than in a list in which most trials elicit a fast response (e.g., an MN list). According to Schmidt's (2013c) temporal learning account, a faster temporal expectancy would cause the difference between easy-to-process stimuli (e.g., neutral items) and hard-to-process stimuli (e.g., incongruent items) to increase because easy stimuli, but not hard stimuli, will speed up because they can be processed fast enough to meet that (fast) temporal expectancy. In the case of an MN list (i.e., a situation in which the temporal expectancy is relatively fast), the result would be a large interference effect. Conversely, a slower temporal expectancy would cause the difference between easy and hard stimuli to decrease because hard stimuli may also speed up because they can be processed fast enough to meet the (slow) temporal expectancy in that situation (although hard stimuli appear to be relatively insensitive to temporal expectancies, at least in some situations: Kinoshita & Mozer, 2006). As a result, an MI list in which the temporal expectancy is relatively slow would produce, if anything, a reduced interference effect. In sum, a PN effect (as well as a PC effect in the standard PC paradigm) could be produced by a temporal learning process rather than by a process of adaptation to conflict frequency.

At present, however, there is little convincing evidence in support of a temporal learning explanation of PC effects. To demonstrate that temporal expectancies could explain PC effects obtained in confound-minimized situations, Schmidt (2013c) re-analyzed the data from one of those situations (Hutchison, 2011) using linear mixed-effects modeling, a type of analysis that, unlike traditional mean-based ANOVAs, allows the evaluation of trial-level predictors. Indeed, in his re-analysis, Schmidt included a trial-level predictor functioning as an index of temporal expectancy, the latency on the most recent trial (i.e., RT on trial n – 1), in addition to the typical predictors of a PC paradigm (i.e., list type and congruency). Schmidt reasoned that, because easy stimuli are more likely to benefit from fast temporal expectancies (i.e., following a fast RT) whereas hard stimuli are more likely to benefit, if anything, from slower temporal expectancies (i.e., following a slow RT), evidence for a temporal learning process being engaged in the Stroop task should take the form of an interaction between temporal expectancy (RT on trial *n* – 1) and the congruency of the stimulus on trial *n*. Specifically, the congruency effect on a given trial should be larger following faster responses than following slower responses, an effect that he did obtain. Because faster responses are, by necessity, more common in an MC list than in an MI list, the implication is that this temporal learning interaction would tend to inflate the congruency effect in the MC list and reduce it in the MI list, resulting in a PC effect.

More recently, however, Cohen-Shikora, Suh, and Bugg (2018) clearly demonstrated that Schmidt's (2013c) results were likely biased because of the nonlinear transformation that he applied to the RT data. While transformations of this sort do a decent job of accommodating the assumption made by linear mixed-effects models that the dependent variable be normally distributed (an assumption that raw RTs typically fail to satisfy), they have the downside of systematically altering the pattern and size of interaction terms, making analyses of interactions unreliable overall (Balota, Aschenbrenner, & Yap, 2013). Indeed, Cohen-Shikora et al. re-analyzed a number of datasets (including Hutchison's, 2011) and were unable to replicate Schmidt's (2013c) temporal learning interaction when untransformed, rather than transformed, RT data were used in a type of mixed-effects model that tolerates deviations from normality in the dependent variable (a generalized linear mixed-effects model: Lo & Andrews, 2015; see also Spinelli et al., 2019). Several additional attempts to evaluate the impact of temporal learning by Cohen-Shikora et al. also yielded no convincing evidence that temporal learning contributes to the PC effect to any extent.

In sum, although there was no control in the present experiment for a potential temporal learning confound (a faster temporal expectancy in the MN list than in the MI list), the extant evidence suggests that this confound does not pose a serious challenge to control-based interpretations of PC/PN effects (although see Schmidt, 2017). In fact, when we re-analyzed the (raw) RT data of the present experiment using RT on trial $n - 1$ as an additional predictor in a generalized linear mixed-effects model, we found no evidence for the temporal learning interaction which, according to Schmidt (2013c), would support a temporal learning interpretation of the PN effect that we obtained. On the contrary, we found a *reversed* temporal learning interaction on both context and transfer items, with congruency effects increasing, rather than decreasing, following slower responses on the preceding trial. (note 4) This reversed pattern, which was occasionally reported in the analyses conducted by Cohen-Shikora et al. (2018; see also Spinelli et al., 2019), is completely inconsistent with Schmidt's (2013c) temporal learning account and makes a strong case that the PN effects that we obtained did not emerge from the temporal learning process that Schmidt hypothesized. (note 5)

Overall, the present results challenge the argument that adaptation to list-wide conflict frequency is not a process that humans use (Schmidt, 2013b, 2019). Note, however, that the evidence supporting the use of this process was obtained when learning contingencies between words and responses was not a viable option, the only type of situation, according to Bugg (2014a), in which a proactive conflict adaptation process is used. Thus, although the present results do indicate that this process exists, they do not argue against the possibility that its usage might be restricted to the type of situation under examination here, i.e., one in which an alternative, contingency learning process is not available.

Footnotes

1. In the following, for simplicity, we refer to adaptation to list-wide conflict frequency as a "proactive process" even though, in a Dual-Mechanism of Control framework (Braver, 2012; Braver et al., 2007; see also Kane & Engle, 2003), this adaptation would entail the engagement of both proactive control (in a frequently conflicting list) and a form of reactive control (in an infrequently conflicting list).
2. It is important to emphasize that all of these competing accounts of the PC effect are, in essence, learning accounts (see, e.g., Egner, 2014). For example, the process of adaptation to list-wide conflict frequency could be described as the process whereby participants *learn* to focus attention to the task-relevant dimension in a frequently conflicting list and *learn* to relax attention in an infrequently conflicting list. What does distinguish these accounts is what is being learned, e.g., associations between words and responses (a contingency learning process) vs. associations between contexts and control settings (a conflict adaptation process).
3. We do not think that using color-unrelated words instead of consonant strings as neutral words would have dramatically changed the types of processes that participants would have used in dealing with the task. However, we do think that detecting an effect of adaptation to list-wide conflict frequency is potentially harder in a situation in which color-unrelated words are used as neutral words. The reason is that, even in an MN list, participants may sometimes feel a need to focus attention to the color dimension to avoid inadvertently reading the frequent color-unrelated words, making the process used in that type of list not particularly different from the process used in an MI list. As a

result, detecting a PN effect in that situation may require a much more sensitive protocol (e.g., larger sample size, stronger PN manipulation) than the one used here.

4. The procedure and statistical software used for this analysis were the same as that used in Spinelli et al. (2019). The R script used for the analysis is publicly available at https://osf.io/yk57z/.
5. A reviewer of an earlier version of this manuscript proposed yet another account that might explain the list-wide PC effect. This account is that repeated practice with one type of stimulus (e.g., the neutral type or the incongruent type) would speed up responses to that stimulus type. Thus, incongruent items would elicit faster responses in the MI list compared to the MN list because participants receive more practice with incongruent items in the MI list, and, conversely, neutral items would elicit faster responses in the MN list compared to the MI list because participants receive more practice with neutral items in the MN list. The result would be a PN effect. Consistent with this explanation, the PN effect that we observed in the present experiment was determined by both a speed-up for the incongruent items in the MI list compared to the MN list and a speed-up for the neutral items in the MN list compared to the MI list. However, the reason that we do not find this explanation particularly compelling is that, first, it is not clear why an effect of practice would benefit neutral stimuli, i.e., stimuli that produce little or no conflict, to any measurable degree. Furthermore, a speed-up for neutral items in situations in which those items are frequently practiced has not been reported in manipulations similar to ours. For example, as noted, Bugg et al. (2011) found that responses to a fixed set of neutral items were not any faster when those

items appeared in an MN list (i.e., a list in which neutral items were frequently practiced) than when they appeared in an MI list (i.e., a list in which neutral items were infrequently practiced). Similarly, Tzelgov et al. (1992) parametrically varied the proportion of neutral items in a list but latencies to neutral items did not decrease overall in lists in which those items were more frequent. Thus, although this practice explanation would seem compatible with our results, it does not appear to gain any support from the relevant findings in the literature.

Chapter 3.5: Interim summary

Recent research has pointed out several explanations for why a list-wide PC effect might emerge in the Stroop task (e.g., Blais et al., 2007; Schmidt, 2013b, 2019), explanations that do not require the process of adaptation to list-wide conflict frequency that traditional explanations assume (e.g., Botvinick et al., 2001; Kane & Engle, 2003; De Pisapia & Braver, 2006). In Chapter 3, most (if not all) of the processes that those explanations involve were accounted for in a Proportion-Neutral manipulation in the Stroop task. Even so, a list-wide Proportion-Neutral effect, similar to the list-wide PC effect, emerged, with larger interference in a list in which conflict was overall infrequent than in a list in which conflict was overall frequent. These results, combined with the results from the picture-word interference task reported in Chapter 2, make a strong case that adaptation to list-wide conflict frequency is possible, at least in situations in which contingency learning is not an option. This idea converges with the conclusions of a few recent studies examining interference effects in the list-wide PC paradigm (Bugg, 2014a; Cohen-Shikora et al., 2018; Hutchison, 2011).

The story is somewhat different for adaptation to item-specific conflict frequency, however. Indeed, it was around the putative form of control in this situation that the debate on the processes underlying PC effects originated (Schmidt & Besner, 2008). As noted, Jacoby et al. (2003) reported an item-specific PC effect in the color-word Stroop task, with a larger congruency effect for color words presented mainly in their congruent color (MC items) than for color words presented mainly in another (incongruent) color (MI items) intermixed in the same list. As Jacoby et al. noted, this effect might reflect either (or both) of two processes: A process (subsequently characterized as an item-specific, reactive control process) whereby the recognition of a certain item would lead to use of the control setting most appropriate for that item (focused attention to task-relevant information for MI items vs. relaxed attention for MC items) or the more general process of learning word-response contingencies.

As described in Chapter 1, most researchers now agree that a contingency learning process is the main process underlying the item-specific PC effect, at least in Jacoby et al.'s (2003) two-item set design (Atalay & Misirlisoy, 2012; Bugg & Hutchison, 2013; Hazeltine & Mordkoff, 2014; Schmidt, 2013a). Because in this design two words and two colors are combined in both the MC set and the MI set, a situation is created in which a high-contingency color exists for both MC words and MI words (i.e., a situation in which not only MC words appear more frequently in their congruent color but also MI words also appear more frequently in one specific incongruent color). According to Bugg and Hutchison (2013), because contingency learning is an especially reliable process in this situation, the item-specific PC effect that is typically obtained would mainly result from the implementation of that process rather than a process of adaptation to item-specific conflict frequency (although such a conflict adaptation process might still be implemented in other situations, for example, in a four-item set design in which no contingencies for MI words can be learned; but see Schmidt, 2014b, 2019, for problems in interpreting results from this design).

In Chapter 4, the conclusion that there is little or no role for adaptation to item-specific conflict frequency in the two-item set design used by Jacoby et al. (2003) was re-examined. A contingency learning account of the item-specific PC effect would imply that the characteristics displayed by the contingency learning process in situations in which contingency learning is not confounded with other factors should replicate in the item-specific PC paradigm as well if the item-specific PC effect were, in fact, a product of contingency learning. One such characteristic was documented by Schmidt, De Houwer, and Besner (2010) in a non-conflict color identification in which color-unrelated words were used and each word appeared in both a high-contingency color (e.g., BRAG appearing more often in green) and in low-contingency colors (e.g., BRAG appearing rarely in blue and yellow). Schmidt et al. showed that the contingency learning effect (i.e., the latency difference between low-contingency and high-contingency items) was reduced for participants who performed the non-conflict color identification task while carrying a high working-memory load, a result that suggests that limited-capacity resources are necessary for learning word-response contingencies. Assuming that this limitation on the contingency learning process would be maintained in the Stroop task in which color names, rather than color-unrelated words, are used, the implication for a contingency learning account of the item-specific PC effect is that this effect should be reduced when participants perform the Stroop task while carrying a high working-memory load.

To test the contingency learning account of the item-specific PC effect, in three experiments, a non-conflict color identification task and an item-specific PC Stroop task were combined with a concurrent working memory load task imposing either a low or high working memory load. A no-load condition in which participants only performed the color identification tasks was also included. To anticipate the results, consistent with Schmidt et al. (2010), it was found that increasing working memory load in the non-conflict color identification task reduced contingency learning, although only when keypress responses and feedback on the accuracy of those responses were used, as in Schmidt et al.'s original study. In contrast, although increasing working memory load did produce higher latencies overall, it did not alter the item-specific PC effect in the Stroop task. This pattern of results was interpreted in the framework of the Dual-Mechanism-of-Control (DMC) account (Braver, 2012; Braver et al., 2007), an account that appears to explain why reduced working memory resources would not impair the item-specific PC effect. Although not as clear-cut as the working memory load manipulation, an individual-differences analysis of the data in Chapter 4 offered some corroboration for this interpretation.

Chapter 4:

Working Memory Load Dissociates Contingency Learning and Item-Specific Proportion-Congruent Effects

Introduction

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 performance on a list composed mainly of congruent items (a mostly-congruent list) to be compared to 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 strategy as posited by the traditional control account. What Jacoby et al. demonstrated was an <u>item-specific</u> proportion-congruent effect. In their manipulation (the "two-item set" design), two color words (e.g., GREEN and YELLOW) were presented mainly in their congruent color (mostly-congruent items, e.g., $\text{GREEN}_{\text{green}}$ appearing more often than $\text{GREEN}_{\text{yellow}}$) and two other color words (e.g., RED and BLUE) were presented mainly in an incongruent color (mostly-incongruent items, e.g., RED_{blue} appearing more often than RED_{red}). 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, it appears that whatever strategy 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 strategy must have been an item-specific one, based on the congruency proportion assigned to each item in the list, that is, a strategy 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 unitary process, i.e., 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. 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). Reactive control can also take the form of a process that uses information about the stimulus to select a specific control strategy for dealing with that stimulus (e.g., focusing attention to the color dimension in response to words that typically appear in incongruent colors; Bugg, Jacoby, & Chanani, 2011; Bugg & Hutchison, 2013; see also Gonthier et al., 2016), a form of reactive control that will be most relevant to our discussion of the item-specific proportion-congruent effect. Although successful behavior likely depends on a mixture of proactive and reactive control engagement, the two modes 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).

In fact, list-wide and item-specific proportion-congruent manipulations appear to provide good illustrative examples of the sorts of manipulations that would lead to the use of both proactive and reactive control strategies, at least in some situations (Gonthier et al., 2016). A list-wide proportion-congruent manipulation puts participants in a situation in which different expectancies concerning the congruency of upcoming stimuli can be formed, with those expectancies favoring reliance on distinct control modes. Specifically, a situation in which frequent conflict is expected between the word and the color, such as when the list is mostly incongruent, is thought to favor the implementation of a proactive, top-down control strategy that minimizes interference from the word by constantly maintaining the color-naming goal. In contrast, a situation in which the word and the color are not expected to be conflicting, such as when the list is mostly congruent, would appear to favor the use of a reactive, bottom-up control strategy whereby the color-naming goal is frequently neglected and is only retrieved upon presentation of the infrequent incongruent words (Botvinick et al., 2001). In sum, a list-wide proportion-congruent manipulation would lead to the engagement of different strategies in the two types of lists, with a mostly-incongruent list favoring a proactive strategy consisting of maintaining the color-naming goal and a mostly-congruent list favoring a reactive strategy consisting of re-activating that goal upon detection of conflict (for a more complete discussion of these issues, see Kane & Engle, 2003).

The situation is a bit different in most item-specific proportion-congruent manipulations. The item-specific proportion-congruent manipulation puts participants in a situation in which they can learn associations between words and their congruency and use those associations to select the control strategy (e.g., relaxed vs. focused attention onto 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., RED, 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., GREEN, 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 GREEN but its color is yellow rather than its usual green color).

It is worth noting that reliance on this reactive strategy, the strategy thought to underlie the item-specific proportion-congruent effect, would not necessarily prevent other strategies from being invoked, although such strategies 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 strategy 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 strategy instead of applying, or while concurrently applying, a reactive strategy of adaptation to item-specific conflict frequency. In other words, it is reasonable to hypothesize that reactive control may be the more prominent strategy, but not necessarily the only strategy that individuals could employ in an item-specific proportion-congruent manipulation. This reasoning, however, 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 strategy of maintaining the color-naming goal were used for both mostly-congruent and mostly-incongruent items, this strategy 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 in a DMC framework (Gonthier et al., 2016).

Although the control account of the item-specific proportion-congruent effect supports a role for control processes in this effect, it does not necessarily negate the possibility that non-control processes may also have an important role in this effect. Indeed, Bugg and colleagues (Bugg, 2015; Bugg et al., 2011; Bugg & Hutchison, 2013), for example, 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 processing strategy (Schmidt, Crump, Cheesman, & Besner, 2007). In color-word 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, 2018).

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 RED appears most often in blue, individuals may learn to associate the word RED with the (incongruent) blue response. Conversely, if the mostly-congruent word GREEN appears most often in green, that would allow participants to learn that GREEN predicts the (congruent) green response. Crucially, if frequent word-color pairs elicit faster responses, relatively fast responding to the high-contingency incongruent item RED_{blue} will lead to a relatively small congruency effect for mostly-incongruent items, whereas fast responding to the high-contingency congruent item $\text{GREEN}_{\text{green}}$ 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, 2014a; 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 <u>list-wide</u> proportion-congruent effects do persist when controlling for both contingency learning (Bugg, 2014a; Bugg & Chanani, 2011; Gonthier et al., 2016; Hutchison, 2011; Spinelli, Perry, & Lupker, 2019) and learning of list-wide temporal expectancies, another non-conflict learning mechanism thought to contribute to generating list-wide proportion-congruent effects (Cohen-Shikora, Suh, & Bugg, in press; 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, 2014b, 2017).

With respect to the <u>item-specific</u> proportion-congruent effect, however, the situation is a bit different. To sum up this issue, 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 RED) while the latter assumes that they associate words with specific responses (e.g., predict a blue response upon presentation of the mostly-incongruent word RED). While both mechanisms might be used, researchers who have tried to directly dissociate the two accounts have only found support for contingency learning processes (Hazeltine & Mordkoff, 2014; Schmidt, 2013a). 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, this result was not observed. Instead, performance on mostly-congruent and mostly-incongruent words was equivalent when those words appeared in the critical incongruent colors, suggesting that no conflict adaptation strategy 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 with 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 strategy 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).

<u>The present research</u>

The present research was an attempt to re-examine the conclusion that performance (i.e., the item-specific proportion-congruent effect) in Jacoby et al.'s (2003) two-item set design is dominated by 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, 2010). Schmidt et al. (2010) had participants perform this non-conflict 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.

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). However, what is also quite possible is 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, as opposed to *proactive* control processes, i.e., top-down control strategies 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 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. Indeed, in an fMRI memory study, Speer, Jacoby, and Braver (2003) obtained evidence consistent with this idea. As shown by the activity dynamics in a distinct set of brain regions, an expected low WM load showed an activation pattern consistent with the idea that participants were using a proactive strategy 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 strategy, whereby study items were not actively maintained and the probe was used as a retrieval cue instead.

A similar type of conclusion would apply in experiments analyzing individual differences in WM resources, typically defined in terms of WM capacity. WM capacity and WM load refer to distinct types of individual variation – for the former, inter-individual variation is determined by the amount of information an individual is able to maintain in working memory while performing a distracting task; for the latter, intra-individual variation is determined by the impact of one task on concurrent performance on another task. Both WM capacity and WM load, however, limit the amount of WM resources available, thus potentially influencing the strategies that individuals would use while performing a task. According to the DMC account (and in earlier versions of this type of account, e.g., Kane & Engle, 2003), WM capacity is, in fact, an important determinant of the extent to which individuals rely on proactive versus reactive modes of control (Braver, 2012). Although all individuals, presumably, have access to proactive and reactive control, low WM-capacity individuals, having less available WM resources to use, would tend to rely more on reactive control, whereas high WM-capacity individuals, having more available WM resources to use, would tend to rely more on proactive control.

The implications of these DMC-based claims concerning proportion-congruent effects in the Stroop task as a function of WM capacity (when there is no WM load) would be as follows. Because individuals with fewer WM resources (i.e., low WM-capacity individuals) should tend to rely on reactive control, they would show the typical markers of this form of control in proportion-congruent manipulations in the Stroop task: Specifically, they would show a disproportionately large congruency effect in the mostly-congruent list in the list-wide proportion-congruent manipulation and a robust item-specific proportion-congruent effect in the item-specific proportion-congruent manipulation. The reason is that, as noted, reactive strategies are involved in both of these situations: In mostly-congruent lists in the list-wide manipulation, use of a reactive strategy of retrieving the task goal upon detection of conflict is what produces the large congruency effect in that list; in the item-specific manipulation, use of a reactive strategy of selecting the appropriate control process upon recognition of specific items is what produces the item-specific proportion-congruent effect in general.

In contrast, because individuals with more WM resources (i.e., high WM-capacity individuals) should be more prepared to maintain a proactive strategy, they would be less likely to show typical markers of reactive control in list-wide and item-specific proportion-congruent manipulations in a Stroop task with no WM load. Specifically, in a list-wide proportion-congruent manipulation, those individuals may produce a larger congruency effect in the mostly-congruent list than in the mostly-incongruent list (resulting in the typical list-wide proportion-congruent effect); however, the congruency effect in the mostly-congruent list would not be as large as the congruency effect produced by low WM-capacity individuals in the same (mostly-congruent) list because, in this situation, high WM-capacity individuals would not rely on reactive control as heavily as low WM-capacity individuals do. In addition, use of a proactive strategy of maintaining attention focused on task-relevant information in high WM-capacity individuals would make those individuals less likely to learn associations between items and conflict frequency, thus resulting in an attenuated item-specific proportion-congruent effect overall. Patterns of results like the ones just described have indeed been observed (Kane & Engle, 2003; Hutchison, 2011), although in error rates more often than in latencies (as will be reviewed in introducing Experiments 3A and 3B below). Furthermore, these results are in line with findings from neuroimaging research indicating that in a memory task, individuals with higher fluid intelligence (a variable highly related to WM capacity: Engle, Tuholsky, Laughlin, & Conway, 1999) show greater reliance on proactive control whereas individuals with lower fluid intelligence preferentially engage in a reactive strategy instead (Burgess & Braver, 2010).

In general, results from both intra-individual variation in WM resources (obtained by manipulating WM load) and inter-individual variation in WM resources (obtained by comparing individuals varying in WM capacity) suggest that reactive control is, in fact, relatively easily implemented when WM resources are scarce, an idea that is well accommodated within the DMC framework (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 one 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 strategy 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 either a large load or a low WM capacity, 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 non-conflict 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, 2014a; Bugg & Hutchison, 2003). To preview, we were not able to replicate the original pattern in the non-conflict 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 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 non-conflict 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 evaluate the idea that lower WM resources are associated with 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 non-conflict, 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 GREEN presented 75% and 25% of the time in green and yellow, respectively) and mostly-incongruent color words (e.g., the word RED presented 75% and 25% of the time in blue and red, 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 strategy for performance (Bugg, 2014a; 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 involving one-third of the participants) or five digits (the high-load group involving one-third of the participants), as in Schmidt et al. (2010).

Colors were responded to vocally, whereas Schmidt et al. (2010) had participants respond to colors via button pressing. Note that Schmidt et al. provided no indication that response modality 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 (non-conflict 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.

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

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 either 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 digits 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 or 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. (note 1) Latency analyses were conducted only on correct responses in the color identification task. (note 2)

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). (note 3) 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). 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 & 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). In addition, the raw data and the SPSS files used for the analyses are publicly available at https://osf.io/rtnw2/.

Table 3.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Experiment 1A – Vocal Non-conflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	589 (16)	1.0 (.3)
Low	601 (18)	1.3 (.5)
Contingency effect	12	.3
Low load		
High	786 (35)	.5 (.2)
Low	801 (37)	.9 (.3)
Contingency effect	15	.4
High load		
High	743 (30)	.5 (.1)
Low	752 (28)	.4 (.2)
Contingency effect	9	-.1

Table 4.

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

Congruency	RTs			Error rates		
	Mostly-congruent items	Mostly-incongruent items	Item type effect	Mostly-congruent items	Mostly-incongruent items	Item type effect
<u>No load</u>						
Congruent	626 (17)	665 (23)	39	.1 (.1)	.8 (.6)	.7
Incongruent	751 (25)	715 (18)	-36	7.1 (.017)	2.4 (.7)	-4.7
Congruency Effect	125	50	-75	7	1.6	-5.4
<u>Low load</u>						
Congruent	726 (23)	757 (24)	31	.1 (.1)	.2 (.2)	.1
Incongruent	855 (30)	794 (22)	-61	1.1 (.006)	.7 (.3)	-.4
Congruency Effect	129	37	-92	1	.5	-.5
<u>High load</u>						
Congruent	793 (33)	821 (30)	28	0 (0)	.5 (.4)	.5
Incongruent	899 (29)	849 (32)	-50	5.3 (1.2)	1.7 (.7)	-3.6
Congruency Effect	106	28	-78	5.3	1.2	-4.1

Experiment 1A (non-conflict color identification task)

RTs. Both the main effects of Contingency (high-contingency faster than low-contingency), $F(1, 57) = 8.40$, $MSE = 4802$, $p = .005$, $\eta_p^2 = .128$, and of WM Load, $F(2, 57) = 13.45$, $MSE = 432326$, $p < .001$, $\eta_p^2 = .321$, 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 that the low-load and the high-load groups did not differ from one another ($p = .487$). Importantly, Contingency and WM Load did not interact ($F(2, 57) = .19$, $MSE = 98$, $p = .825$, $\eta_p^2 = .007$), 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 $BF_{01} = 6.51$, meaning that the data were 6.51 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), $F(1, 57) = 99.64$, $MSE = 374025$, $p < .001$, $\eta_p^2 = .636$, and of WM Load, $F(2, 57) = 9.94$, $MSE = 466469$, $p < .001$, $\eta_p^2 = .259$. 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 that 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, $F(1, 57) = 56.18$, $MSE = 99133$, $p < .001$, $\eta_p^2 = .496$, 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, $F(2, 57) = .230$, $MSE = 407$, $p = .795$, $\eta_p^2 = .008$, suggesting that the item-specific proportion-congruent effect was equivalent in all load groups. The Bayes Factor, $BF_{01} = 6.29$, indicated "moderate" evidence for the absence of the three-way interaction.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), $F(1, 57) = 33.39$, $MSE = .045$, $p < .001$, $\eta_p^2 = .369$, Item Type (mostly incongruent more accurate than mostly congruent), $F(1, 57) = 12.71$, $MSE = .009$, $p = .001$, $\eta_p^2 = .182$, and WM Load, $F(2, 57) = 6.68$, $MSE = .009$, $p = .002$, $\eta_p^2 = .190$. 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, $F(1, 57) = 23.36$, $MSE = .017$, $p < .001$, $\eta_p^2 = .291$. However, Congruency also interacted with WM Load, $F(2, 57) = 4.82$, $MSE = .007$, $p = .012$, $\eta_p^2 = .145$, and the three-way interaction was also significant, $F(2, 57) = 4.08$, $MSE = .003$, $p = .022$, $\eta_p^2 = .125$.

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%), $F(1, 38) = 9.21$, $MSE = .012$, $p = .004$, $\eta_p^2 = .195$, or the high-load group (2.4%), $F(1, 38) = 7.56$, $MSE = .006$, $p = .009$, $\eta_p^2 = .166$, whereas the no-load and high-load groups did not differ from one another, $F(1, 38) = .60$, $MSE = .001$, $p = .442$, $\eta_p^2 = .016$. 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, $F(1, 38) = 7.08$, $MSE = .005$, $p = .011$, $\eta_p^2 = .157$, and high-load groups, $F(1, 38) = 6.89$, $MSE = .003$, $p = .012$, $\eta_p^2 = .153$, but no difference was found between the no-load and high-load groups, $F(1, 38) = .39$, $MSE = .000$, $p = .538$, $\eta_p^2 = .010$. Separate analyses for each load group showed that the reason for this 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, $F(1, 19) = 10.90$, $MSE = .014$, $p = .004$, $\eta_p^2 = .365$, and the high-load group, $F(1, 19) = 13.56$, $MSE = .008$, $p = .002$, $\eta_p^2 = .416$, there was no significant interaction for the low-load group, $F(1, 19) = .82$, $MSE = .000$, $p = .\ 376$, $\eta_p^2 = .041$. In general, it appears that the 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 because there was 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 non-conflict color identification task and to extend those findings to the Stroop task using vocal responses. Surprisingly, however, the non-conflict 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 (note 4). Although this pattern of results supports the idea that item-specific proportion-congruent effects in the Stroop task and contingency learning effects in the non-conflict 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 has little to do with the WM load manipulation employed.

Second, the contingency learning effect in Experiment 1A was small (12 ms in the no-load condition) compared to what is typically reported in the literature (e.g., Schmidt et al., 2007, reported a 60-ms contingency learning effect with a design similar to the one used here). The possibility exists, therefore, that there may be an important difference between using vocal versus manual responding in the color identification task. Indeed, Forrin and MacLeod (2017) and Spinelli, Perry, and Lupker (under review) recently reported smaller contingency learning effects for vocal than for manual responding to the color of color-unrelated words. The difference in the magnitude of contingency learning effects for manual versus vocal responding might be explained in terms of whether responses to stimuli are undertrained (manual) or overtrained (vocal), with stimulus-response associations playing a larger role in the former situation than the latter (Schmidt, 2018; see also Spinelli et al., under review). More important for the present discussion, however, is that if vocal responding typically elicits small contingency learning effects, observing a significant reduction in their size might be challenging (for a similar point, see Kinoshita et al., 2018). Insofar as manual responding elicits larger baseline contingency learning effects, use of that response modality 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, a hypothesis that provides the motivation for Experiments 2A and 2B. (note 5)

Experiments 2A & 2B (manual responses)

Experiments 2A and 2B were identical to Experiments 1A and 1B, except that manual responding to colors was used in an attempt to increase the size of baseline contingency learning effects and thus provide a better opportunity to observe modulations of such effects. Another difference was that participants were given feedback on their performance. Feedback was given for two reasons: First, to replicate Schmidt et al.'s (2010) original experiment more closely; second, because evidence from our lab suggests that feedback can modulate the size of contingency learning effects, with larger contingency learning effects when color identification is feedback-assisted (Spinelli et al., under review).

Method

Participants

Sixty-three participants took part in Experiment 2A (non-conflict 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 were identical to those in Experiments 1A and 1B, respectively.

Procedure

The procedure was the same as in Experiments 1A and 1B, with some minor 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 correct responses in the color identification task. 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.

Table 5.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Experiment 2A – Manual Non-conflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	712 (26)	2.5 (.5)
Low	769 (25)	3.6 (.8)
Contingency effect	57	1.1
Low load		
High	829 (28)	2.7 (.5)
Low	857 (27)	3.6 (.7)
Contingency effect	28	.9
High load		
High	843 (29)	3.5 (.6)
Low	855 (27)	4.4 (.9)
Contingency effect	12	.9

Table 6.

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

Congruency	RTs			Error rates		
	Mostly-congruent items	Mostly-incongruent items	Item type effect	Mostly-congruent items	Mostly-incongruent items	Item type effect
<u>No load</u>						
Congruent	766 (29)	817 (33)	51	.8 (.5)	1.1 (.5)	.3
Incongruent	940 (41)	885 (36)	-55	2.1 (.9)	1.3 (.6)	-.8
Congruency Effect	174	68	-106	1.3	.2	-1.1
<u>Low load</u>						
Congruent	794 (30)	857 (36)	63	2.1 (.7)	2 (.7)	-.1
Incongruent	947 (33)	904 (35)	-43	7.7 (1.3)	3.6 (.7)	-4.1
Congruency Effect	153	47	-106	5.6	1.6	-4
<u>High load</u>						
Congruent	877 (25)	916 (26)	39	2.8 (.7)	2.7 (1.2)	-.1
Incongruent	1032 (28)	923 (24)	-109	5.8 (1.3)	3 (.5)	-2.8
Congruency Effect	155	7	-148	3	.3	-2.7

Experiment 2A (non-conflict color identification task)

RTs. Both the main effects of Contingency (high-contingency faster than low-contingency), $F(1, 57) = 41.86$, $MSE = 31731$, $p < .001$, $\eta_p^2 = .423$, and of WM Load, $F(2, 57) = 5.164$, $MSE = 149122$, $p = .009$, $\eta_p^2 = .153$, 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, $F = 6.80$, $MSE = 5157$, $p = .002$, $\eta_p^2 = .193$. 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), $F(1, 38) = 5.33$, $MSE = 3898$, $p = .024$, $\eta_p^2 = .127$, and high-load groups (12 ms), $F(1, 38) = 15.10$, $MSE = 10118$, $p < .001$, $\eta_p^2 = .284$, whereas the low- and high-load groups did not significantly differ from each other, $F(1, 38) = 1.62$, $MSE = 1456$, $p = .211$, $\eta_p^2 = .041$. 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, $F(1, 57) = 5.66$, $MSE = .003$, $p = .021$, $\eta_p^2 = .090$.

Experiment 2B (Stroop task)

RTs. There was a significant main effect of Congruency (congruent faster than incongruent), $F(1, 57) = 136.36$, $MSE = 609659$, $p < .001$, $\eta_p^2 = .705$. Latencies also tended to slow down as load increased, however the WM Load effect was only marginal, $F(2, 57) = 2.52$, $MSE = 153814$, $p = .089$, $\eta_p^2 = .081$. The only significant interaction was that between Congruency and Item Type, $F(1, 57) = 74.61$, $MSE = 215459$, $p < .001$, $\eta_p^2 = .567$, 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, $F(2, 57) = 1.01$, $MSE = 2917$, $p = .371$, $\eta_p^2 = .034$, suggesting that the item-specific proportion-congruent effect was equivalent in the three load groups. The Bayes Factor, $BF_{01} = 5.32$, 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), $F(1, 57) = 16.93$, $MSE = .024$, $p < .001$, $\eta_p^2 = .229$, Item Type (mostly incongruent more accurate than mostly congruent), $F(1, 57) = 6.76$, $MSE = .010$, $p = .012$, $\eta_p^2 = .106$, and WM Load, $F(2, 57) = 7.73$, $MSE = .015$, $p = .001$, $\eta_p^2 = .213$. 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%),

$F(1, 57) = 9.16$, $MSE = .010$, $p = .004$, $\eta_p^2 = .138$. 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, $F(2, 57) = 2.89$, $MSE = .004$, $p = .064$, $\eta_p^2 = .092$. The three-way interaction between Congruency, Item Type, and WM load was not significant, $F(2, 57) = .91$, $MSE = .001$, $p = .41$, $\eta_p^2 = .031$, and the Bayes Factor analysis revealed that there was, indeed, "moderate" evidence for the absence of this interaction, $BF_{01} = 3.97$.

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 non-conflict 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 with vocal responses to colors in Experiment 1A (12 ms), replicating recent findings that 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 hypothesis, 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, 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 non-conflict 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.

As in Experiments 1A and 1B, 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. This 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. Because different participants took part in Experiments 2A and 2B, the most likely explanation for the results in the high-load group in that experiment is merely random noise. Nonetheless, 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 non-conflict 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 non-conflict 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 strategy, playing a crucial role instead. Indeed, the finding obtained in both Experiments 1B and 2B that WM load had 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 strategy that generates a proportion-congruent effect, continues to be a useful option when available WM resources are decreased.

Notably, part of the evidence supporting the DMC account (and accounts that use conceptually similar notions, e.g., Kane & Engle, 2003) comes from individual-differences research, i.e., research that investigates variation in WM resources due to differences naturally occurring among individuals rather than to variation in WM resources generated by an experimental manipulation (Burgess & Braver, 2010; Speer et al., 2003). As mentioned in the Introduction, according to this type of account, the availability of WM resources as determined by WM capacity should impact performance in a way that is similar to that of a concurrent memory load, by altering the relative reliance on reactive and proactive control. Although all individuals presumably have access to proactive and reactive modes of control, individual differences do appear to exist in terms of the extent to which people rely on those modes, with individuals who score higher and individuals who score lower in fluid intelligence and WM capacity eliciting patterns of behavior and brain activity that are suggestive of a preference for proactive and reactive control, respectively. More importantly for present purposes, evidence consistent with such a pattern of results was recently reported by Hutchison (2011) for proportion-congruent effects in the Stroop task.

Hutchison (2011) 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. (For low WM-capacity participants, the congruency effect was 10.8% for mostly-congruent items vs. 7.1% for mostly-incongruent items, whereas for high WM-capacity participants, the congruency effect was 5.1% for mostly-congruent items vs. 5.4% for mostly-incongruent items) (note 6). Hutchison interpreted this finding as consistent with the idea that high and low WM-capacity individuals differ in the extent to which they rely on proactive and reactive modes of control, with the increased reliance on reactive control in low WM-capacity individuals producing a more pronounced pattern of proportion-congruent effects. However, although interesting, those results do raise a number of questions.

The first question is, why should a difference between individuals differing in WM capacity emerge in error rates, but not in latencies (in a Stroop task with no WM load manipulation)? It is important to note that, in the Stroop task, latencies and error rates do not necessarily index the same processes. For example, Kane and Engle (2003) interpreted interference errors (i.e., incongruent – neutral) as an index of participants' inability to successfully maintain the task goal (see also MacLeod, 1991) and interference in the latencies as an index of conflict resolution, a process that occurs only when the task goal has been successfully maintained. Error rates appear to be a critical variable in discriminating Stroop performance in low and high WM-capacity individuals. Across several experiments employing a list-wide proportion-congruent manipulation, although Kane and Engle found little group differences in interference in the latencies, differences emerged more clearly in interference errors. Specifically, compared to high WM-capacity participants, low WM-capacity participants showed disproportionately high interference errors when responding to stimuli in a mostly-congruent list. Moreover, those interference errors produced by low WM-capacity individuals were mainly fast, word-reading responses. According to Kane and Engle, these results reflect a failure of low WM-capacity individuals to maintain the task goal in a context that favors goal neglect (i.e., a mostly-congruent condition in which frequent congruent trials bias attention away from the color-naming task), a failure that typically results in a word-reading error rather than an increased latency (see also MacLeod, 1991).

The second question is, how exactly would differences in reliance on proactive versus reactive control among low and high WM-capacity individuals explain Hutchison's (2011) results? Like Hutchison, we propose that low and high WM-capacity individuals differ in the extent to which they rely on proactive and reactive modes of control. Low WM-capacity individuals would mainly rely on the reactive control mode, a mode that uses word-specific conflict frequency to determine which strategy is best to implement with those words, and only minimally on the proactive control mode, a mode whereby the color-naming goal is maintained over trials and one which would prevent inadvertent word reading. As a result, those individuals might be especially inclined to focus attention to the color in response to mostly-incongruent words and to relax attention in response to mostly-congruent words. Crucially, with the latter words, a relaxation of attention would frequently result in neglecting the color-naming goal, especially if little effort is being made to proactively maintain that goal. The result is increased word-reading errors to the infrequent incongruent words in the mostly-congruent condition. On the other hand, high WM-capacity participants may be more likely to engage in a proactive strategy to maintain the color-naming goal throughout the task.

As noted in the Introduction, use of a proactive strategy should presumably attenuate item-specific proportion congruent effects. 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). On the other hand, the fact that Hutchison's (2011) high WM-capacity individuals showed a regular item-specific proportion-congruent effect in the latencies suggests that possessing a superior WM capacity (and, therefore, being able to engage proactive control) does not make those individuals so prepared for conflict that item-specific conflict frequency would have no impact whatsoever in their performance. That is, even though those individuals may preferentially engage proactive control, they would still have access to reactive control, i.e., they would still be able to learn about item-specific conflict frequency and adapt to it. However, it is possible that proactive maintenance of the task goal would reduce the impact of this process of adaptation to item-specific conflict frequency in some way.

Specifically, proactive control could prevent high WM-capacity individuals from neglecting the color-naming goal when dealing with all words, including mostly-congruent words, words for which use of reactive control would lead to a relaxation of attention to the color and would typically cause a disproportionate number of errors for the incongruent words in that condition. In other words, high WM-capacity individuals would be able to use a reactive strategy leading them to relax attention with mostly-congruent items and to focus attention to the color with mostly-incongruent items (which results in an item-specific proportion-congruent effect in the latencies); however, because they are also applying a proactive strategy, they would never let their guard down on the color-naming goal even when their attention is relatively relaxed (i.e., when dealing with mostly-congruent items), with the result being that incongruent words would produce no more inadvertent word-reading errors when presented in mostly-congruent items than in mostly-incongruent items (i.e., no item-specific proportion-congruent effect in the errors would emerge). In sum, assuming that 1) both low and high WM-capacity individuals can use a reactive strategy of adaptation to item-specific conflict frequency, 2) that proactive maintenance of the task goal prevents inadvertent word reading, and 3) that this proactive strategy is more readily applied by high WM-capacity individuals, appears to do a decent job of explaining why those individuals show an item-specific proportion-congruent effect in the latencies (as a result of a reactive strategy) but not in the error rates (as a result of a proactive strategy), an effect which is found in both dependent measures in low WM-capacity individuals.

In any case, although Hutchison's (2011) finding of an overall more robust item-specific proportion-congruent effect for low-WM capacity individuals than for high WM-capacity individuals is intriguing, it is important to emphasize that this result comes from a design that is rather peculiar in that contingency learning, list-wide proportion congruency, and item-specific proportion congruency were all manipulated in a single experiment. Such a situation is rather different from those traditionally used to study item-specific proportion-congruent effects. Therefore, it will be helpful to know whether a WM-capacity difference in the item-specific proportion-congruent effect would emerge in a manipulation that is closer to the original paradigm showing an item-specific proportion-congruent effect, that is, one in which congruent and incongruent items are equally probable in the list (Jacoby et al., 2003) (which is the type of design employed in the present experiments). Thus, another objective of Experiments 3A and 3B was to determine whether Hutchison's (2011) WM-capacity differences in the item-specific proportion-congruent effect are indeed the norm by evaluating them in the context of Jacoby et al.'s (2003) item-specific proportion-congruent design.

To address this question, 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. It was hypothesized that in the no-load condition in the Stroop task (Experiment 3B) individuals with lower WM capacity would show a more pronounced pattern of item-specific proportion-congruent effects than would individuals with higher WM capacity, paralleling the results reported by Hutchison (2011). This difference was expected to be more prominent in the error rates than in the latencies, as Hutchison also reported, because, as discussed, 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).

The pattern of results expected from the DMC account (as just described) does appear to differ considerably from the one that would be expected from the contingency learning account. 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, similar to participants in general performing it with a taxing concurrent task, 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. Crucially, according to this logic, the same should be true of any effect that is allegedly due to the contingency learning process, e.g., the item-specific proportion-congruent effect. In other words, this account would predict that the item-specific proportion-congruent effect, being based on contingency learning, should be smaller for low WM-capacity participants than for high WM-capacity participants. As noted, the reviewed data from Hutchison (2011) showed the opposite tendency (i.e., a regular item-specific proportion-congruent effect in low WM-capacity individuals and a null item-specific proportion-congruent effect in high WM-capacity individuals), although only in the error rates.

In a similar vein, it is also worth noting that there was no evidence in Hutchison's (2011) data that learning of contingencies between color words and responses to colors on incongruent trials (e.g., faster responses to the word RED when appearing in its usual (incongruent) black color compared to when it appeared in its unusual (incongruent) yellow color) were reduced for low WM-capacity individuals. 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, $F(1, 84) = 14.13$, $MSE = 25814$, $p < .001$, $\eta_p^2 = .144$, but not in the error rates, $F < 1$. Most importantly, contingency learning did not interact with WM-capacity in either analysis (both Fs < 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). (note 7) Therefore, Hutchison's results offer no support for the hypothesis that these types of contingency learning effects decrease with decreasing WM resources as indexed by an individual's WM capacity, as would be expected according to a contingency learning account.

This null result should be taken with caution, however, as it comes from an experiment in which vocal responses to colors were used. Vocal responding 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 use of manual responses, implemented in Experiment 3A (and 3B), provides a better way of determining if 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 (non-conflict 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 was half 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).

Procedure

Participants completed Experiments 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 (non-conflict 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 successive 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. (note 8)

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 correct responses in the color identification task. 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.

Table 7.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Experiment 3A – Manual Non-conflict Color Identification Task

Contingency	RTs	Error rates
No load		
High	665 (7)	2.1 (.2)
Low	728 (10)	3.6 (.4)
Contingency effect	63	1.5
Low load		
High	783 (18)	2.8 (.4)
Low	814 (18)	3.9 (.8)
Contingency effect	31	1.1
High load		
High	846 (25)	3 (.4)
Low	863 (23)	2.7 (.7)
Contingency effect	17	-.3

Table 8.

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

	RTs			Error rates		
Congruency	Mostly-congruent items	Mostly-incongruent items	Item type effect	Mostly-congruent items	Mostly-incongruent items	Item type effect
<u>No load</u>						
Congruent	699 (10)	771 (14)	72	1.8 (.3)	2.9 (.6)	1.1
Incongruent	885 (14)	829 (13)	-56	6.1 (.8)	4.1 (.4)	-2
Congruency Effect	186	58	-128	4.3	1.2	-3.1
<u>Low load</u>						
Congruent	825 (24)	878 (24)	53	2.2 (.5)	3.3 (1)	1.1
Incongruent	1012 (26)	948 (21)	-64	7.4 (1.4)	4.8 (.7)	-2.6
Congruency Effect	187	70	-117	5.2	1.5	-3.7
<u>High load</u>						
Congruent	869 (26)	920 (28)	51	2.6 (.4)	2.3 (.9)	-.3
Incongruent	1022 (24)	974 (21)	-48	4.8 (1.1)	3.8 (.7)	-1
Congruency Effect	153	54	-99	2.2	1.5	-.7

In addition, similar to Hutchison (2011), an extreme-groups analysis was performed for the no-load condition participants comparing performance for low and high WM-capacity individuals. For this analysis, twenty-eight participants were removed from the no-load group because their accuracy on the distractor component of one or more of the complex span tasks was below 75%. (note 9) For each of the remaining 98 participants, a single composite score was computed from the average of the partial scores obtained in each of the three complex span tasks. A quartile split was then conducted on this composite score. 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. Again, 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 9 and 10 for Experiments 3A and 3B, respectively, without splitting the data by Order.

Table 9.

Mean RTs and Percentage Error Rates (and Corresponding Standard Errors) for Low and High WM-Capacity groups in Experiment 3A – Manual Non-conflict 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 10.

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	Item type effect	Mostly-congruent items	Mostly-incongruent items	Item type effect
Low WM capacity						
Congruent	702 (25)	761 (32)	59	2.7 (.8)	3.9 (1.1)	1.2
Incongruent	918 (35)	844 (30)	-74	11.2 (2.4)	5.8 (1.2)	-5.4
Congruency Effect	216	83	-133	8.5	1.9	-6.6
Medium-low WM capacity						
Congruent	742 (23)	817 (31)	75	1.5 (.6)	1.3 (.8)	-.2
Incongruent	902 (34)	876 (29)	-26	4.8 (1.4)	2.7 (.8)	-2.1
Congruency Effect	160	59	-101	3.3	1.4	-1.9
Medium-high WM capacity						
Congruent	660 (19)	750 (30)	90	1.1 (.5)	4.2 (2.1)	3.1
Incongruent	834 (27)	798 (25)	-36	5.7 (1.6)	3.4 (1)	-2.3
Congruency Effect	174	48	-126	4.6	-.8	-5.4

<u>High WM capacity</u>						
Congruent	666 (21)	721 (27)	55	1.5 (.6)	.7 (.5)	-.8
Incongruent	874 (36)	777 (25)	-97	3.5 (1.2)	3.3 (.8)	-.2
Congruency Effect	208	56	-152	2	2.6	.6

In addition to the extreme-groups analysis, a procedure that involves arbitrarily splitting the sample and removing half of that sample from the analyses, another analysis was conducted using all 98 individuals in the no-load group who maintained an accuracy on the distractor component of each of the complex span tasks above the 75% threshold. This analysis was conducted using mixed-effects modelling, a type of analysis which permits use of both continuous and categorical variables, even in repeated-measures designs (the fixed effects), while controlling for variance among the participants and the items being used (the random effects; Baayen, 2008; Baayen, Davidson, & Bates, 2008). Specifically, 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, 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) were entered as fixed effects. 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. In other words, in these analyses, rather than comparing low vs. high WM-capacity individuals, the full range of WM capacity sampled was analyzed using the composite score from the complex span tasks as a continuous index of participants' WM capacity (the higher the score, the higher the WM capacity; for similar analyses in the context of Stroop and Stroop-like tasks, see Meier & Kane, 2013, 2015).

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 generalized linear mixed-effects model. The model was 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 one million iterations (Bolker, 2019). The ggplot2 package, version 3.1.0 (Wickham, 2016), was used to generate graphs.

Finally, note that in the latency analysis, we used a generalized linear mixed-effects model (GLMM) instead of a (more commonly used) linear mixed-effects model (LMM) because generalized linear models, unlike linear models, do not assume a normally distributed dependent variable. Therefore, generalized linear models can accommodate the typically positively skewed distribution of RT data with no need to apply nonlinear transformations to normalize those data. These transformations, often applied when using linear mixed-effects models, have the downside of systematically altering the pattern and size of interaction terms, making it difficult to interpret those terms (Balota, Aschenbrenner, & Yap, 2013; Lo & Andrews, 2015). A Gamma distribution was used to fit the raw RTs, with an identity link between fixed effects and the dependent variable (Lo & Andrews, 2015). The R scripts used to perform the analyses are available at https://osf.io/rtnw2/.

Experiment 3A (non-conflict color identification task)

WM load analysis

RTs. There were main effects of Contingency (high-contingency faster than low-contingency), $F(1, 202) = 49.20$, $MSE = 109472$, $p < .001$, $\eta_p^2 = .196$, Order (overall faster latencies for participants who performed Experiment 3A following Experiment 3B than for participants who performed Experiment 3A first), $F(1, 202) = 9.42$, $MSE = 206034$, $p = .002$, $\eta_p^2 = .045$, and WM Load, $F(2, 202) = 40.39$, $MSE = 883268$, $p < .001$, $\eta_p^2 = .286$. 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, $F(2, 202) = 9.10$, $MSE = 20243$, $p < .001$, $\eta_p^2 = .083$. 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 group (31 ms), $F(1, 165) = 7.97$, $MSE = 17504$, $p = .005$, $\eta_p^2 = .046$, and the high-load group (17 ms), $F(1, 161) = 14.89$, $MSE = 32173$, $p < .001$, $\eta_p^2 = .085$, whereas low- and high-load groups did not significantly differ, $F(1, 78) = .76$, $MSE = 1828$, $p = .387$, $\eta_p^2 = .010$. 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, $F(1, 202) = 2.98$, $MSE = 6639$, $p = .086$, $\eta_p^2 = .015$, 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), $F(1, 202) = 6.08$, $MSE = .005$, $p = .014$, $\eta_p^2 = .029$. 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, $F(2, 202) = 2.77$, $MSE = .002$, $p = .065$, $\eta_p^2 = .027$, and the Bayes Factor indicated no real preference for either the model with the interaction or the model without it, $BF_{01} = 1.15$.

WM capacity analysis – extreme-groups ANOVA

RTs. Contingency (high-contingency faster than low-contingency) was the only significant effect, $F(1, 44) = 50.36$, $MSE = 100277$, $p < .001$, $\eta_p^2 = .534$. Although individuals with high WM Capacity were numerically faster than individuals with low WM Capacity, WM Capacity was not statistically significant, $F(1, 44) = .52$, $MSE = 6573$, $p = .47$, $\eta_p^2 = .012$. In addition, WM Capacity did not interact with Contingency, $F(1, 44) = .07$, $MSE = 131$, $p = .80$, $\eta_p^2 = .001$, indicating equivalent contingency learning effects for the low (63 ms) and the high (67 ms) WM Capacity groups. Bayesian analyses revealed that there was “moderate” evidence for the absence of the interaction, $BF_{01} = 3.47$.

Error rates. There was a main effect of Contingency (high-contingency more accurate than low-contingency), $F(1, 44) = 18.52$, $MSE = .018$, $p < .001$, $\eta_p^2 = .296$, and WM Capacity (high WM-capacity group more accurate than low WM-capacity group), $F(1, 44) = 7.66$, $MSE = .012$, $p = .008$, $\eta_p^2 = .148$. There was no interaction between Contingency and WM Capacity, however, $F(1, 44) = .84$, $MSE = .001$, $p = .37$, $\eta_p^2 = .019$, indicating that the contingency learning effects for the low (3.4%) and high (2.1%) WM Capacity groups were equivalent. In the Bayesian analyses, the evidence for the absence of this interaction, however, was only "anecdotal", $BF_{01} = 2.33$.

WM capacity analysis – full-sample GLMM

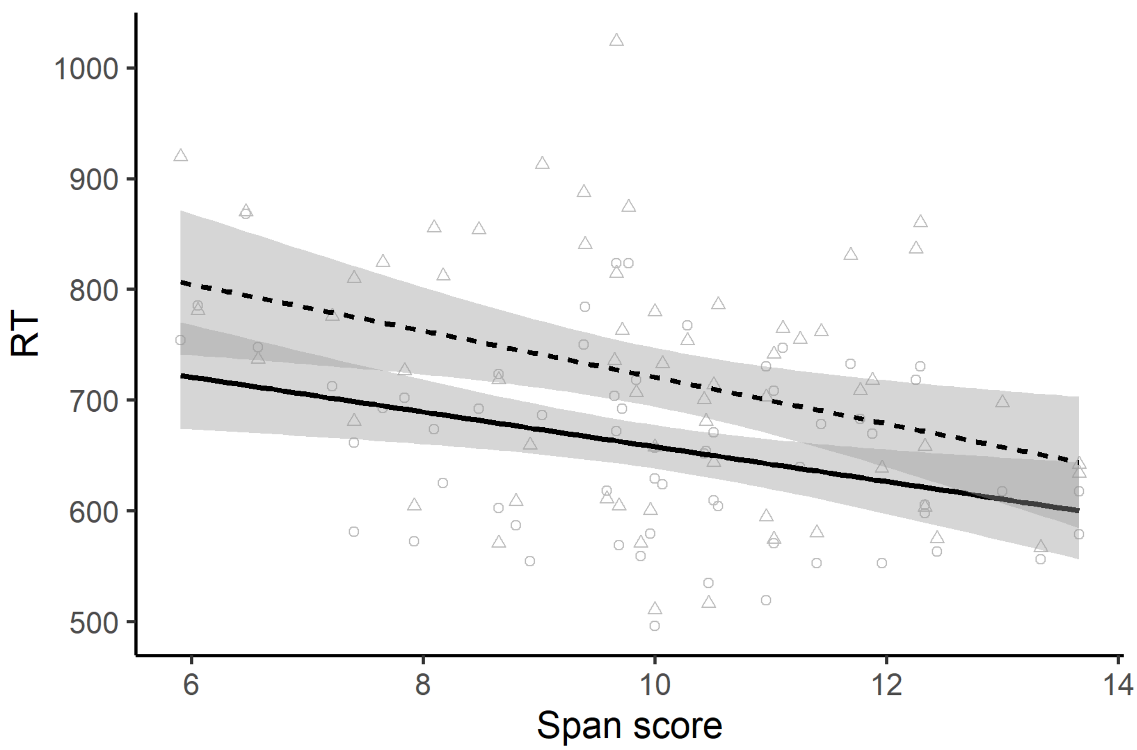
RTs. There were main effects of Contingency (high-contingency faster than low-contingency), $ß = -31.60$, $SE = 1.94$, $z = -16.30$, $p < .001$, Order (participants who performed Experiment 3A following Experiment 3B were faster than participants who performed Experiment 3A first), $ß = -9.11$, $SE = 3.31$, $z = -2.77$, $p = .006$, and Span Score (latencies decreased with higher scores), $ß = -15.08$, $SE = 3.37$, $z = -4.48$, $p < .001$. There was also an interaction between Order and Span Score, $ß = -17.08$, $SE = 4.14$, $z = -4.12$, $p < .001$, and a three-way interaction between Contingency, Order, and Span Score, $ß = 6.82$, $SE = 1.99$, $z = 3.47$, $p < .001$.

To explore these interactions, the data were split by Order and analyzed separately. For participants performing Experiment 3A first, there were main effects of both Contingency (high-contingency faster than low-contingency), $ß = -29.02$, $SE = 2.39$, $z = -12.13$, $p < .001$, and Span Score (latencies decreased with higher scores), $ß = -33.59$, $SE = 3.84$, $z = -8.75$, $p < .001$, as well as a Contingency by Span Score interaction, $ß = 5.77$, $SE = 2.39$, $z = 2.29$, $p = .022$. This interaction indicated that contingency learning effects tended to diminish with higher Span

Score and 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.

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

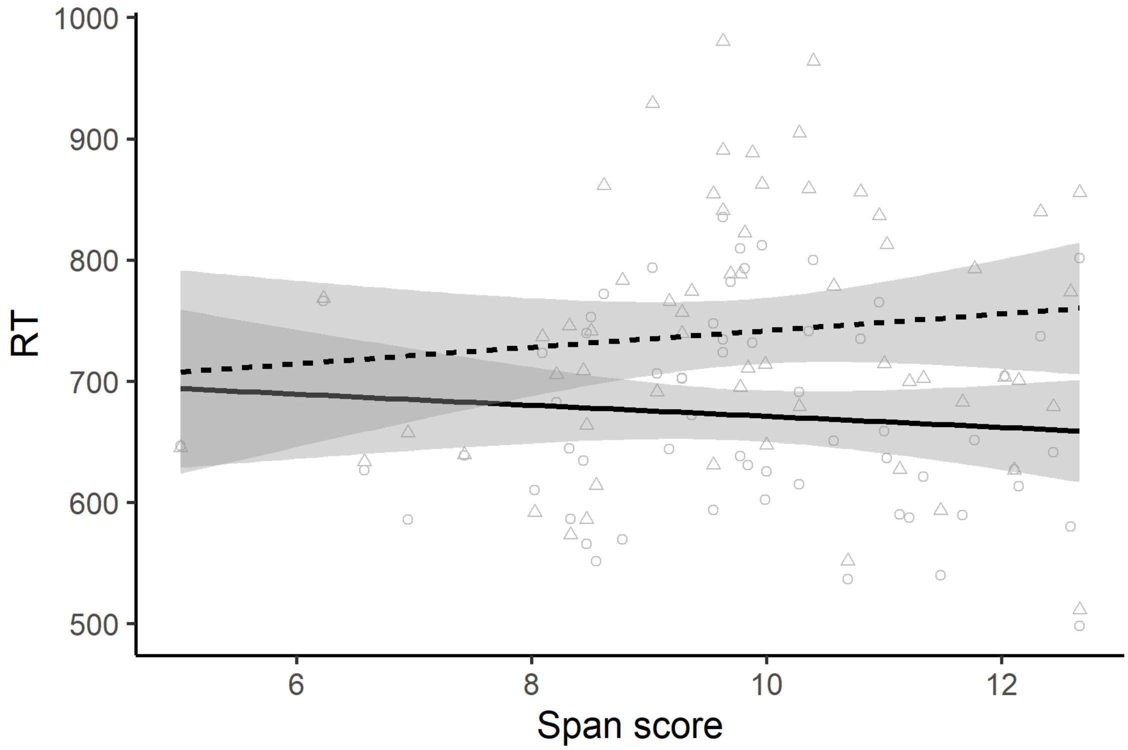


Note. For each participant, the mean latency for high- and low-contingency items is marked with a circle and a triangle, respectively. Regression slopes (with 95% confidence interval bands) for high- and low-contingency items are marked with a solid line and a dashed line, respectively.

For participants performing Experiment 3A following Experiment 3B, there was the usual effect of Contingency, *ß* = -33.28, *SE* = 2.67, *z* = -12.48, *p* < .001, but not of Span Score, *ß* = 1.92, *SE* = 5.54, *z* = .35, *p* = .73. Contingency and Span Score interacted in this case as well, *ß* = -7.74, *SE* = 2.85, *z* = -2.71, *p* = .007. However, as represented in Figure 2, the pattern of this interaction was the opposite of that found in participants who did Experiment 3A first: Here, the contingency learning effect tended to *increase* with higher Span Score (and Span Score did not reduce latencies overall).

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

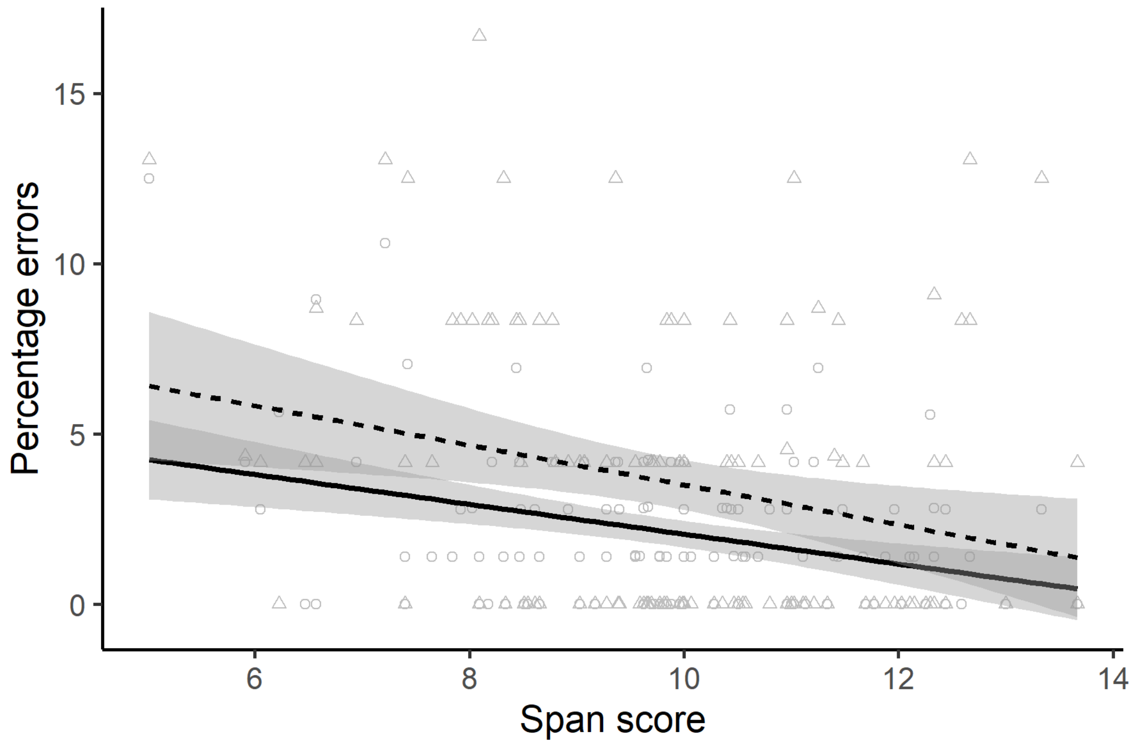


Note. For each participant, the mean latency for high- and low-contingency items is marked with a circle and a triangle, respectively. Regression slopes (with 95% confidence interval bands) for high- and low-contingency items are marked with a solid line and a dashed line, respectively.

Errors. The only significant effects were those of Contingency (high-contingency more accurate than low-contingency), *ß* = .29, *SE* = .06, *z* = 4.63, *p* < .001, and Span Score (accuracy increased with higher scores), *ß* = .31, *SE* = .08, *z* = 3.77, *p* < .001. This general pattern is represented in Figure 3, with higher error rates for low-contingency than high-contingency items and regression lines going down for both items types with higher span scores. There was, however, also a marginal three-way interaction between Contingency, Order, and Span Score, *ß* = -.11, *SE* = .06, *z* = -1.87, *p* = .062, indicating a numerical tendency for a pattern similar to that found in the latencies: Contingency learning effects tended to decrease with higher scores for participants who did Experiment 3A first but they tended to increase with higher scores for participants who did Experiment 3A following Experiment 3B.

Figure 3.

The Impact of Span Score on the Contingency learning Effect in Error Rates for No-Load Participants in Experiment 3A



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.

Experiment 3B (Stroop task)

WM load analysis

RTs. There were main effects of Congruency (congruent faster than incongruent), $F(1, 202) = 301.62$, $MSE = 2222028$, $p < .001$, $\eta_p^2 = .599$, Order (overall faster latencies for participants who performed Experiment 3B following Experiment 3A than for participants who performed Experiment 3B first), $F(1, 202) = 7.08$, $MSE = 445799$, $p = .008$, $\eta_p^2 = .034$, and WM Load, $F(2, 202) = 28.63$, $MSE = 1802531$, $p < .001$, $\eta_p^2 = .221$. 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 the low-load and high-load groups did not differ significantly from one another ($p = .513$). There was an interaction between Congruency and Order, $F(1, 202) = 7.15$, $MSE = 52661$, $p < .001$, $\eta_p^2 = .034$, 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, $F(1, 202) = 98.41$, $MSE = 521158$, $p < .001$, $\eta_p^2 = .328$, indicating that a regular item-specific proportion-congruent effect was found, 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, $F(2, 202) = .63$, $MSE = 3327$, $p = .54$, $\eta_p^2 = .006$, 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} = 13.09.

Error rates. There were main effects of Congruency (congruent more accurate than incongruent), $F(1, 202) = 39.76$, $MSE = .108$, $p < .001$, $\eta_p^2 = .164$, and Order (participants who did Experiment 3B following Experiment 3A were overall more accurate than those who did Experiment 3B first), $F(1, 202) = 4.41$, $MSE = .025$, $p = .037$, $\eta_p^2 = .021$. Congruency also interacted with Item Type, $F(1, 202) = 8.63$, $MSE = .024$, $p = .004$, $\eta_p^2 = .041$, indicating that congruency effects were larger for mostly-congruent items (4.1%) than for mostly-incongruent items (1.3%). This item-specific proportion-congruent effect was not modulated by WM Load, as no three-way interaction was found between Congruency, Item Type, and WM Load, $F(2, 202) = .923$, $MSE = .003$, $p = .40$, $\eta_p^2 = .009$. Once again, in the Bayesian analyses, there was “moderate” evidence in support of the model without the interaction, BF_{01} = 6.35.

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, $F(2, 202) = 5.28$, $MSE = .015$, $p = .023$, $\eta_p^2 = .025$. 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%), $F(1, 102) = 10.64$, $MSE = .039$, $p = .002$, $\eta_p^2 = .094$. In contrast, no significant item-specific proportion-congruent effect was observed for participants who did Experiment 3B after Experiment 3A (congruency effect for mostly-congruent items: 3%;

congruency effect for mostly-incongruent items: 1.8%), $F(1, 100) = .30$, $MSE = .001$, $p = .59$, $\eta_p^2 = .003$.

WM capacity analysis – extreme-groups ANOVA

RTs. There was a main effect of Congruency (congruent faster than incongruent), $F(1, 44) = 179.64$, $MSE = 940783$, $p < .001$, $\eta_p^2 = .803$, and an interaction between Congruency and Item Type, $F(1, 44) = 39.57$, $MSE = 238229$, $p < .001$, $\eta_p^2 = .473$. The interaction indicated that, as usual, there was 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, $F(1, 44) = 1.44$, $MSE = 89535$, $p = .237$, $\eta_p^2 = .032$. 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, $F(1, 44) = .33$, $MSE = 1966$, $p = .571$, $\eta_p^2 = .007$. The Bayes Factor, $BF_{01} = 3.19$, 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, $F(1, 44) = 3.98$, $MSE = 20840$, $p = .052$, $\eta_p^2 = .083$. 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), $F(1, 44) = 21.96$, $MSE = .069$, $p < .001$, $\eta_p^2 = .333$, WM Capacity (high WM-capacity group more accurate than low WM-capacity group), $F(1, 44) = 15.56$, $MSE = .070$, $p < .001$, $\eta_p^2 = .261$, and a marginal effect of Item Type, $F(1, 44) = 3.50$, $MSE = .010$, $p = .068$, $\eta_p^2 = .074$, indicating a tendency for mostly-incongruent items to be more accurate than mostly-congruent items. WM Capacity interacted with Order, $F(1, 44) = 5.51$, $MSE = .025$, $p = .023$, $\eta_p^2 = .111$, 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, $F(1, 44) = 3.73$, $MSE = .012$, $p = .060$, $\eta_p^2 = .078$. Congruency marginally interacted with WM Capacity as well, $F(1, 44) = 3.88$, $MSE = .012$, $p = .055$, $\eta_p^2 = .081$, 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, $F(1, 44) = 5.25$, $MSE = .017$, $p = .027$, $\eta_p^2 = .107$.

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, $F(1, 22) = 14.80$, $MSE = .070$, $p = .001$, $\eta_p^2 = .402$, and a Congruency by Item Type interaction, $F(1, 22) = 5.42$, $MSE = .029$, $p = .030$, $\eta_p^2 = .198$. 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, $F(1, 22) = 7.34$, $MSE = .012$, $p = .013$, η_p^2

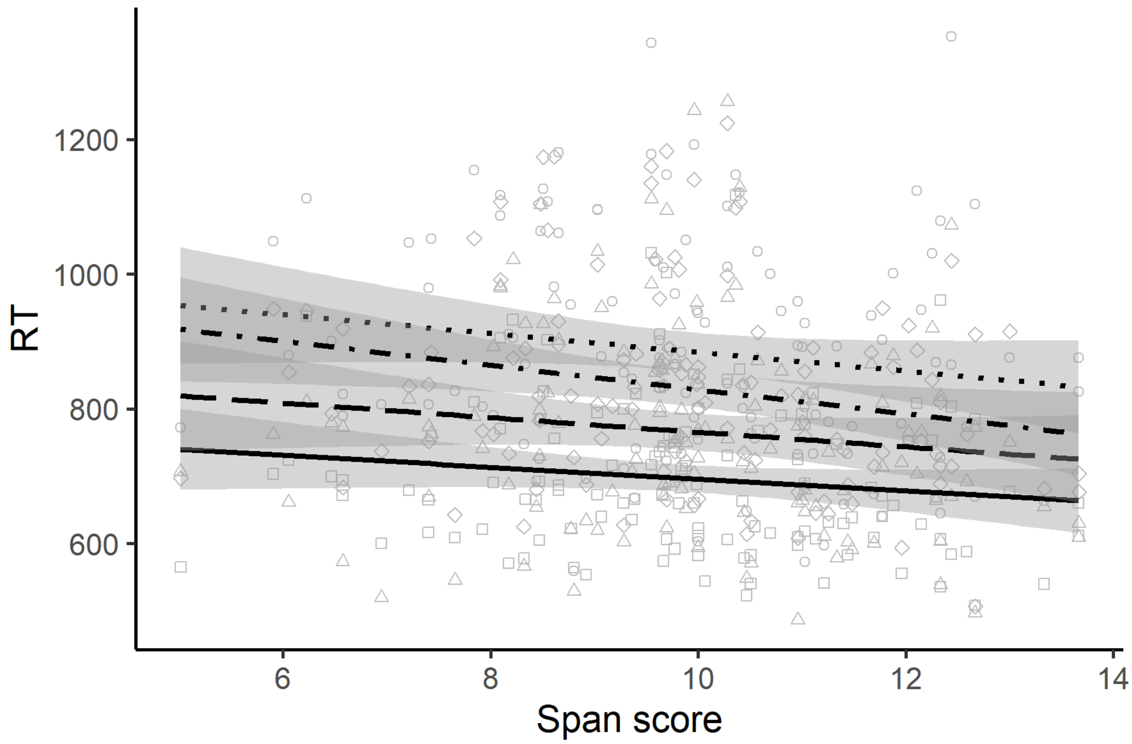
= .250, with no evidence of an item-specific proportion-congruent effect, *F*(1, 22) = .18, *MSE* = .000, *p* = .674, η_p^2 = .008. Indeed, the congruency effect for mostly-congruent items (2%) was slightly smaller than the congruency effect for mostly-incongruent items (2.6%).

WM capacity analysis – full-sample GLMM

RTs. There were main effects of Congruency (congruent faster than incongruent), *ß* = -61.00, *SE* = 1.84, *z* = -33.18, *p* < .001, Order (overall faster latencies for participants who did Experiment 3B following Experiment 3A than for participants who did Experiment 3B first), *ß* = -42.43, *SE* = 3.27, *z* = -12.97, *p* < .001, and Span Score (latencies decreased with higher scores), *ß* = -18.42, *SE* = 2.80, *z* = -6.58, *p* < .001. Congruency interacted with Item Type, *ß* = -29.83, *SE* = 1.82, *z* = -16.41, *p* < .001, indicating a regular item-specific proportion-congruent effect. Congruency also interacted with Span Score, *ß* = 4.77, *SE* = 2.11, *z* = 2.26, *p* = .024, indicating that, overall, congruency effects tended to decrease with higher scores. However, there was no three-way interaction between Congruency, Item Type, and Span Score, *ß* = -1.63, *SE* = 1.93, *z* = -.85, *p* = .40, suggesting that the item-specific proportion-congruent effect, overall, did not change across the range of scores, a pattern represented in Figure 4. 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 if latencies diminish and all lines tend to come together, indicating reduced congruency effects with higher scores.

Figure 4.

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.

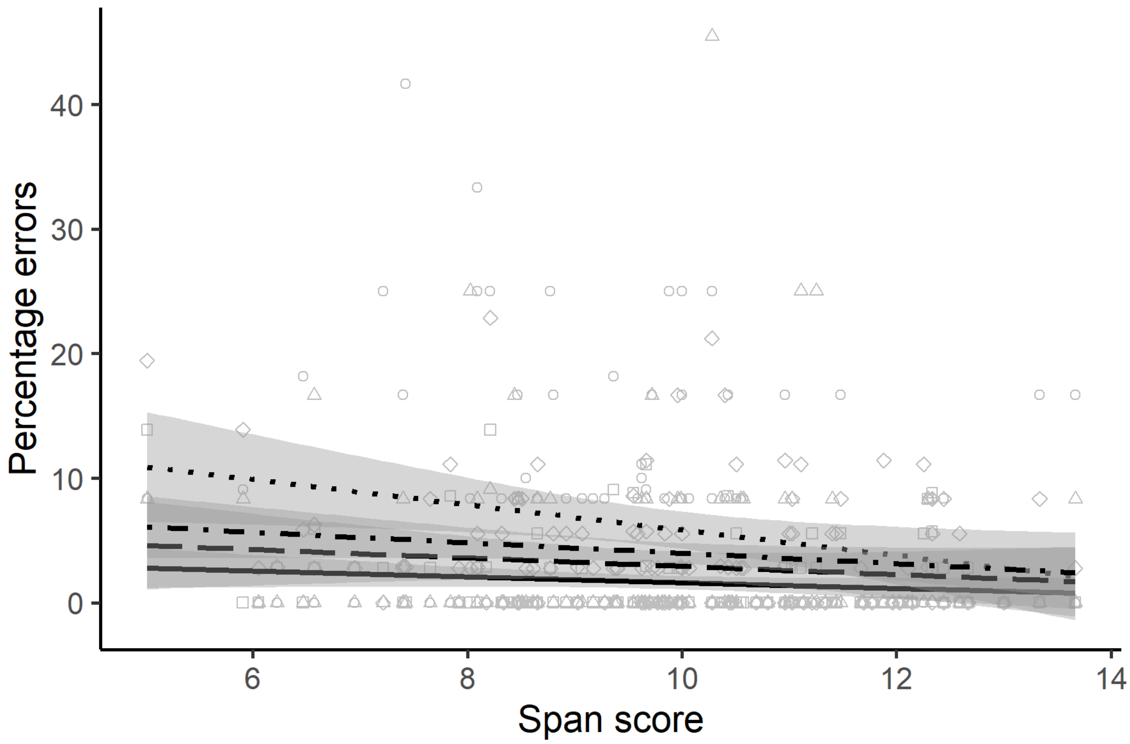
On the other hand, the impact of Span Score on overall performance and the item-specific proportion-congruent effect was modulated by Order. Specifically, there was an interaction between Order and Span Score, *ß* = -16.32, *SE* = 2.86, *z* = -5.71, *p* < .001, a three-way interaction between Congruency, Order, and Span Score, *ß* = 5.35, *SE* = 1.77, *z* = 3.02, *p* = .003, and a marginal four-way interaction between Congruency, Item Type, Order, and Span Score, *ß* = 3.90, *SE* = 1.99, *z* = 1.96, *p* = .050. To explore these interactions, the data were split by Order and analyzed separately. For participants performing Experiment 3B first, the only significant effects were the main effect of Congruency, *ß* = -57.84, *SE* = 2.93, *z* = -19.78, *p* < .001, and the Congruency by Item Type interaction, *ß* = -31.57, *SE* = 3.11, *z* = -10.16, *p* < .001, reflecting a regular item-specific proportion-congruent effect. There was also a numerical tendency for this item-specific proportion-congruent effect to increase with higher Span Score, however, the three-way interaction between Congruency, Item Type, and Span Score reflecting this tendency did not reach significance, *ß* = -5.23, *SE* = 3.14, *z* = -1.67, *p* = .096. For participants performing Experiment 3B following Experiment 3A, in addition to the main effect of Congruency, *ß* = -63.83, *SE* = 2.65, *z* = -24.10, *p* < .001, Span Score also had an effect, *ß* = -36.14, *SE* = 4.40, *z* = -8.22, *p* < .001, with higher scores leading to faster latencies overall. There was also an interaction between Congruency and Item Type (a regular item-specific proportion-congruent effect), *ß* = -27.60, *SE* = 2.69, *z* = -10.27, *p* < .001, and an interaction between Congruency and Span Score, *ß* = 10.69, *SE* = 2.66, *z* = 4.02, *p* < .001, with the last interaction indicating decreasing congruency effects with higher scores. There was, however, no indication that Span Score modulated the item-specific proportion-congruent effect, i.e., there was no three-way interaction between Congruency, Item Type, and Span Score, *ß* = 2.27, *SE* = 2.72, *z* = .83, *p* = .40.

The absence of a significant three-way interaction between Congruency, Item Type, and Span Score in both order versions (Experiment 3B first and Experiment 3B following Experiment 3A) suggests that although Order did appear to modulate the direction of the relation between the item-specific proportion-congruent effect and Span Score (the item-specific proportion-congruent effect tended to increase with higher scores for participants who did Experiment 3B first whereas it tended to decrease for participants who did Experiment 3B following Experiment 3A), there was little evidence in either order version that this relation, in one direction or the other, actually existed.

Errors. There were main effects of Congruency (congruent more accurate than incongruent), $ß = .47$, $SE = .06$, $z = 7.18$, $p < .001$, and Span Score (fewer errors with higher scores), $ß = .28$, $SE = .10$, $z = 2.92$, $p = .003$. There was also an interaction between Congruency and Item Type, $ß = .27$, $SE = .06$, $z = 4.39$, $p < .001$, reflecting a regular item-specific proportion-congruent effect. However, this item-specific proportion-congruent effect was not modulated by Span Score, i.e., there was no three-way interaction between Congruency, Item Type, and Span Score, $ß = -.02$, $SE = .06$, $z = -.27$, $p = .79$. The relation between Congruency, Item Type, and Span Score is represented in the scatterplot in Figure 5. 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.

Figure 5.

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.

Finally, Span Score interacted with Order, *ß* = -.21, *SE* = .10, *z* = -2.19, *p* = .029. Splitting the data by Order revealed that this interaction emerged because higher span reduced errors in participants who did Experiment 3B first, *ß* = .47, *SE* = .15, *z* = 3.19, *p* = .001, but not in participants who did Experiment 3B following Experiment 3A, *ß* = .07, *SE* = .12, *z* = .59, *p* = .55.

<u>Discussion</u>

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 non-conflict 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 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 extreme-groups comparison, there was no evidence that the process of contingency learning in the non-conflict color identification task (Experiment 3A) was any different in low WM-capacity individuals compared to high WM-capacity individuals. In contrast, a difference between low and high WM-capacity individuals did emerge in item-specific proportion-congruent effects in the Stroop task (Experiment 3B). Specifically, in the Stroop task, high WM-capacity participants showed an item-specific proportion-congruent effect in their latencies but not in their error rates, whereas low WM-capacity participants showed a clear item-specific proportion-congruent effect in both dependent measures. This pattern of findings is virtually identical to that reported by Hutchison (2011), suggesting that the peculiarities of Hutchison's experiment (e.g., the non-standard design for examining contingency learning and item-specific proportion-congruent effects, the simultaneous list-wide proportion-congruent manipulation) and the response modality he employed (i.e., vocal responding to colors) had little or no role in producing his results.

A more complicated picture emerged in the full-sample analysis, however, an analysis in which a continuous measure of WM capacity was used. In this analysis, the order in which the two tasks (the non-conflict color identification task and the Stroop task) were completed had an important role in modulating the relation between WM capacity and contingency learning in the non-conflict color identification task. For participants who completed the non-conflict color identification task first, higher WM capacity was associated with smaller contingency learning effects. For participants who completed the non-conflict color identification task following the Stroop task, the opposite pattern was found, with higher WM capacity leading to larger contingency learning effects. In sum, unlike in the extreme-groups analysis, there was some evidence in this analysis for a relation between WM capacity and contingency learning effects, although this relation does not appear to be as straightforward as the relation between WM load and contingency learning appears to be.

The results from the full-sample analysis of the Stroop task also differed from those from the extreme-groups analysis in one important way. Specifically, there was no evidence in the full-sample analysis that WM capacity modulated item-specific proportion-congruent effects in the error rates. In the latencies, there was a tendency for the item-specific proportion-congruent effect to increase with higher WM capacity for participants who completed the Stroop task first and an opposite tendency for that effect to decrease with higher WM capacity for participants who completed the Stroop task following the non-conflict color identification task. However, neither of these tendencies was significant, suggesting that in the latencies as well as in the error rates, overall, item-specific proportion-congruent effects did not vary as a function of WM capacity.

In sum, the results of the WM-load analysis of Experiments 3A and 3B, similar to the results of Experiments 2A and 2B, clearly indicate that contingency learning may not be the sole process underlying the item-specific proportion-congruent effect: Reducing WM resources by means of a concurrent WM load reduces the contingency learning effect in the non-conflict color identification task but not the item-specific proportion-congruent effect in the Stroop task. In contrast, the results of the WM-capacity analyses of Experiments 3A and 3B seem to suggest that, overall, inter-individual variability in WM resources may not have an especially strong impact on either contingency learning or adaptation to item-specific conflict frequency, although additional factors, such as practice effects associated with the order in which tasks were performed, may have an important role in these patterns, a point to which we return in the General Discussion.

General Discussion

The item-specific proportion-congruent effect: Does it 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 support the existence of a mechanism of adaptation to conflict frequency (Schmidt, 2013b). This contingency learning 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 the present research, non-conflict 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 non-conflict 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. Specifically, 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 those results, however, is merely that the vocal responding procedure fails to replicate Schmidt et al.'s (2010) original finding in the non-conflict 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.

Manual responding to colors (plus the addition of feedback on each trial), however, not only increased baseline contingency learning effects in the non-conflict 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). Importantly, this pattern was obtained even when the same participants were tested in both the non-conflict task and the Stroop task (Experiments 3A and 3B). (note 10)

An aspect of our WM load manipulation that should be noted is that in no case did concurrent WM 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 not effective. However, although other load manipulations might have been possible (e.g., an *n*-back task; see also footnote 5), the goal that our manipulation was required to achieve was to impair contingency learning, i.e., the critical process that supposedly underlies the item-specific proportion-congruent effect in the Stroop task (according to the contingency learning account of this effect, that is). To the extent that this goal was achieved (as demonstrated by reduced contingency learning effects under load in a task, the non-conflict color identification task, in which contingency learning was unambiguously the only process being engaged), the fact that our load manipulation spared the basic congruency effect in the Stroop task does not appear to be at all problematic.

Indeed, the fact that our load manipulation selectively impaired contingency learning (i.e., and not the overall congruency effect) may have its own merits. A load manipulation leading to an increased congruency effect would seem to imply that, under a high load, individuals would experience higher conflict overall. What would be possible, then, is that individuals would apply different strategies when dealing with that high conflict than when dealing with the lower conflict experienced in a normal situation (i.e., with no concurrent WM load). For example, an item-specific conflict adaptation mechanism may be preferred when dealing with high conflict (with a high concurrent WM load) whereas a contingency learning mechanism may be preferred when dealing with lower conflict (with low or no load). As a result, whatever result is obtained when maintaining a WM load would tell little about the processes involved in the item-specific proportion-congruent effect in normal circumstances (i.e., when no WM load is maintained). Clearly, this problem does not arise in a load procedure that selectively impairs contingency learning such as ours, because the level of Stroop conflict experienced at all load levels would be roughly the same. Thus, in this type of situation, drawing inferences about the processes normally involved in the item-specific proportion-congruent effect based on the pattern of results obtained with a concurrent WM load appears more justified and straightforward.

In sum, the overall pattern of results that we obtained 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. Note that the fact that contingency learning and item-specific proportion-congruent effects were examined in two different tasks – a non-conflict color identification task and the Stroop task, respectively – has little relevance for the contingency learning account. According to this account, contingency learning is a general cognitive process that has nothing to do with conflict (Schmidt et al., 2007; Schmidt & Besner, 2008) and functions with conflicting and non-conflicting stimuli in the same way (Levin & Tzelgov, 2016). As such, one cannot simply attribute the different patterns observed in the non-conflict and the Stroop task to the nature of the words (noncolor vs. color) that those tasks employ.

A Dual-Mechanisms-of-Control account of the item-specific proportion-congruent effect

An explanation that better accommodates the present results is one that assumes that 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 setting, with the mostly-congruent words leading to relaxed attention (as the irrelevant dimension is typically not conflicting) and the mostly-incongruent words leading 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 strategy. 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 strategy 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 control strategies, that is, effortful strategies that involve sustained maintenance of task goals. In other situations, increasing WM load may, instead, bias individuals to use reactive control strategies, that is, strategies 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 strategy (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.

Because the evidence supporting the DMC framework comes not only from studies investigating differences in WM resources that are induced experimentally (e.g., by use of a concurrent WM load) but also from studies investigating differences in WM resources that occur naturally across individuals (i.e., their WM capacity; Braver, 2012; see also Kane & Engle, 2003), the WM-capacity analyses of Experiment 3B are potentially relevant in evaluating these conclusions. Unfortunately, however, those analyses only offered partial support to the idea that adaptation to item-specific conflict frequency in the Stroop task is the dominant strategy when WM resources are reduced. This partial support comes from the extreme-groups analysis. This analysis, consistent with previous findings (Hutchison, 2011), revealed that while both low and high WM-capacity individuals showed an item-specific proportion-congruent effect in latencies, high WM-capacity individuals produced no evidence for adaptation to item-specific conflict frequency in the error rates, a pattern low WM-capacity individuals clearly showed.

Errors in the Stroop task are typically interpreted as an index of participants' inability to successfully maintain the task goal (Kane & Engle, 2003; MacLeod, 1991). From the perspective of the DMC account, it is reasonable to assume that this inability depends, at least in part, on the degree to which individuals rely on proactive control, i.e., the degree to which they successfully maintain the color-naming goal throughout the task. As such, what the extreme-groups analysis of WM capacity would suggest is that, on one hand, both low and high WM-capacity individuals have access to a reactive strategy to adapt to item-specific conflict frequency (as demonstrated by the regular item-specific proportion-congruent effect obtained in the latencies). On the other hand, low and high WM-capacity individuals might differ considerably in their ability to concurrently implement a proactive strategy of goal maintenance. Specifically, because they have insufficient WM resources, low WM-capacity individuals would have limited access to this sort of strategy. As a result, they would be biased to use reactive control as the main strategy in performing the task. Mainly relying on (reactive) adaptation to item-specific conflict frequency, those individuals might be especially vulnerable to inadvertent word reading in a situation that leads to a relaxation of attention, i.e., mostly-congruent words. In this situation, and with no proactive strategy being properly implemented to prevent goal neglect, relaxing attention would cause them to inadvertently read the word in many cases, with the likely result being a word-reading error in some of the infrequent instances in which conflict does occur (i.e., the infrequent incongruent words in the mostly-congruent condition). The implication is that incongruent words in the mostly-congruent condition would be much more problematic than incongruent words in the mostly-incongruent condition, thus resulting in an item-specific proportion-congruent effect in the error rates for low WM-capacity individuals.

The same would not be true for high WM-capacity individuals. Being able to engage in proactive maintenance of the task goal while concurrently adapting to item-specific conflict frequency, those individuals would allow item-specific conflict frequency to influence their performance (as shown by the item-specific proportion-congruent effect in the latencies), but they would not allow words which elicit a relaxation of attention to cause them to inadvertently read the word. The result is that the infrequent incongruent words in the mostly-congruent condition would not produce especially high error rates. In particular, those words would not produce any more errors than the incongruent words in the mostly-incongruent condition, thus resulting in the absence of an item-specific proportion-congruent effect in error rates for high WM-capacity individuals.

Casting some doubt on this interpretation of the data are the results of the full-sample analysis of WM capacity. In this analysis, rather than comparing individuals scoring low and individuals scoring high on our complex span tests, we used the score on those tests as a continuous measure of individuals' WM capacity (Meier & Kane, 2013, 2015). Although, similar to what was found in the extreme-groups analysis, 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 in this analysis that WM capacity had an impact on the item-specific proportion-congruent effect in the latencies or, most importantly, the error rates. That is, in this analysis, unlike in the extreme-groups analysis, increasing WM capacity did reduce errors overall but did not reduce the item-specific proportion-congruent effect.

A potential explanation for the different pattern of results obtained in the two analyses is that individuals located at one (or both) of the ends of the WM-capacity continuum may be somewhat "special". For example, it may be assumed that only individuals with the lowest WM capacity are unable to continuously implement proactive control. Possessing this ability is what would prevent individuals with higher WM capacity from making word-reading errors even when dealing with words (i.e., mostly-congruent words) that induce relaxed attention. However, because individuals with very low WM capacity may be considerably poorer at constantly maintaining proactive control, they would inadvertently read words more easily in general, but especially so when dealing with those mostly-congruent words. Consistent with this idea, individuals in the bottom quartile of WM capacity in Experiment 3B (i.e., the low WM-capacity individuals in the extreme-groups analysis) were those who committed the most errors (11.2%) when responding to incongruent words in the mostly-congruent condition. In comparison, individuals in the two middle quartiles and individuals in the top quartile (i.e., the high WM-capacity individuals in the extreme-groups analysis) produced less than half that number of errors in the same condition (medium-low WM capacity: 4.8%; medium-high WM capacity: 5.7%; high WM capacity: 3.5%; see Table 10). Thus, it would appear that, compared with individuals with the lowest WM capacity, individuals with higher WM capacity are able to minimize word-reading errors, even when dealing with mostly-congruent words. With the small number of errors committed by those individuals, observing an item-specific proportion-congruent effect in all cases may not be possible. In fact, analyses conducted separately for each of the four quartiles (i.e., the two middle quartiles in addition to the bottom and top quartiles analyzed in the extreme-groups analysis) revealed that not only the high WM-capacity group but even the medium-low WM-capacity group failed to show a significant item-specific proportion-congruent effect in the errors, maintaining a low error rate across all the conditions. Indeed, in the present experiments, another situation in which error rates were very low (i.e., the no-load group in Experiment 1B, where errors were less than 1% overall), also failed to produce an item-specific proportion-congruent effect in the errors.

In sum, it is possible that the inefficient application of proactive control in individuals in the bottom quartile (i.e., the low WM-capacity individuals) created a situation in which word-reading errors were frequently made when attention was relaxed (i.e., with mostly-congruent words), resulting in a clear item-specific proportion-congruent effect in the error rates. However, errors quickly descended to the floor in individuals with higher WM-capacity (i.e., the individuals in the top three quartiles) because those individuals were more efficient at applying proactive control in general, creating a situation in which an item-specific proportion-congruent effects in the error rates was less easily observable. The idea that individuals with the lowest WM capacity do not cluster with the rest of the individuals would seem to explain the different results obtained in the extreme-groups and the full-sample analyses reported above. Because individuals with the lowest WM capacity are distinguished from the other individuals, an impact of WM capacity on item-specific proportion-congruent effects in error rates would be easy to observe in an analysis in which those low WM-capacity individuals are separated from other individuals, as in our extreme-groups analysis. In contrast, a WM-capacity difference would be less easily observed in an analysis in which the item-specific proportion-congruent effect is assumed to be monotonically related to a continuous measure of WM capacity, as in our full-sample analysis.

Overall, given the mixed pattern of results obtained, it would appear incautious to draw strong conclusions on the nature of the processes leading to the elimination of the item-specific proportion-congruent effect in errors for high WM-capacity individuals. While the present WM-capacity analyses do show some consistency with the extant literature (Hutchison, 2011; Kane & Engle, 2003), 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 Experiment 3B was relatively underpowered for a WM-capacity analysis, both in terms of the number of items used and the size of the sample tested. Better powered investigations of the relation between WM capacity and adaptation to item-specific conflict frequency appear to be required.

Contingency learning: The roles of response modality and WM capacity

Although the focus of the present research was the nature of the item-specific proportion-congruent effect in the Stroop task, interesting results emerged concerning the contingency learning process in the non-conflict color identification task. First, we replicated the finding that a concurrent memory load impairs individuals' ability to learn stimulus-response contingencies in this type of task, consistent with the idea that capacity-limited resources may be necessary for encoding and retrieving those contingencies (Schmidt et al., 2010). Second, this pattern of results occurred when responses to colors were manual (Experiments 2A and 3A) as in Schmidt et al.'s (2010) original experiments, but not when they were vocal (Experiment 1A). The likely reason for this difference is that, with vocal responses, the contingency learning effect at baseline (i.e., in the no-load condition) was so small that a further reduction would be especially challenging to obtain.

A relevant question that this finding raises is what causes reduced contingency learning effects in vocal versus manual responding, a pattern of results that has also been observed in other recent studies (Forrin & MacLeod, 2017; Spinelli et al., under review). (note 11) This 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 in vocal than manual responding. Indeed, because it may favor word processing, vocal responding was one of the factors that Bugg et al. (2011) considered influential in biasing use of contingency learning in the two-item set design of the item-specific proportion-congruent effect. On the other hand, reduced contingency learning effects for vocal responding do seem consistent 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). 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.

Unfortunately, the present data are insufficient to allow us to conclude that stimulus-response compatibility is the crucial factor in determining the different magnitude of contingency learning effects in vocal versus manual responding. The reason is that, in the present research, vocal responding (Experiment 1A) and manual responding (Experiments 2A and 3A) differed not only in the type of response that was required but also in whether responding was assisted with feedback, which was absent with vocal responses but present with keypress responses in these experiments. Indeed, a more complicated story emerged when we tried to address this concern in a series of non-conflict 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 smaller contingency learning effect for vocal than for manual 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.

The interpretation that we offered for those results is that word-response contingency learning may be reduced in situations that require allocation of attention away from the word dimension. These situations would include those requiring a vocal response (either with or without feedback) and those requiring a manual response without feedback. The reason that attention to the word dimension may be reduced in vocal responding is that words automatically trigger reading, the task they are strongly associated with, which typically produces a phonological code (Monsell, Taylor, & Murphy, 2001; Rogers & Monsell, 1995). Therefore, to perform the color naming task efficiently, participants in a vocal responding situation will typically need to inhibit that automatic reading behavior. As a result, learning of contingencies between words and vocal responses will be harder to achieve, resulting in a modest contingency learning effect. Attention to the word dimension may also be reduced in manual responding without feedback, although for a different reason. This reason is that, in the absence of feedback, individuals may need to allocate a certain amount of limited-capacity attentional resources to the process of making sure that mappings between colors (i.e., the relevant stimulus dimension) and keypress responses are correctly implemented. As a result, fewer attentional resources would be available to learn word-response contingencies, leading to a reduced contingency learning effect in this situation. The situation would be different in manual responding with feedback because, in the presence of feedback, individuals might find it useful to rely on that feedback and relax active monitoring of correct stimulus-response mapping implementation. The result would be more attentional resources available for the contingency learning process and, hence, a larger contingency learning effect.

What this interpretation implies for the present results is that the reason for the smaller baseline contingency learning effect in Experiment 1A (vocal responding without feedback) compared to Experiments 2A and 3A (manual responding with feedback) is that the former experiment put participants in a situation in which contingency learning was relatively inefficient because vocal responding reduced the attention that was paid to the word dimension (in order to help suppress the word reading response). This was not the case for the latter experiments, in which attention to the word dimension did not have to be reduced as much because 1) there was no strong need to inhibit word reading (unlike with vocal responding), and 2) there was no strong need to employ many attentional resources to actively monitor the correct implementation of stimulus-response mappings (unlike with manual responding without feedback). In other words, because attention was relaxed, word processing proceeded relatively normally in Experiments 2A and 3A (in the no-load condition), resulting in a large contingency learning effect. Naturally then, obtaining a reduction in this effect by imposing a memory load was easier in this situation than in a situation like that in Experiment 1A in which attention was, to some extent, diverted from the word dimension to begin with.

Another interesting result emerging in the present research concerns the relation between contingency learning and WM capacity. Based on Schmidt et al.'s (2010) idea that the amount of limited-capacity resources available determines the magnitude of contingency learning effects, we derived the expectation that low WM-capacity individuals (i.e., individuals with fewer WM resources) would produce smaller contingency learning effects than high WM-capacity individuals (i.e., individuals with more WM resources), thus producing results that paralleled the results from the WM-load manipulations. A different expectation, however, would be derived based on a control account such as the DMC framework (Braver, 2012; Braver et al., 2007), an account that assumes that reactive control is more easily implemented when WM resources are scarce. The reason that such an account is relevant to the process of learning contingencies in a non-conflict situation is that the notion of reactive control could, in theory, encompass not only control over associations between words and control settings (i.e., adaptation to item-specific conflict frequency) but also control over associations between words and responses in general (Abrahamse et al., 2016; Egner, 2014; Hutchison, 2011). If the learning of contingencies is interpreted as a component of a reactive control strategy, one would then expect that low WM-capacity individuals' preference for reactive control would result not only in more robust item-specific proportion-congruent effects but also in stronger, rather than weaker, contingency learning effects. The reason is that low WM-capacity participants may find themselves processing words to a deeper level and may thus be advantaged in learning word-response contingencies (Hutchison, 2011). In sum, based on the contingency learning account, one would expect larger contingency learning effects in high WM-capacity individuals because those individuals would have more limited-capacity resources available to learn contingencies in the task. On the other hand, based on the DMC account, one would expect larger contingency learning effects in low WM-capacity individuals because those individuals would be more prone to use reactive control, the type of control that would underlie learning of contingencies.

Unfortunately, both the re-analysis of Hutchison's contingency learning data and the results of Experiment 3A, overall, failed to find evidence for either hypothesis, with extreme-groups analyses revealing equal-sized contingency learning effects for low and high WM-capacity individuals. The situation was somewhat different in our full-sample analysis, however, as this analysis revealed an important role of the order in which Experiment 3A (the non-conflict color identification task) was performed. For the participants in Experiment 3A who completed the noncolor identification task as the first task (the condition that should be considered the normal one), higher WM capacity led to *reduced* contingency learning effects, consistent with the DMC account. Nonetheless, the opposite pattern (larger contingency learning effects for higher WM-capacity individuals) was found in the group of participants who performed the nonconcolor identification task following the Stroop task, a pattern that is more easily reconciled with the contingency learning account.

Overall, what these results suggest is that a complete explanation for the relation between contingency learning and WM capacity is unlikely to be found either in the original contingency learning account (e.g., Schmidt et al., 2010) or in control accounts, such as the DMC account, extended to explain non-conflict associations (e.g., Abrahamse et al., 2016). In particular, it would seem that both types of account would require additional notions to explain the pattern of results that emerged from our full-sample analysis. These notions could include, for example, the idea that the process of learning contingencies might be more effective when individuals are discouraged from engaging in the concurrent process of making sure that stimulus-response mappings are being correctly implemented (Spinelli et al., under review). In turn, 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 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 after an entire block of practice they may relax the monitoring process and be better able to pick up on those contingencies.

In any case, these hypotheses are, of course, purely speculative at this point. In addition, the data suffer from the same power limitation described for the Stroop task above (and even more so in the presence of an order effect that essentially cuts the sample in half), casting some doubt on the reliability of the effects reported. Overall, further research is needed to clarify whether and how WM capacity influences color-word 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 a manifestation of a reactive control strategy of adaptation to item-specific conflict frequency rather the result of 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 strategy 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 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 $\text{GREEN}_{\text{yellow}}$ represents both a mostly-incongruent word and a mostly-incongruent color, thus providing a strong bias toward word inhibition. In contrast, word-specific and color-specific conflict frequency provide conflicting 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, however, in the crucial conditions in that experiment, they appeared in both blue and green, colors that were associated with infrequent conflict. 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 crucial for adjudicating between conflict adaptation and contingency learning accounts of the item-specific proportion-congruent effect.

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 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; by necessity, a high-contingency color existed for mostly-congruent words in both versions).

In both versions of the task, 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 used 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 compared to words that were mostly congruent 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 a 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 designs: In the four-item set design, conflict adaptation would be the dominant mechanism, as demonstrated by the fact that, 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 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 cause this inconsistency?

It is worth noting that, in Bugg and Hutchison's (2013) manipulation, there seems to be little necessity for individuals to transfer knowledge about item-specific conflict frequency acquired from the set of stimuli appearing in the first part of the experiment to the new set of stimuli appearing in the final block, even if the words appearing in the final block are the same as those used in the first part. To use an example from a real-life situation, the fact that a certain Joe is a trustworthy person does not mean that another Joe should also be considered trustworthy. Going back to Bugg and Hutchison's experiment, there was indeed little reason for participants to apply a conflict adaptation strategy in the final block because, in that block, words that used to be mostly congruent in the first part and words that used to be mostly incongruent in the first part appeared with congruent and incongruent colors equally often (i.e., the item-specific proportion-congruent manipulation was not maintained in that block). Of course, the fact that there was no necessity to transfer information about item-specific conflict frequency from the first part to the final block of the experiment does not mean that participants would not transfer that information nonetheless. However, this fact does imply that 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 a conflict adaptation strategy had not 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). Because the item-specific conflict information learned in the first part of the experiment does not clearly help 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. 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. 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, 2014b, 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 contributions of contingency learning and of 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. 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.
2. In addition to the regular analyses of 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 whether the WM-load effects of interest would emerge in terms of proportional changes from baseline. Again, the results were virtually identical in all cases.
3. 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, although not commonly used for Stroop experiments (although see Schmidt & Besner, 2008), offers a direct parallel to the analysis for the non-conflict color identification task because it allows an evaluation of the interaction between Contingency and WM Load in both types of tasks (although note that, in the Stroop task, the *F* value of that two-way interaction is equivalent to the *F* value of the three-way interaction between Congruency, Item Type, and WM Load in the analysis with Item Type as a factor). 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.

4. The error rates in the low-load group in Experiment 1B, for which no item-specific proportion-congruent effect was found, are an exception to this pattern. However, that group made very few errors and showed overall smaller congruency effects in the error rates than the other groups, suggesting that results from their accuracy data may reflect a floor effect and, hence, should be interpreted cautiously.
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 that load procedure in the following experiments. To foreshadow those results, the contingency learning effects in Experiments 2A and 3A were similar in size to those reported by Schmidt et al. (2010), suggesting that the reduced effect sizes in Experiment 1A were not due to the use of the two-alternative forced choice procedure.

6. The reported means are based on Hutchison's (2011) original data but were recalculated collapsing low-contingency and high-contingency incongruent stimuli in the mostly-incongruent item condition (for more details see Hutchison, 2011).
7. We would like to thank Keith Hutchison for sharing his data with us.
8. We would like to thank Ken Paap for sharing his questionnaire with us.
9. 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.
10. 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, $F(2, 265) = 14.81$, $MSE = 27950$, $p < .001$, $\eta_p^2 = .101$, and marginally so in the error rates, $F(2, 265) = 2.52$, $MSE = .002$, $p = .083$, $\eta_p^2 = .019$. 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, $F(2, 265) = .22$, $MSE = 1053$, $p = .80$, $\eta_p^2 = .002$ for the latencies, $F(2, 265) = .93$, $MSE = .002$, $p = .40$, $\eta_p^2 = .007$ for the error rates. In fact, the Bayes Factors for both the latencies, $BF_{01} = 20.11$, and the error rates, $BF_{01} = 10.89$, indicated "strong" evidence for the absence of the three-way interaction.

11. A related question concerns the implications that reduced contingency learning in vocal responding might have for interpreting the results of item-specific proportion-congruent manipulations in the Stroop task in which this response modality is used. Specifically, 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. On the other hand, the results of a recent item-specific proportion-congruent manipulation in our lab suggest a more cautious conclusion

(Spinelli & Lupker, in press). In that experiment, the design permitted us to dissociate the independent contributions of contingency learning and adaptation to item-specific conflict frequency in 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).

Chapter 5: Summary and Conclusions

Conflict adaptation is not an illusion: Support for proactive and reactive adaptation to conflict frequency in the Stroop task

The ability to resolve conflict from information that is irrelevant to one's current goal is a primary characteristic of an efficient control system. Another important ability that might contribute to an efficient control system is the ability to learn to regulate attention between task-relevant and task-irrelevant information based on the frequency with which task-irrelevant information conflicts with the current goal. Such a conflict adaptation function is indeed a core property of the conflict-monitoring model (Botvinick et al., 2001) and of other popular theories of cognitive control (e.g., Braver, 2012; Braver et al., 2007; Kane & Engle, 2003). In recent years, however, there has been increasing research interest in the idea that what have traditionally been considered markers of processes of adaptation to conflict frequency in tasks such as the Stroop (1935) task might reflect, in fact, more general learning processes that are not directly related to conflict (Schmidt, 2013b, 2019). In the present research, I aimed to provide an initial answer to the fundamental question that this idea entails, i.e., is conflict adaptation an illusion (Schmidt et al., 2015)? Tackling this issue using a range of approaches and methodologies, the present research has accumulated converging evidence showing that processes of adaptation to conflict frequency cannot be dismissed as easily as non-conflict learning accounts suggest.

First, it was demonstrated that humans can and do adapt to the overall frequency with which conflict occurs in a list of trials. This process, examined in the list-wide Proportion-Congruent (PC) paradigm in Stroop and Stroop-like tasks, implies that the latency difference between non-conflicting items (e.g., congruent) and conflicting items (e.g., incongruent) would be relatively small in a list in which conflicting items are frequent (e.g., a Mostly Incongruent [MI] list) because, to deal with the high frequency of conflict in this situation, attention to task-relevant information would be more focused. On the other hand, that latency difference would be relatively large in a list in which conflicting items are infrequent (e.g., a Mostly Congruent [MC] list) because, as conflict is infrequently experienced in this situation, attention to task-relevant information can be relaxed.

While this PC effect is indeed the pattern typically obtained, several other processes exist that could explain this effect without assuming a process of adaptation to list-wide conflict frequency. Specifically, the PC effect could be produced by adaptation to item-specific, as opposed to list-wide, conflict frequency (Blais et al., 2007), learning of word-response contingencies (Schmidt & Besner, 2008), adaptation to the informativeness of the stimuli (Schmidt, 2014b, 2019), and/or learning of temporal expectancies for the emission of a response (Schmidt, 2013c). Any of these processes can be engaged in traditional list-wide PC paradigms, making it difficult to determine what a list-wide PC effect would reflect in those circumstances. Although recent research has attempted to dissociate the process of adaptation to list-wide conflict frequency, the process that traditional explanations of the list-wide PC effect assume (e.g., Botvinick et al., 2001), from other, mainly conflict-unrelated processes, little of that research has clearly reported a list-wide PC effect uniquely attributable to list-wide conflict adaptation (Schmidt, 2013b, 2019; but see Cohen-Shikora et al., 2018). The present research began to fill this gap by employing two approaches.

The first approach, reported in Chapter 2, was to use a variant of the Stroop task, the picture-word interference task, to construct a list-wide PC manipulation in which no individual target or distractor stimulus was repeated. The advantage that this characteristic of the task affords is that it completely eliminates processes that are typically made possible in PC manipulations in which stimuli are repeated (i.e., adaptation to item-specific conflict frequency, contingency learning, and adaptation to stimulus informativeness). Even so, a list-wide PC effect emerged in both a task requiring a picture naming response and a task requiring a picture categorization response. Furthermore, in line with a recent assessment of the temporal learning account of the list-wide PC effect (Cohen-Shikora et al., 2018), both the analyses of the picture-word interference tasks and the results of an additional picture naming task failed to show evidence that a non-conflict temporal learning process contributed to the obtained list-wide PC effects.

The second approach, reported in Chapter 3, was to devise a manipulation of conflict frequency in the classic color-word Stroop task which, similar to what was achieved in the picture-word interference tasks in Chapter 2, would negate the possibility of participants applying processes related to item-specific conflict frequency, contingency learning, or stimulus informativeness, even though individual stimuli were, by necessity, repeated multiple times. This result was obtained by manipulating the frequency of neutral and incongruent items rather than congruent and incongruent items as in the standard list-wide manipulation. Similar to the standard manipulation, both a frequently conflicting list in which incongruent items were frequent and neutral items were infrequent (an MI list) and an infrequently conflicting list in which neutral items were frequent and incongruent items were infrequent (a Mostly Neutral [MN] list) were used. For some critical items in this manipulation, no processes other than adaptation to list-wide conflict frequency could have had a differential impact in the two lists, a situation that cannot be easily implemented in the standard list-wide PC paradigm. Nevertheless, a list-wide Proportion-Neutral effect, similar to the list-wide PC effect in the standard paradigm, emerged, as the contrast between incongruent and neutral items revealed a larger difference (i.e., more interference) in the MN list in which conflict was infrequent than in the MI list in which conflict was frequent. Furthermore, an analysis aimed to control for temporal learning revealed that no such process could explain the pattern of results obtained. Taken together, the findings from Chapters 2 and 3 provide converging evidence that in Stroop and Stroop-like tasks, the finding that congruency and/or interference effects are larger for infrequently conflicting vs. frequently conflicting lists likely reflects not only non-conflict learning processes and/or item-specific control processes, but also a list-wide control process whereby attention between task-relevant and task-irrelevant information is adjusted to the frequency of conflict in the list.

Another control process that this research has illuminated is a process whereby attention to task-relevant vs. task-irrelevant information is adjusted based on the conflict frequency associated with specific items in the list as opposed to the list as a whole. This item-specific conflict adaptation process was one of the explanations that Jacoby et al. (2003) proposed for their finding that frequently conflicting items (MI items) produced a smaller congruency effect than infrequently conflicting items (MC items) when intermixed in the same list – an item-specific PC effect. The other explanation that Jacoby et al. considered for this effect was a contingency learning explanation according to which the item-specific PC effect results from the process of learning contingencies between each word and its most likely response.

Subsequent research has tended to favor a contingency learning account of the item-specific PC effect, at least in Jacoby et al.’s (2003) paradigm in which a contingency learning process can be engaged for both MI and MC words (Bugg & Hutchison, 2013; Schmidt, 2013a). The research reported in Chapter 4, however, challenges this conclusion. Based on the assumption that limited-capacity resources are necessary for learning contingencies (Schmidt et al., 2010), we tested the contingency learning account of the item-specific PC effect by combining Stroop and non-conflict versions of a color identification task with a concurrent working memory load task. Consistent with Schmidt et al., some evidence emerged suggesting that increasing working memory load reduces people’s ability to learn contingencies in a non-conflict color identification task. In contrast, no impact of concurrent working memory load was found for the item-specific PC effect in the classic Stroop task. These results pose a challenge for a contingency learning account of the item-specific PC effect. If contingency learning were the only process driving the item-specific PC effect, as has been argued (Schmidt & Besner, 2008), then carrying a high working memory load should have reduced that effect in the Stroop task, paralleling the results obtained for the contingency learning effect in the non-conflict color identification task. The fact that, across three experiments, this pattern was not observed, suggests that contingency learning might not be the only process driving the item-specific PC effect. Instead, adaptation to item-specific conflict frequency might have an important role in producing this effect.

Overall, the results of the present research consistently refute the argument that conflict adaptation is an illusion (Schmidt et al., 2015). On the contrary, evidence in favor of this process emerged in two popular paradigms used to study adaptation to conflict frequency: the list-wide and the item-specific PC paradigms. Although these paradigms are similar in that they both involve manipulating the frequency of conflict in a certain context, they are believed to index distinct forms of control (e.g., Blais et al., 2007; Bugg et al., 2008; Gonthier et al., 2016). Specifically, according to the Dual-Mechanisms of Control account (Braver, 2012; Braver et al., 2007; De Pisapia & Braver, 2006; Gonthier et al., 2016), the list-wide PC effect would involve two forms of control. On the one hand, a form of proactive or preparatory control that would maintain attention focused to task-relevant information in a situation in which conflict is frequently experienced (i.e., in an MI list). On the other hand, a form of reactive control (i.e., control applied after stimulus onset) would be engaged when conflict is detected in a situation in which conflict is infrequently experienced (i.e., in an MC or an MN list) and attention, as a result, is more relaxed. This reactive form of control would cause the task goal to be reactivated in order to deal with that unexpected conflict. In contrast, the item-specific PC effect would mainly reflect the action of reactive control, although this control would be somewhat different from the reactive control applied in infrequently conflicting lists in the list-wide PC paradigm. Specifically, the reactive control engaged in the item-specific PC paradigm would retrieve the control settings that are appropriate to the item being presented, i.e., more focused attention to task-relevant information when the presented item is an MI item and relaxed attention when the presented item is an MC item.

This Dual Mechanisms of Control account would also explain why, in Chapter 4, the item-specific PC effect was not reduced by a concurrent working memory load. According to this account, reduced working memory resources would favor reliance on reactive control, the mode of control that the item-specific PC effect is a manifestation of (e.g., Burgess & Braver, 2010). As a result, a concurrent working memory load would, if anything, encourage participants to reactively adapt to item-specific conflict frequency, with the resulting item-specific PC effect being no smaller while carrying a high working memory load than while carrying low or no load.

This interpretation received partial support from the individual-differences analyses reported in Chapter 4. In an extreme-groups comparison, individuals with a low working memory capacity showed a larger item-specific PC effect than individuals with a high working memory capacity (albeit only in the error rates, similar to what found in previous reports: Hutchison, 2011; Kane & Engle, 2003). This result would be consistent with the idea that, because it reflects item-specific reactive control, adaptation to item-specific conflict frequency is preferentially engaged when working memory resources are reduced. On the other hand, another individual-differences analysis in which the whole range of working memory capacity was considered (unlike in the extreme-groups analysis) failed to replicate this result (working memory capacity did not modulate the item-specific PC effect in a monotonic fashion). In any case, what these results suggest is that, consistent with the Dual Mechanisms of Control account, reactive control is not impaired when working memory resources are reduced, either because an individual's working memory capacity is lower or because a load is maintained in working memory. In general, the present research is consistent with the view that humans can and do use processes of adaptation to conflict frequency, and that these processes occur at multiple levels of control (Bugg et al., 2008).

Limitations and future directions

The main objective of the present research was to determine whether PC effects in Stroop and Stroop-like tasks are completely explained by non-conflict learning confounds, such that, when those confounds are eliminated, so are the PC effects (Schmidt et al., 2015). In that sense, this research was successful in that it demonstrated that PC effects are still observable in situations in which non-conflict learning processes are impaired or made impossible altogether. However, a question that this research leaves unanswered is whether those situations would represent a special case compared to the standard PC paradigms. The reason this question is relevant is that Bugg and Hutchison (Bugg, 2014a; Bugg & Hutchison, 2013) proposed that in both list-wide and item-specific PC paradigms, adaptation to conflict frequency would be the main process driving PC effects only if the situation examined is one in which contingency learning is not a reliable process overall (e.g., in an MI list [in the list-wide paradigm] and in item-specific PC manipulations in which no contingencies can be learned for most of the items or in the list). Notably, the situations examined in the present research were all situations of this sort. In fact, contingency learning was not an option at all in the list-wide conflict frequency manipulations reported in Chapters 2 and 3. Consequently, from these results, it is not clear whether adaptation to list-wide conflict frequency would be possible in a situation in which contingency learning can be concurrently used, a type of situation that many list-wide PC paradigms create. Similarly, although contingency learning was possible in the item-specific PC paradigm reported in Chapter 4, the concurrent working memory load manipulation was used precisely to impair that process. Thus, although the fact that an item-specific PC effect was obtained with a concurrent working memory load does suggest that, when working memory resources are reduced, this effect likely reflects a process of adaptation to item-specific conflict frequency (because contingency learning was supposedly impaired), this effect may not reflect item-specific conflict adaptation in normal circumstances (when there is no concurrent working memory load and, therefore, contingency learning is not impaired).

Based on these considerations, an obvious goal for follow-up research is to determine whether, as proposed by Bugg and Hutchison (Bugg, 2014a; Bugg & Hutchison, 2013), adaptation to conflict frequency, albeit possible, would only be a “last resort”, i.e., a process that is only engaged when contingency learning cannot be used to minimize interference in the task. In fact, some data already exist that seem to run counter to this idea. For example, Hutchison (2011) obtained a list-wide PC effect even when the MI list (in addition to the MC list) was constructed in such a way that contingency learning was possible for most of the words in that list. In addition, Shedden et al. (2013) obtained evidence from Event-Related Potentials that in Jacoby et al.’s (2003) item-specific PC paradigm, MC items and MI items are distinguished early in processing, a result that is more consistent with the idea that two distinct processes (i.e., a process leading to the relaxation of attention for MC items vs. a focusing of attention for MI items), rather than one and the same process (i.e., contingency learning), are applied to those items. In sum, a closer examination of the idea that adaptation to conflict frequency may be applied even when other options (e.g., contingency learning) are available in the task appears necessary.

One way in which this idea might be explored is to modify PC paradigms so as to create a situation which allows researchers to separate conflict adaptation processes from non-conflict learning processes even though the latter processes can be engaged in the task. For example, Schmidt (2013a) constructed an item-specific PC manipulation in which item-specific conflict adaptation and contingency learning could be evaluated independently. Although Schmidt found no effect of item-specific conflict adaptation, as discussed in Chapter 4, his manipulation was problematic because for most of the critical items used to measure conflict adaptation, colors and words provided inconsistent information about the conflict frequency of the item. When we fixed this problem in an improved version of his design, an item-specific conflict adaptation effect emerged (Spinelli & Lupker, in press). Thus, although the evidence in support of a role of adaptation to conflict frequency even when word-response contingencies can be concurrently learned is still scarce, this evidence holds considerable promise for the idea that processes of adaptation to conflict frequency may not be merely a “last resort”.

Last word

The present research, overall, provides a significant contribution to our understanding of cognitive control engagement in response to situations varying in conflict frequency, a contribution that could be of benefit for both the theory and the practical (e.g., clinical) applications of the paradigms examined. Far from being a mere illusion, adaptation to conflict frequency might be an important resource in coping with tasks that frequently vs. infrequently require people to deal with conflict. Although learning about what to respond (contingency learning and stimulus informativeness) and when to do it (temporal learning) might be important aspects of successful goal-oriented behavior, learning how to respond (i.e., learning the appropriate attentional strategy to achieve the goal) is another human ability that needs to be acknowledged.

References

Aben, B., Verguts, T., & Van den Bussche, E. (2017). Beyond trial-by-trial adaptation: A quantification of the time scale of cognitive control. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 509-517. http://dx.doi.org/10.1037/xhp0000324

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*, 693-728. https://doi.org/10.1037/bul0000047

Abrahamse, E., Ruitenberg, M., Duthoo, W., Sabbe, B., Morrens, M., & van Dijck, J. (2016). Conflict adaptation in schizophrenia: reviewing past and previewing future efforts. *Cognitive Neuropsychiatry*, *21*, 197-212. https://doi.org/10.1080/13546805.2016.1167679

Ahmed, L., & De Fockert, J. W. (2012). Focusing on attention: the effects of working memory capacity and load on selective attention. *PloS One*, *7*, e43101. https://doi.org/10.1371/journal.pone.0043101

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. https://doi.org/10.1037/a0028458

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. https://doi.org/10.3389/fpsyg.2014.01410

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

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. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

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. https://doi.org/10.1037/a0032186

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48. https://doi.org/10.18637/jss.v067.i01

Beckers, T., De Houwer, J., & Matute, H. (2007). Human contingency learning. *Quarterly Journal of Experimental Psychology*, *60*, 289-290. https://doi.org/10.1080/17470210601000532

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*, 2758-2767. https://doi.org/10.1162/jocn.2009.21365

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. https://doi.org/10.1037/0033-295x.114.4.1076

Bolker, B. (2018). *GLMM FAQ*. Retrieved from https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106-113. https://doi.org/10.1016/j.tics.2011.12.010

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. https://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Bonnin, C. A., Houeto, J. L., Gil, R., & Bouquet, C. A. (2010). Adjustments of conflict monitoring in Parkinson's disease. *Neuropsychology*, *24*, 542-546. https://doi.org/10.1037/a0018384

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. https://doi.org/10.1037/0033-295X.108.3.624

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., Egner, T. (in press). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*. Advance online publication.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106-113. https://doi.org/10.1016/j.tics.2011.12.010

Bugg, J. (2014a). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*, 567-587. https://doi.org/10.1037/a0035032

Bugg, J. M. (2014b). Evidence for the sparing of reactive cognitive control with age. *Psychology and Aging*, *29*, 115-127. https://doi.org/10.1037/a0035270

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. https://doi.org/10.3758/s13414-014-0752-x

Bugg, J. M., & Chanani, S. (2011). List-wide control is not entirely elusive: Evidence from picture–word Stroop. *Psychonomic Bulletin & Review, 18*, 930-936. https://doi.org/10.3758/s13423-011-0112-y

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. https://doi.org/10.3389/fpsyg.2012.00367

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. https://doi.org/10.1037/a0029145

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. https://doi.org/10.1037/a0019957

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*, 1484-1494. https://doi.org/10.3758/MC.36.8.1484

Bugg, J. M., McDaniel, M. A., Scullin, M. K., & Braver, T. S. (2011). Revealing list-level control in the Stroop task by uncovering its benefits and a cost. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1595-1606. http://dx.doi.org/10.1037/a0024670

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. https://doi.org/10.1371/journal.pone.0012861

Chateau, D., & Lupker, S. J. (2003). Strategic effects in word naming: Examining the route-emphasis versus time-criterion accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 139-151. https://doi.org/10.1037/0096-1523.29.1.139

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (in press). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000670

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. https://doi.org/10.3758/bf03196772

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. https://doi.org/10.3758/BF03193850

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception and Psychophysics*, *11*, 274-276. https://doi.org/10.3758/BF03210377

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722. https://doi.org/10.3758/BF03196827

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*, 1322-1326. https://doi.org/10.1016/j.neucom.2005.12.100

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition*, *28*, 1437-1449. https://doi.org/10.3758/BF03211844

Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247. https://doi.org/10.3389/fpsyg.2014.01247

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. https://doi.org/10.1037//0096-3445.128.3.309

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

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. http://dx.doi.org/10.1037/0033-2909.125.6.777

Forrin, N. D., & MacLeod, C. M. (2017). Relative speed of processing determines color–word contingency learning. *Memory & Cognition*, *45*, 1206-1222. https://doi.org/10.3758/s13421-017-0721-4

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124. https://doi.org/10.3758/BF03195503

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. https://doi.org/10.3758/s13421-014-0461-7

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*, 778-788. https://doi.org/10.3758/s13421-016-0591-1

Grandjean, J., d'Ostilio, K., Fias, W., Phillips, C., Balteau, E., Degueldre, C., ... & Collette, F. (2013). Exploration of the mechanisms underlying the ISPC effect: evidence from behavioral and neuroimaging data. *Neuropsychologia*, *51*, 1040-1049. https://doi.org/10.1016/j.neuropsychologia.2013.02.015

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468-484. https://doi.org/10.1016/0013-4694(83)90135-9

Hazeltine, E., & Mordkoff, J. T. (2014). Resolved but not forgotten: Stroop conflict dredges up the past. *Frontiers in Psychology*, *5*, 1327. https://doi.org/10.3389/fpsyg.2014.01327

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 1-17. https://doi.org/10.1007/s00426-003-0132-y

Huber-Huber, C., & Ansorge, U. (2017). The role of RT carry-over for congruence sequence effects in masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 757-780. https://doi.org/10.1037/xlm0000331

Huber-Huber, C., & Ansorge, U. (2018). Unconscious conflict adaptation without feature-repetitions and response time carry-over. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 169-175. https://doi.org/10.1037/xhp0000450

Hutcheon, T. G., & Spieler, D. H. (2014). Contextual influences on the sequential congruency effect. *Psychonomic Bulletin & Review*, *21*, 155-162. https://doi.org/10.3758/s13423-013-0473-5

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. https://doi.org/10.1037/a0023437

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. https://doi.org/10.3758/s13414-016-1066-y

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*, 638-644. https://doi.org/10.3758/BF03196526

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience & Biobehavioral Reviews*, *46*, 30-43. https://doi.org/10.1016/j.neubiorev.2014.06.001

Jiménez, L., & Méndez, A. (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 271-284. http://dx.doi.org/10.1037/a0027734

Jiménez, L., & Méndez, A. (2014). Even with time, conflict adaptation is not made of expectancies. *Frontiers in Psychology*, *5*, 1042. https://doi.org/10.3389/fpsyg.2014.01042

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. https://doi.org/10.3758/s13423-014-0735-x

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. https://doi.org/10.1037/0096-3445.132.1.47

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. https://doi.org/10.1037/0096-3445.133.2.189

King, J. A., Korb, F. M., & Egner, T. (2012). Priming of control: Implicit contextual cuing of top-down attentional set. *Journal of Neuroscience*, *32*, 8192-8200. https://doi.org/10.1523/JNEUROSCI.0934-12.2012

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44,* 1730-1742. https://doi.org/10.1037/xlm0000552

Kinoshita, S., & Mozer, M. C. (2006). How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency-blocking effects. *Memory & Cognition*, *34*, 726-742. https://doi.org/10.3758/BF03193591

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General*, *140*, 622-636. https://doi.org/10.1037/a0024230

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, *77*, 576-588. http://dx.doi.org/10.2307/1420768

La Heij, W. (1988). Components of stroop-like interference in picture naming. *Memory & Cognition, 16*, 400-410. https://doi.org/10.3758/BF03214220

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. https://doi.org/10.1016/j.actpsy.2015.12.009

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. https://doi.org/10.1037/xlm0000436

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. https://doi.org/10.3389/fpsyg.2015.01171

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. https://doi.org/10.3758/BF03197535

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*, 485-495. https://doi.org/10.3758/BF03198265

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 570-590. https://doi.org/10.1037/0278-7393.23.3.570

Lupker, S. J., & Katz, A. N. (1981). Input, decision, and response factors in picture–word interference. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 269-282. https://doi.org/10.1037/0278-7393.7.4.269

Lupker, S. J., & Katz, A. N. (1982). Can automatic picture processing influence word judgments?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 418-434. https://doi.org/10.1037/0278-7393.8.5.418

Lupker, S. J., Kinoshita, S., Coltheart, M., & Taylor, T. E. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language*, *49*, 556-575. https://doi.org/10.1016/S0749-596X(03)00094-9

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*, 163-203. https://doi.org/10.1037/0033-2909.109.2.163

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. http://dx.doi.org/10.1037/a0029200

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. http://dx.doi.org/10.1037/xlm0000147

Melara, R. D., & Algom, D. (2003). Driven by information: a tectonic theory of Stroop effects. *Psychological Review*, *110*, 422-471. http://dx.doi.org/10.1037/0033-295X.110.3.422

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. https://doi.org/10.3758/BF03197195

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. http://dx.doi.org/10.1037/0278-7393.19.4.789

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.

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. https://doi.org/10.1523/jneurosci.2966-15.2016

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

Melara, R. D., & Algom, D. (2003). Driven by information: a tectonic theory of Stroop effects. *Psychological review*, *110*, 422-471. https://doi.org/10.1037/0033-295X.110.3.422

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. https://doi.org/10.1523/JNEUROSCI.2966-15.2016

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. https://doi.org/10.1037/0894-4105.7.2.119

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of the item frequency of the previous trial. *Quarterly Journal of Experimental Psychology*, *56*, 385-402. https://doi.org/10.1080/02724980244000387

Proctor, R. W. (1978). Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, *23*, 413-419. https://doi.org/10.3758/BF03204145

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. https://doi.org/10.3758/BF03192979

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

Rastle, K., Kinoshita, S., Lupker, S. J., & Coltheart, M. (2003). Cross-task strategic effects. *Memory & Cognition*, *31*, 867-876. https://doi.org/10.3758/BF03196441

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. https://doi.org/10.1027/1015-5759/a000123

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231. https://doi.org/10.1037//0096-3445.124.2.207

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 515-537. http://dx.doi.org/10.1037/0096-1523.27.3.515

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*, 103-113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

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. https://doi.org/10.1016/j.actpsy.2012.11.004

Schmidt, J. R. (2013b). Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20*, 615-630. https://doi.org/10.3758/s13423-012-0373-0

Schmidt, J. R. (2013c). Temporal learning and list-level proportion congruency: conflict adaptation or learning when to respond?. *PLoS One*, *8*, e82320. https://doi.org/10.1371/journal.pone.0082320

Schmidt, J. R. (2014a). List-level transfer effects in temporal learning: Further complications for the list-level proportion congruent effect. *Journal of Cognitive Psychology*, *26*, 373-385. https://doi.org/10.1080/20445911.2014.896367

Schmidt, J. R. (2014b). Contingencies and attentional capture: the importance of matching stimulus informativeness in the item-specific proportion congruent task. *Frontiers in Psychology,* *5*, 540. https://doi.org/10.3389/fpsyg.2014.00540

Schmidt, J. R. (2016). Temporal learning and rhythmic responding: no reduction in the proportion easy effect with variable response-stimulus intervals. *Frontiers in Psychology, 7*, 634. https://doi.org/10.3389/fpsyg.2016.00634

Schmidt, J. R. (2017). 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. https://doi.org/10.1037/cep0000106

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. https://doi.org/10.3758/s13421-017-0755-7

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

Schmidt, J. R., Augustinova, M., & De Houwer, J. (2018). Category learning in the color-word contingency learning paradigm. *Psychonomic Bulletin & Review, 25*, 658-666. https://doi.org/10.3758/s13423-018-1430-0

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. https://doi.org/10.1037/0278-7393.34.3.514

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. https://doi.org/10.1016/j.concog.2006.06.010

Schmidt, J. R., & De Houwer, J. (2012). Learning, awareness, and instruction: Subjective contingency awareness does matter in the colour-word contingency learning paradigm. *Consciousness and Cognition*, *21*, 1754-1768. https://doi.org/10.1016/j.concog.2012.10.006

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. https://doi.org/10.1016/j.lmot.2016.09.002

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. https://doi.org/10.1016/j.concog.2009.12.016

Schmidt, J. R., Notebaert, W., & van den Bussche, E. (2015). Is conflict adaptation an illusion?. *Frontiers in Psychology*, *6*, 172. https://doi.org/10.3389/fpsyg.2015.00172

Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PLoS One*, *9*, e102337. https://doi.org/10.1371/journal.pone.0102337

Schmidt, J. R., & Weissman, D. H. (2016). Congruency sequence effects and previous response times: conflict adaptation or temporal learning?. *Psychological Research*, *80*, 590-607. https://doi.org/10.1007/s00426-015-0681-x

Schneider, D. W. (2015). Attentional control of response selection in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1315-1324. https://doi.org/10.1037/xhp0000091

Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a postperceptual effect. *Psychonomic Bulletin & Review*, *19*, 301-308. https://doi.org/10.3758/s13423-011-0190-x

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. https://doi.org/10.1016/j.concog.2013.10.002

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., ... & Eskandar, E. N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature*, *488*, 218-221. https://doi.org/10.1038/nature11239

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. https://doi.org/10.2139/ssrn.2160588

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.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture–word processing. *Journal of Experimental Psychology: General*, *109*, 373-392. https://doi.org/10.1037/0096-3445.109.4.373

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. https://doi.org/10.3758/cabn.3.3.155

Spinelli, G., & Lupker, S. L. (in press). 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*.

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. J., (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. http://dx.doi.org/10.1037/xhp0000656

Starreveld, P. A., & La Heij, W. (2017). Picture-word interference is a Stroop effect: A theoretical analysis and new empirical findings. *Psychonomic Bulletin & Review*, *24*, 721-733. https://doi.org/10.3758/s13423-016-1167-6

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662. https://doi.org/10.1037/h0054651

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., . . . Bates, E. (2005). Timed action and object naming. *Cortex*, *41*, 7-25. https://doi.org/10.1016/S0010-9452(08)70174-6

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 117-138. https://doi.org/10.1037/0278-7393.27.1.117

Taylor, T. E., & Lupker, S. J. (2006). Time perception and word recognition: An elaboration of the time-criterion account. *Perception & Psychophysics*, *68*, 933-945. https://doi.org/10.3758/bf03193356

Taylor, T. E., & Lupker, S. J. (2007). Sequential effects in time perception. *Psychonomic Bulletin & Review*, *14*, 70-74. https://doi.org/10.3758/bf03194030

Torres-Quesada, M., Funes, M. J., & Lupiáñez, J. (2013). Dissociating proportion congruent and conflict adaptation effects in a Simon–Stroop procedure. *Acta Psychologica*, *142*, 203-210. https://doi.org/10.1016/j.actpsy.2012.11.015

Torres-Quesada, M., Lupiáñez, J., Milliken, B., & Funes, M. J. (2014). Gradual proportion congruent effects in the absence of sequential congruent effects. *Acta Psychologica*, *149*, 78-86. https://doi.org/10.1016/j.actpsy.2014.03.006

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*, 127-154. https://doi.org/10.1016/0749-596x(89)90040-5

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*, 727-735. https://doi.org/10.3758/bf03202722

Underwood, G. (1976). Semantic interference from unattended printed words. *British Journal of Psychology, 67*, 327-338. https://doi.org/10.1111/j.2044-8295.1976.tb01519.x

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. https://doi.org/10.3758/bf03192720

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture—word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, *16*, 987-999. https://doi.org/10.3758/PBR.16.6.987

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological Review*, *115*, 518-525. http://dx.doi.org/10.1037/0033-295X.115.2.518

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, *13*, 304-319. https://doi.org/10.3758/bf03202499

Weissman, D. H., Jiang, J., & Egner, T. (2014). Determinants of congruency sequence effects without learning and memory confounds. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 2022-2037. https://doi.org/10.1037/a0037454

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*, 179-189. https://doi.org/10.1111/1469-8986.3720179

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. New York: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wilk, H. A., Ezekiel, F., & Morton, J. B. (2012). Brain regions associated with moment-to-moment adjustments in control and stable task-set maintenance. *Neuroimage*, *59*, 1960-1967. https://doi.org/10.1016/j.neuroimage.2011.09.011

Yang, H., Chen, J., Spinelli, G., & Lupker, S. J. (2019). The impact of text orientation on form priming effects in four-character Chinese words. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 45*, 1511-1526. https://doi.org/10.1037/xlm0000655

Ethics Approval

Western
Research

Research Ethics

Western University Non-Medical Research Ethics Board
NMREB Delegated Initial Approval Notice

Principal Investigator: Prof. Stephen Lupker
Department & Institution: Social Science\Psychology,Western University

NMREB File Number: 108956
Study Title: Choosing words for speaking

NMREB Initial Approval Date: February 28, 2017
NMREB Expiry Date: February 28, 2018

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Letter of Information & Consent	Letter of Information and consent - 30 minutes slot	2017/01/20
Letter of Information & Consent	Letter of Information and Consent - 60 minutes slot	2017/01/20
Western University Protocol	Received January 30, 2017.	
Recruitment Items	SONA description - 30 minutes slot	2017/01/20
Recruitment Items	SONA description - 60 minutes slot	2017/01/20
Other	Debriefing Form	2017/02/09

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941. [redacted]

Western University, Research, Support Services Bldg., Rm. 5150
London, ON, Canada N6G 1G9 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/ethics

Western Research

Date: 15 February 2018

To: Prof. Stephen Lupker

Project ID: 108956

Study Title: Choosing words for speaking

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Date Approval Issued: 15/Feb/2018

REB Approval Expiry Date: 28/Feb/2019

Dear Prof. Stephen Lupker,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted]

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Western Research

Date: 13 February 2019

To: Prof. Stephen Lupker

Project ID: 108956

Study Title: Choosing words for speaking

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Meeting Date: 01/Mar/2019

Date Approval Issued: 13/Feb/2019

REB Approval Expiry Date: 28/Feb/2020

Dear Prof. Stephen Lupker,

The Western University Non-Medical Research Ethics Board has reviewed this application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted]

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Curriculum Vitae

EDUCATION

2015-now	Ph.D. Psychology
	University of Western Ontario
	Expected completion: Fall 2019
2012-2014	M.A. Linguistics
	Sapienza University of Rome
	Grade Average: 30/30
	Final Grade: 110 cum laude/110
2009-2012	B.A. Humanities
	University of Florence
	Grade Average: 29.92/30
	Final Grade*:* 110 cum laude/110
2004-2009	High School Diploma
	Liceo Scientifico “Amedeo di Savoia Duca d’Aosta” of Pistoia
	Final Grade: 100 cum laude/100

WORKSHOPS

2016	Brain and Mind Institute Summer Workshop in EEG
	University of Western Ontario
2015	Teaching Assistant Training Program
	University of Western Ontario
2014	1st School of Statistics for Linguists
	Sapienza University of Rome
2013	3rd International School of Textual Data Analysis and Text Mining
	Sapienza University of Rome

WORK EXPERIENCE

Sept. 2019-now	Data analyst
	Ivey Business School
2018-now	Research Assistant
	University of Western Ontario

TEACHING EXPERIENCE

2018-2019	Research Assistant
	University of Western Ontario
2016-2019	Honors Student Supervisor and Co-supervisor
	University of Western Ontario
Fall 2016	Teaching Assistant
	Course: Special Topics in Psychology: Autobiographical Memory
	University of Western Ontario
Fall 2015	Teaching Assistant
	Course: Introduction to Psychology
	University of Western Ontario

AD HOC REVIEWER CONTRIBUTIONS

Journal of Experimental Psychology: Learning, Memory, and Cognition

Journal of Research in Reading

Quarterly Journal of Experimental Psychology

International Journal of Bilingualism

HONORS AND AWARDS

2018	Psychonomic Society Accommodation Award
	Psychonomic Society's 3rd International Meeting
2015	Ontario Trillium Scholarship
	University of Western Ontario
2013	Paid internship
	Linguistics Library, Sapienza University of Rome
2013	"Wanted the Best" Fellowship
	Sapienza University of Rome
2011-2012	Best Graduate in Faculty of Arts and Philosophy
	University of Florence
2009	Excellent High School Students Grant
	Ministry of Education, Universities and Research
2008-2009	Best High School Students in Tuscany
	Cavalieri del Lavoro, Tuscan Division

PUBLICATIONS

Spinelli, G., Forti, L., & Jared, D. (under review). Learning to assign stress in a second language: The role of second-language vocabulary size and transfer from the native language in second-language readers of Italian.

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

Spinelli, G., & Lupker, S. J., (under review). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and stimulus-informativeness confounds.

Spinelli, G., Krishna, K., Perry, J. R., & Lupker, S. J., (under review). Working memory load dissociates contingency learning and item-specific proportion-congruent effects.

Lupker, S. J., Spinelli, & Davis, C. J. (in press). Masked form priming as a function of letter position: An evaluation of current orthographic coding models. *Journal of Experimental Psychology. Learning, Memory, and Cognition.*

Colombo, L., Spinelli, G., & Lupker, S. J. (in press). The impact of consonant-vowel transpositions on masked priming effects in Italian and English. *Quarterly Journal of Experimental Psychology.*

Spinelli, G., & Lupker, S. J., (in press). 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*.

Spinelli, G., Perry, J. R., & Lupker, S. J., (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*. Advance online publication.

Yang, H., Chen, J., Spinelli, G., & Lupker, S. J. (2018). The impact of text orientation on form priming effects in four-character Chinese words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. Advance online publication.

Sulpizio, S., Spinelli, G., & Burani, C. (2017). STRESYL: An Italian stress-in-syllables database for reading research. *Written Language & Literacy*, *20*, 80-103.

Spinelli, G., Sulpizio, S., & Burani, C. (2017). Q2Stress: A database for multiple cues to stress assignment in Italian. *Behavior Research Methods*, *49*, 2113-2126.

Sulpizio, S., Spinelli, G., & Burani, C. (2016). Stress still affects articulatory planning in reading aloud: A reply to White and Besner (2016). *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 2129-2131.

Spinelli, G., Sulpizio, S., Primativo, S., & Burani, C. (2016). Stress in context: Morpho-syntactic properties affect lexical stress assignment in reading aloud. *Frontiers in Psychology*, *7*, 942.

Sulpizio, S., Spinelli, G., & Burani, C. (2015). Stress affects articulatory planning in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 453-461.

TALKS

Spinelli, G., & Lupker, S. J. Proportion congruent effects do have something to do with congruency: Adaptation to item-specific conflict frequency in the Stroop task. *CSBBCS's 29th Annual Meeting*, Waterloo, Canada, June 7-9, 2019.

Spinelli, G., Forti, L., & Jared, D. Learning to assign stress in a second language: The role of L2 vocabulary size and transfer from L1 in second-language readers of Italian. *CSBBCS's 29th Annual Meeting*, Waterloo, Canada, June 7-9, 2019.

Spinelli, G., Perry, J. R., & Lupker, S. J. Not so fast: Delaying fast responses reduces the congruence effect on immediate-response trials. *Cognitive Science Arena 2018*, Bressanone-Brixen, Italy, February 23-24, 2018.Spinelli, G., Burani, C., Forti, L., Micheli, M. S., Sulpizio, S., Talamo, L., & Wilson, M. A. Towards a morphological database of Italian. *Giornata di Linguistica Sperimentale e Computazionale*, Padova, Italy, February 23, 2018.

Spinelli, G., Sulpizio, S., Primativo, S., & Burani, C. Il ruolo del contesto sintattico nell'assegnazione dell'accento di parola in italiano [The role of the syntactic context in lexical stress assignment in Italian]. *XX Congresso Nazionale AIP (Associazione Italiana di Psicologia), sezione Psicologia Sperimentale*, Pavia, Italy, September 15-17, 2014.

POSTERS

Spinelli, G., & Lupker, S. J. Contingency learning, yet adaptation to conflict frequency: Evidence from the item-specific and list-wide proportion-congruent paradigms. *Psychonomic Society's 60th Annual Meeting*, Montreal, Canada, November 14-17, 2019.

Colombo, L., Spinelli, G., & Lupker, S. J. Consonant-vowel structure and transposition effects in masked priming. *Psychonomic Society's 60th Annual Meeting*, Montreal, Canada, November 14-17, 2019.

Spinelli, G., & Lupker, S. J. Can you adapt to unconscious conflict? Evidence from the congruency-sequence and proportion-congruent paradigms. *ASSC's 23rd Annual Meeting*. London, Canada, June 25-28, 2019.

Spinelli, G., Lupker, S. J., & Katz, A. N. No better memory for unexpected conflict: Expectancy-based modulations of interference effects in the picture-word Stroop task do not affect later memory performance. *Psychonomic Society's 59th Annual Meeting*, New Orleans, LA, USA, November 15-18, 2018.

Spinelli, G., Krishna, K., Perry, J. R., & Lupker, S. J. WM load and WM span dissociate contingency learning and conflict adaptation. *Ninth European Working Memory Symposium (EWOMS),* Pavia, Italy, August 29-31, 2018.

Spinelli, G., Krishna, K., Perry, J. R., & Lupker, S. J. Conflict adaptation under load: Proportion congruent effects persist when contingency learning fails. *Psychonomic Society's 3rd International Meeting*, Amsterdam, The Netherlands, May 10-12, 2018.

Spinelli, G., Perry, J. R., & Lupker, S. J. Conflict adaptation without contingency and temporal learning: Evidence from the picture-word interference task. *Psychonomic Society's 58th Annual Meeting*, Vancouver, BC, Canada, November 9-12, 2017.

McPhedran, M., Taikh, A., Spinelli, G., & Lupker, S. J. The impact of visible intervenors on form and identity priming. *Psychonomic Society's 57th Annual Meeting*, Boston, MA, USA, November 17-20, 2016.

Spinelli, G., Sulpizio, S., & Burani, C. Lexical stress assignment in Italian: Resources and findings. *The 10th International Conference on the Mental Lexicon*, Ottawa, ON, Canada, October 19-21, 2016.

LANGUAGES

Italian: Mother tongue

Spanish: Intermediate

English standardized tests (2014):	TOEFL IBT	108
	GRE	Verbal Reasoning 164
		Quantitative Reasoning 162
		Analytical Writing 4.5

COMPUTER SKILLS

Programming: Python, R, MATLAB, Javascript

Operating Systems: Mac OS X, Windows 7/10, Linux (Ubuntu)

Spreadsheet: Microsoft Excel

Statistical software: R, SPSS

Experimental software: E-Prime, DMDX, Checkvocal, jspsych

NLP tools: TreeTagger, AntConc, TalTac 2, HFST