Published in final edited form as:

J Exp Psychol Hum Percept Perform. 2020 September ; 46(9): 1029–1050. doi:10.1037/xhp0000801.

What can be Learned in a Context-Specific Proportion Congruence Paradigm? Implications for Reproducibility

Julie M. Bugg, **Jihyun Suh**, **Jackson S. Colvett**, **Spencer G. Lehmann**
Washington University in St. Louis

Abstract

Crump and Milliken (2009) reported a context-specific proportion congruence (CSPC) effect for inducer and diagnostic sets, the strongest evidence to date of context-specific control. Attempts to replicate/reproduce this evidence have failed including Experiment 1. Using a picture-word Stroop task, we tackled the question of how to interpret such failures by testing the consistency hypothesis (Hutcheon & Spieler, 2017) and two novel hypotheses inspired by our theorizing about learning opportunities in the CSPC paradigm. Experiment 2 found a CSPC effect when there was no diagnostic set, supporting the consistency hypothesis. Experiment 3 produced novel evidence for item-PC learning in a CSPC paradigm. In contrast, Experiment 4 did not produce strong evidence for location-item conjunctive learning. Our findings suggest failures to replicate/reproduce the CSPC effect do not necessarily indicate a Type 1 error or instability but instead may indicate episodic representations were organized based on item and not location. This item-PC learning hypothesis uniquely predicted Experiment 3 findings and accommodates findings of all but one prior attempt to replicate/reproduce the CSPC effect for inducer and diagnostic sets, including Experiment 1. Predicting whether future attempts are successful will require deeper understanding of the factors that promote learning of item-PC versus location-PC associations.

Keywords

Cognitive Control; Context-Specific Proportion Congruence; Learning; Reproducibility

Cognitive control refers to a collection of processes that bias attention in favor of goal-relevant information while minimizing the influence of goal-irrelevant information. Cognitive control is often investigated using conflict tasks that pit a task-relevant response against a relatively automatic but inappropriate response, such as naming an ink color while ignoring a color word in the Stroop task (Stroop, 1935). There is ample evidence that cognitive control adapts to features of the environment and one prominent feature is the likelihood of encountering conflict. This feature is manipulated via proportion congruence manipulations, which vary the relative frequency of congruent (e.g., RED in red ink) and incongruent trials (e.g., BLUE in red ink). Proportion congruence has been manipulated across lists (i.e., blocks of trials), contexts (e.g., locations on screen), and items (e.g.,

Correspondence concerning this article should be sent to Julie M. Bugg, Department of Psychological and Brain Sciences, Campus Box 1125, Washington University in St. Louis, St. Louis, MO 63130, jbugg@wustl.edu.
Julie M. Bugg, Jihyun Suh, Jackson S. Colvett, & Spencer Lehmann, Department of Psychological and Brain Sciences, Washington University in St. Louis.

different stimuli). In each case the manipulation leads to a proportion congruence effect--less susceptibility to conflict (as indicated by smaller congruency effects) in the mostly incongruent (MI) condition compared to the mostly congruent (MC) condition (for a user's guide to the proportion congruence manipulation, see Bugg, 2017; for reviews, see Bugg, 2012; Bugg & Crump, 2012). The present study investigates the context-specific proportion congruence (CSPC) effect (Crump & Milliken, 2009). Of theoretical importance (Braver, Gray, & Burgess, 2007; Bugg, 2012), this effect represents a key piece of evidence demonstrating that control operates locally (reactively) post-stimulus onset to bias attention to goal relevant information and not just globally (proactively; see Bugg, 2014; Bugg & Chanani, 2011; Gonthier et al., 2016; Hutchison, 2011). At the same time, however, the stability of the CSPC effect has been called into question based on failures to replicate or reproduce the effect (Hutcheon & Spieler, 2017; Crump, Brosowsky, & Milliken, 2017).[1] The present study was motivated by the following question: should we attribute such failures to a Type 1 error or instability, or are there theoretical explanations that could explain the absence of CSPC effects? We tackle this question from the angle of asking about and seeking evidence for three distinct associations that can be learned in CSPC paradigms, two of which have yet to be tested empirically to our knowledge. Before describing these associations, we first provide background on the CSPC effect and extant discussions regarding the reproducibility of the effect.

A Brief History of CSPC Effects

Although a CSPC effect can be produced by manipulating the proportion congruence of a variety of contextual features (for a review, see Bugg & Crump, 2012), the most dominant feature to date has been the (screen) location in which stimuli appear. In one of the first demonstrations of a CSPC effect, a word stimulus (e.g., RED) was presented centrally in white followed by a to-be-named color-patch probe that was congruent (e.g., a red square) or incongruent (e.g., a blue square) with the word. The probe appeared randomly in one of two locations on screen (Crump, Gong, & Milliken, 2006; see also Corballis & Gratton, 2003). In one location (e.g., upper) the words and probe colors were MC, and in the second location (e.g., lower) they were MI. The congruency effect (i.e., RT slowing on incongruent trials compared to congruent trials) was significantly smaller in the MI location as compared to the MC location. This CSPC effect was taken as an indicator of the flexible and context-specific modulation of cognitive control on a trial-by-trial basis.

Direct evidence in support of this control-based account was later provided (Crump & Milliken, 2009). The key feature of the design that enabled the researchers to draw conclusions about control independent of lower-level mechanisms such as contingency-learning (Logan, 1988; Schmidt, 2017) was the inclusion of both an inducer set of stimuli that carried the CSPC manipulation, and a diagnostic set of stimuli that assessed the effects of the induction (see Braem et al., 2019). Just as in the original study, one location was MC (e.g., upper) and one was MI (e.g., lower). However, only the inducer set varied in

[1]Following Crump et al. (2017), we use the term replicate when referring to designs that attempted to directly replicate the CSPC effects of Crump and Milliken (2009) and reproduce when referring to design variants that attempted to measure the same general phenomenon of CSPC effects.

proportion congruence (e.g., 92% congruent in the upper location; 8% congruent in the lower location). In contrast, the diagnostic set was 50% congruent and frequency-matched in both locations. Additionally, the inducer set (e.g., words RED and GREEN in red or green ink) did not share either the relevant or irrelevant dimension with the diagnostic set (e.g., words BLUE and YELLOW in blue or yellow ink). Consequently, when Crump and Milliken (2009) observed a CSPC effect for the diagnostic set (i.e., smaller congruency effect in lower, MI location compared to upper, MC location), they could unambiguously attribute it to context-specific control. As such, this evidence has represented the strongest to date in support of a local, reactive control mechanism that operates on a trial-by-trial fashion based on the probability of conflict in a context (location).[2]

Theoretically, the episodic retrieval account represents the currently dominant account of context-specific control (Crump & Milliken, 2009; see also the instance-based memory account of context-specific control, Crump, 2016). According to this account, not only are features such as the stimulus, response, congruency (i.e., conflict status), and context (e.g., stimulus location) stored in episodes (i.e., event-files, instances; Hommel, 1998; Logan, 1988) reflecting experiences on each trial, but so too are the attentional settings that are used when responding to stimuli in each location. For example, an MI location encourages participants to frequently adopt a "focused" attentional setting to bias attention toward the relevant stimulus dimension (color) and away from the distracting dimension (word) since word and colors tend to conflict. In contrast, the MC location encourages participants to frequently adopt a "relaxed" attentional setting since words tend to coincide with the color. Consequently, when a stimulus appears in a particular location (e.g., MI location), it triggers retrieval of the associated attentional setting (i.e., focused). Crump and Milliken's (2009) observation of a CSPC effect for the diagnostic set, in addition to the inducer set, demonstrated reactive triggering of learned attentional settings by the location of the stimulus that generalized to the diagnostic items (see Gottschalk & Fischer, 2017, for evidence the stimulus must be processed as relevant to the task for such triggering to occur).

Recently, the stability of this theoretically important CSPC effect has been called into question (see Table 1 for a summary of relevant results from conflict paradigms). Hutcheon and Spieler (2017, Experiments 1 and 2) were unable to reproduce the CSPC effect for the diagnostic set in a color-word Stroop task when using either a direct replication of the design of Crump and Milliken (2009, Experiment 2, as described above) or a modified version of the design with a larger stimulus set (four items in the inducer and diagnostic sets). Not only was the CSPC effect for the diagnostic set not evident, but surprisingly they also did not observe a CSPC effect for the inducer set. Based on these patterns, Hutcheon and Spieler concluded that the CSPC effect (Crump & Milliken, 2009) may have been a Type 1 error. Further, they proposed that the CSPC effect may depend on consistency in the informativeness of the contextual cue (e.g., location). In other words, when two different types of items are presented in the same location (i.e., inducer which are MC [or in the opposite location, MI] and diagnostic which are 50% congruent), the location is not a

[2]Hereafter, we use the term "CSPC effect" as a neutral term that refers to the finding of a smaller congruency effect in the MI location compared to the MC location regardless of the mechanism that produced it and reserve the term context-specific control to imply that a CSPC effect was found under conditions that isolated control from other mechanisms.

consistent indicator of the likelihood that an item will have a word that is congruent. The idea is that participants may not use the location to guide attentional adjustments because the location signals that some items (inducer) are MC (or MI) but other items presented in the exact same location (diagnostic) are only 50% congruent. We here refer to this hypothesis as the *consistency hypothesis* (Hutcheon & Spieler, 2017), the key tenet of which is that a location may fail to function as an effective signal for the up- or down-regulation of control (i.e., may not become associated with a distinct attentional setting or may fail to trigger its retrieval) when that location does not consistently inform the PC of all items (inducer and diagnostic) in that location. Hutcheon and Spieler (Experiment 3) tested this hypothesis by presenting only an inducer set during initial (training) blocks, and in line with the hypothesis, they did observe a CSPC effect. When a diagnostic set was added to the mix in later blocks, a CSPC effect was not found. To take stock, whenever the design included an intermixing of inducer and diagnostic sets resulting in locations that did not consistently predict the PC of an item (e.g., inducer stimuli were MC and diagnostic stimuli were 50% congruent in the MC location), CSPC effects were not found (Hutcheon & Spieler, 2017).

In response to these observations, Crump et al. (2017) also tested the stability of the CSPC effect (see Table 1). Using the combined inducer/diagnostic design hitherto considered, they found mixed evidence across two experiments. Using the Stroop task, they did not replicate their original finding of a CSPC effect for the inducer set; in addition, there was no CSPC effect for the diagnostic set. Switching to a flanker task, they did not find a CSPC effect for the inducer set, but they did find an effect for the diagnostic set. The latter is not easily explained nor fully consistent with their prior findings (Crump & Milliken, 2009). In sum, these observations coincided with those of Hutcheon and Spieler (2017) and the consistency hypothesis.

However, contrasting these patterns, in two additional experiments, Crump et al. (2017) attempted to replicate the CSPC effect they previously observed for the diagnostic set using a different design (see Crump & Milliken, 2009, Experiment 1). The design differed from the design hitherto described in that the inducer set was 100% congruent or 100% incongruent (instead of MC [92% congruent] or MI [8% congruent]), which consequently yields a slightly stronger CSPC manipulation when intermixed with the 50% diagnostic set.[3] In this case, they found a CSPC effect for the diagnostic set in both the Stroop and flanker tasks indicating that this finding is stable. Unfortunately, a drawback of this design is that a CSPC effect cannot be calculated for the inducer set. Nonetheless, the finding of a CSPC effect for the diagnostic set challenges the consistency hypothesis. Ultimately, Crump et al. concluded that their original observations did not represent Type 1 errors, and that the instability of the CSPC effect in the design type first described above likely reflected the weaker CSPC manipulation (see Footnote 3), low power and/or low trial counts (including for the inducer set).

In summary, to date, failures to observe CSPC effects for the inducer set and/or diagnostic set have been attributed to statistical and methodological limitations such as errors in

[3]In the version where inducer items are 100% congruent or incongruent, the locations overall are 75% congruent or incongruent; in the version where inducer items are 92% congruent or incongruent, the locations overall are 71% congruent or incongruent.

statistical inference (Type 1 error; Hutcheon & Spieler, 2017) and use of underpowered designs that include cells with relatively few observations (Crump et al., 2017). One theoretical explanation, the consistency hypothesis (Hutcheon & Spieler, 2017), has also been proposed to explain the inconsistent findings but this hypothesis has received mixed support.

The purpose of the present study was to further address the overarching question of how to interpret failures to replicate and reproduce the CSPC effect for the inducer set and diagnostic set. The consistency hypothesis predicts a boundary condition for CSPC effects based on location consistency. When locations are consistently informative of the PC of all item types in each location, CSPC effects should be found (i.e., participants learn location-PC associations); however, when locations are inconsistently informative, CSPC effects should not be found. We further examined this hypothesis in the present study. However, our primary aim was to address the question of what *is* learned when locations are *not* consistently informative of the PC of all items in each location, as in the failed replication/reproduction attempts. Specifically, we tested two theoretically motivated learning hypotheses that may in part account for variation in the magnitude/significance of the CSPC effect from one study to the next in studies that have included both MC/MI inducer items and diagnostic items, resulting in location inconsistency and no CSPC effect for inducer or diagnostic items. Both hypotheses stem from our view that there are other opportunities for learning in the CSPC paradigm (besides associating locations with proportion congruence) that have not received adequate attention[4] and failure to consider these opportunities may offer an incomplete perspective on the stability of the CSPC effect (see Abrahamse, Braem, Notebaert, & Verguts, 2016; Egner, 2014 for broader discussions of role of learning in control).

(At Least) Three Opportunities for Learning in the CSPC Paradigm

The primary hypothesis researchers have tested when using the CSPC paradigm is whether participants learn the association between location and proportion congruency (PC). If this association is learned, then researchers assume that a CSPC effect will be observed for *both* the inducer set and the diagnostic set (Crump & Milliken, 2009; but see Hutcheon & Spieler, 2017, for alternative prediction). Hereafter we refer to this as the *location-PC learning hypothesis*. However, there are at least two other associations that could be learned in the CSPC paradigm, which have to date garnered little (if any) empirical attention.

One such association is the relationship between items and their probability of conflict. The *item-PC learning hypothesis* posits that participants may not encode the location in which items (stimuli) appear but rather they may encode each item and associate it with its PC. This hypothesis has its roots in prior studies that have demonstrated that participants do in fact learn to associate specific items (e.g., colors or pictures in a Stroop task) with their

[4]Complex stimulus-response (S-R) learning processes have, in contrast, garnered empirical attention as an alternative to the episodic retrieval account that posits that CSPC effects reflect learning of associations between a location and an attentional setting based on PC (e.g., Schmidt & Besner, 2008). Learning of such S-R associations was not the focus of the present study, however, as such learning should have worked *in favor* of observing CSPC effects for inducer items and not made such effects difficult to observe or reproduce in prior studies (as in Crump et al., 2017; Hutcheon & Spieler, 2017).

PCs, resulting in item-specific proportion congruence effects (i.e., Bugg et al., 2011; Bugg & Hutchison, 2013; Bugg & Dey, 2018; Chiu et al., 2017). For example, in a picture-word Stroop task, participants learn that pictures of some animals (e.g., birds) are MC and pictures of other animals (e.g., dogs) are MI, resulting in smaller congruency effects for MI items. In the CSPC paradigm, collapsing across locations, the inducer set comprises items that are 50% congruent (the same is true of course for items in the diagnostic set). Accordingly, if participants have learned that any item (inducer or diagnostic) is 50% congruent (i.e., each item is 50% congruent), then on any given trial, they should retrieve the exact same attentional setting (i.e., an intermediate setting between relaxed and focused; see Diede & Bugg, 2017; 2019 for evidence that participants can retrieve an attentional setting representing the "average" across different stimuli). Critically, this should result in no CSPC effect for the inducer set, a prediction that is unique to this learning hypothesis (see Table 1). Additionally, it should result in no CSPC effect for the diagnostic set. As noted in Table 1, this is the most common pattern that has been observed in past studies that failed to replicate/reproduce the CSPC effect (see e.g., Crump et al., 2017, Experiment 2; Hutcheon & Spieler, 2017, Experiments 1 and 2).

Another association that could be learned in the CSPC paradigm is the conjunctive relationship between a location-item pairing and PC (cf. Schmidt & Lemercier, 2019 for a compound-stimulus contingency learning account based on learning of highly contingent responses). The *conjunctive learning hypothesis* predicts that participants may learn that a subset of items (i.e., stimuli in the inducer set) is 75% congruent in one location but 25% congruent in the other, and a second subset of items (i.e., stimuli in the diagnostic set) is 50% congruent in one location and 50% congruent in the other location. If these associations are learned, then a CSPC effect should be found for the inducer set but a CSPC effect should not be found for the diagnostic set. In other words, different attentional settings should be retrieved for the items in the inducer set depending on location, but the same attentional setting should be retrieved for items in the diagnostic set in each location. The findings from one prior replication attempt have observed a pattern suggestive of conjunctive learning (Hutcheon & Spieler, 2017, Experiment 3; see Table 1). In that study, the first four blocks presented only an inducer set (MC vs. MI locations) and a CSPC effect was found. A diagnostic set was added in later blocks (with the inducer set, which was 100% congruent or incongruent in these blocks) and no CSPC effect was found. Although the goal of the experiment was not to test for conjunctive learning, it was theoretically possible that participants would generalize within the locations—that is, generalize their learning about the inducer items in the earlier blocks (which was statistically significant) to the diagnostic items in the same location in the later blocks (see e.g., Weidler & Bugg, 2016, for an example of a CSPC design in which participants generalized learning from inducer items in initial blocks to [a different type of] diagnostic items that were presented only in later blocks). However, participants did not generalize and consequently, the resulting pattern (CSPC effect for inducer but not diagnostic items) aligns best with the conjunctive learning hypothesis.

It is valuable to consider these opportunities for learning in the CSPC paradigm in relationship to the episodic retrieval account (Crump & Milliken, 2009). What is learned and consequently the pattern of CSPC effects across the inducer and diagnostic sets is

theoretically a consequence of how participants encode and organize (i.e., "bin") experiences on each trial during the CSPC paradigm (see Diede & Bugg, 2019). This should impact what gets stored and how different dimensions (e.g., location, item, etc.) are weighted within the episodes that are presumably formed on each trial. Our above reasoning posits that participants may bin their experiences during the task based on location (or, weight this dimension most heavily), item, or a conjunction of location and item. Importantly, how experiences are organized into bins will critically affect whether a CSPC effect is found for the inducer set and diagnostic set, neither set, or only the inducer set as illustrated in Panels A, B, and C, respectively, of Figure 1.

Present Approach and Overview of Experiments

In recent research on the CSPC effect, much attention has been paid to the diagnostic set because the goal has been to identify a design that enables conclusions to be made about context-specific control, which is possible when a CSPC effect is observed for the diagnostic set (under confound minimized conditions; see Braem et al., 2019, for further details). Accordingly, recent research has adopted the combined inducer/diagnostic design of Crump and Milliken (2009) in which CSPC effects can be calculated for inducer and diagnostic sets. However, this design has a serious shortcoming in that it does not yield a consistent pattern of results across studies (see Table 1), and presently it is unclear how to interpret these failures to replicate or reproduce the CSPC effect. Thus, we posit that a different approach is needed to better understand these failures.

The approach we adopted is to focus much of our attention theoretically and empirically on the inducer set. The rationale was as follows. Firstly, as can be seen in Table 1, the failed replication/reproduction attempts using the combined inducer/diagnostic design that permits calculation of CSPC effects for both item sets have one thing in common—a CSPC effect was not found for the inducer set. In our view, it is as striking and arguably more surprising that the CSPC effect for the inducer set, and not just the CSPC effect for the diagnostic set, has failed to replicate or reproduce in this design given that items in the inducer set are biased, allowing for all potential contributing mechanisms [e.g., control; S-R learning] to work together to produce the effect. Second, the combined inducer/diagnostic design is predicated on the assumption that participants learn attentional settings for each location based on the inducer set, and this learning induces participants to adopt location-based control that affects not only inducer items but also generalizes to diagnostic items. Yet, the failed replication/reproduction attempts indicate that participants are not learning location-PC associations much of the time. This begs the question of what participants *are* learning based on the inducer items. If we can understand what participants learn about the inducer items and the conditions under which such learning occurs, then we will be better positioned to anticipate when researchers will or will not observe CSPC effects not just for inducer items but also for diagnostic items.

This approach of considering the learning processes that support cognitive control and not just control per se gels with current theorizing in the control literature (Abrahamse et al., 2016; Egner, 2014). For example, Abrahamse et al. (2016) developed a theoretical perspective that grounds control in associative networks that bind together goals with

perceptual and motor components of a task. Attentional control, for example, was suggested to emerge from learning of associations between goals (e.g., filtering conflicting information) and features (e.g., contexts, items). While such associative learning is often context-specific, generalization of this learning (as must occur for a CSPC effect to be observed for diagnostic items) depends on feature overlap. The assumption is that a learned association can generalize to new items (e.g., within a context) when there is feature overlap (e.g., the location of the diagnostic items is the same as the inducer items that promoted the learning of the associations). Applying this perspective to the replication/reproduction debate in the CSPC literature, a clear possibility is that the feature (i.e., location) that researchers assumed to be associated with a given goal (i.e., what was learned/on what basis experiences were binned) may not be the feature that participants are associating with goals based on their experience with the inducer set. Given the reasonable assumption that it should be more difficult to observe a CSPC effect for the diagnostic set if no location-PC learning is observed for the inducer set, the present study attempted to shed light on the important question of why CSPC effects have been difficult to observe by examining what is being learned for the inducer set.

In the present experiments, we utilized a picture-word Stroop task (Bugg, Jacoby, & Chanani, 2011; Bugg & Dey, 2018) and a novel variant of the CSPC design. To our knowledge this task has not previously been used in a CSPC paradigm. However, there is ample evidence that performance on the task is sensitive to PC manipulations, including those that produce item[5]-PC learning (Bugg et al., 2011; Bugg & Chanani, 2011; Bugg & Dey, 2018; see also Bugg & Chanani, 2011; Gonthier, Braver, & Bugg, 2016). The design variant we used is also novel in the CSPC literature. In this variant, the inducer and diagnostic sets partially cross (i.e., an inducer picture of a cat can appear not only with the words from the inducer set [CAT, BIRD] but additionally with the words from the diagnostic set [FISH, DOG]; see Method for details; cf. Bugg et al., 2011; Bugg & Dey, 2018). This is not ideal if the goal is to isolate context-specific control independent of bottom-up, feature-based influences on performance in the diagnostic set (Braem et al., 2019); however, for the reasons described above, that was not the central goal of the present investigation. The rationale for using this variant extends that of Hutcheon and Spieler (2017, Experiment 2) who used four-item sets instead of two-item sets to potentially discourage S-R learning (cf. Bugg & Hutchison, 2013). When the inducer set is only two items (e.g., birds and cats) and these items only appear with the corresponding words from that set (BIRD and CAT) then participants can learn to produce the congruent response option (e.g., say “bird” when a bird is shown) in the MC location and the incongruent response option (e.g., say “bird” when a cat is shown) in the MI location (see Schmidt & Lemercier, 2019 for evidence of compound-stimulus contingency learning). Such S-R learning could theoretically preclude learning of location-PC associations, resulting in no CSPC effect for the inducer set (and consequently the diagnostic set; cf. Bugg, 2014). While the findings from Hutcheon and Spieler’s four-item set design did not confirm this speculation, their design used fully separate sets (the colors in each set appeared only

[5]We use the term “item” or “item-specific” for consistency with prior literature; however, note it has been established that the items are encoded and represented at the category level and not the exemplar level (Bugg & Dey, 2018). Whether we use “item” or “category” does not fundamentally change the conclusions of our study.

with words from that set). We reasoned that if an animal picture only appears with words within its item set (inducer or diagnostic), the boundaries between item sets may be quite distinctive. Consequently, this may favor certain learning opportunities (perhaps especially the learning of location-item conjunctions but additionally item PC learning), opportunities that have in fact dominated in past replication/reproduction attempts (see Table 1). By using the current design in which the sets do partially cross (relative to Hutcheon & Spieler's 4-item set design), we sought to bias learning of location-PC associations, that is, to encourage participants to bin experiences based on location rather than salient item or location-item distinctions. Thus, a secondary aim of the present research was to determine whether there may be a more optimal design for promoting learning of location-PC associations, which would be a first step in trying to produce CSPC effects for the inducer set (and ultimately a diagnostic set).

An overview of the designs used in each experiment is presented in Figure 2. In Experiment 1 we used our variant of the combined inducer/diagnostic design to try to reproduce the CSPC effects observed by Crump and Milliken (2009), test the consistency hypothesis (which predicted no CSPC effects), and determine which of the three learning hypotheses was most consistent with the results. To foreshadow, we also did not find a CSPC effect for the inducer set nor did we find a CSPC effect for the diagnostic set (cf. Crump et al., 2017; Hutcheon & Spieler, 2017). In other words, Experiment 1 represented another failed reproduction attempt, supported the consistency hypothesis, and provided preliminary support for the item-PC learning hypothesis. In Experiment 2, we eliminated the diagnostic set, thereby ensuring locations were consistent signals of PC, a pre-requisite for CSPC effects according to the consistency hypothesis. Critically, Experiment 2 also enabled us to further test the item-PC learning hypothesis, which predicted no CSPC effect. However, a CSPC effect was found (i.e., there was not evidence of item-PC learning under conditions of location consistency). In sum, in the first two experiments, we a) reproduced key prior CSPC findings previously observed with the color-word Stroop task in picture-word Stroop task variants suggesting additional experimentation with these variants was appropriate for furthering our theoretical understanding of the general phenomenon of CSPC effects in conflict tasks, and b) found preliminary support for the view that selectively under conditions of location inconsistency, conditions where evidence for location-PC learning has not been observed in past reproduction/replication attempts (Crump et al., 2017; Hutcheon & Spieler, 2017), participants learn item-PC associations.

The final two experiments directly tested the two learning hypotheses that serve as contenders to the location-PC learning hypothesis, the item-PC learning hypothesis and the conjunctive learning hypothesis (see Figure 2 for general depiction of designs, which are further detailed along with competing hypotheses in the introduction to each experiment). In Experiment 3, we pitted unique predictions of the item-PC learning hypothesis against the predictions of the location-PC and conjunctive learning hypotheses. In Experiment 4, we devised a design that enabled us to seek evidence for conjunctive learning that could not be explained by either location-PC or item-PC learning. While indirect support exists for the item-PC and conjunctive learning hypotheses (see Table 1; for item-PC, see also Experiment 1), direct support for either hypothesis would strengthen the view that participants may learn other associations in the CSPC paradigm besides the customary association between location

and PC, resulting in the absence of CSPC effects for both sets of items or one set of items, respectively. Critically, such support would have important implications for the ongoing discussion of the stability of CSPC effects.

Experiment 1

Experiment 1 used a variant of the combined inducer/diagnostic CSPC design (cf. Hutcheon & Spieler, 2017, Experiment 2) in a picture-word Stroop task in which two items served in the inducer set and two in the diagnostic set and the sets partially crossed. Using this variant, we attempted to reproduce the CSPC effect originally observed by Crump and Milliken (2009, Experiment 2) for which there is currently weak evidence (Crump et al., 2017; Hutcheon & Spieler, 2017). Assuming learning of location-PC associations, there should be a CSPC effect. However, the consistency hypothesis predicts a null CSPC effect for inducer and diagnostic sets given the mixing of these sets (with different PCs) in each location. To ensure that a null, if observed, would be informative, we tested 60 participants (power analyses performed by Crump et al., 2017 and Hutcheon & Spieler, 2017 based on estimates from Crump & Milliken, 2009, Experiment 2, indicated that a sample size of 32 was needed to achieve power of .80 to detect a CSPC effect for the diagnostic set). The predictions of the item-PC learning hypothesis (no CSPC effects for either item set) and conjunctive learning hypothesis (a CSPC effect for the inducer set but not for the diagnostic set) are the standard predictions for the combined inducer/diagnostic CSPC design, as illustrated in the bottom portion of Figure 1.

Method

Participants.—60 undergraduates (40 females; Age $M = 19.48$, $SD = 1.17$) from Washington University in Saint Louis participated in this study. Participants earned class credit for participation. All participants were native English speakers and had normal or corrected-to-normal vision. The experiment crashed for one participant between the third and fourth block and their data were excluded.

Design and Stimuli.—A 2 × 2 × 2 within-subjects design was used with factors of trial type (congruent or incongruent), location PC (MC or MI), and set type (inducer or diagnostic). The picture-word Stroop stimuli were line drawings of four birds, four cats, four fish, and four dogs with animal words (BIRD, CAT, FISH, DOG) superimposed (see Bugg et al., 2011; Bugg & Dey, 2018). On congruent trials, the identity of the picture and the word matched (i.e., a picture of a cat with the word CAT superimposed) whereas on incongruent trials the identity of the picture and the word conflicted (i.e., a picture of a cat with the word DOG superimposed). Stimuli appeared in the upper part of the screen or the lower part of the screen. The upper location was presented 6.9 cm above the center of the screen; the lower location was presented 6.9 cm below the center of the screen. Both locations were centrally aligned on the horizontal dimension. One location was MC and the other was MI (counterbalanced across participants).

The inducer set comprised two of the animal categories (e.g., bird and cat pictures) and the diagnostic set comprised the other two animal categories (e.g., dog and fish pictures). Assignment was counterbalanced across individuals. The inducer set established the PC of

the locations, and the diagnostic set assessed the effects of the induction. As shown in Table 2, the inducer set was 75% congruent in one location (e.g., MC) and 25% congruent in the other location (e.g., MI). The diagnostic set was unbiased (PC: 50% congruent) and frequency matched across locations but presented less frequently overall than the inducer set so that the locations remained biased when the sets were combined (i.e., locations were 70% congruent and 30% congruent, respectively). Deviating from prior CSPC designs, the two sets partially crossed (cf. Bugg et al., 2011; Bugg & Dey, 2018). As can be seen in Table 2, If bird and cat pictures were in the inducer set, those pictures also appeared with the words FISH and DOG in addition to the words BIRD and CAT; conversely, if fish and dogs were in the diagnostic set, those pictures also appeared with the words BIRD and CAT in addition to the words FISH and DOG. However, the pictures did not cross sets (i.e., pictures of birds and cats, in this example, never served as diagnostic items).

Collapsing across locations, 50% of trials were congruent and 50% were incongruent, and 50% of trials were presented in each location meaning participants could not predict congruency or location on a trial-by-trial basis. See Figure 2 for an illustration of the design and an overview of the design of all subsequent experiments.

Procedure.—Participants consented to participate and were then given instructions for the task. In keeping with previous studies (Bugg et al., 2011; Bugg & Dey, 2018), participants were instructed to name aloud the animal in the picture and not the word written on the animal. Participants were instructed to respond as quickly as possible without sacrificing accuracy by using the general name of the animal (e.g., bird) and not a more specific name (e.g., robin).

In this and all subsequent experiments, E-prime 2.0 software was used to present stimuli on a 17-inch LCD monitor (Psychological Software Tools, Pittsburgh, PA) and vocal responses were detected by a microphone connected to a voice-key using the PST serial response box (Psychological Software Tools, Pittsburgh, PA). On each trial, a fixation cross appeared centrally for 500 ms followed by the stimulus in either the upper or lower location. The stimulus remained on screen until a vocal response was detected by a voice key. After the voice key was triggered, the experimenter coded the participant's response using a keyboard (i.e., indicated what animal name was emitted). Sounds that unintentionally triggered the voice key (e.g., "um", a cough) or vocal responses that were otherwise imperceptible or unintelligible were coded as "scratch trials" by the experimenter and excluded from analysis. After experimenter coding, a blank interstimulus interval appeared for 500 ms. Reaction time (ms) and error rate were recorded.

Participants completed 16 practice trials. The PCs of the inducer and diagnostic sets in each location were consistent with that of the experimental lists. Practice lists were not included in data analyses. After completion of the practice lists, participants were presented with 480 experimental trials that were randomly presented without replacement according to the frequencies listed in Table 2. A brief break was provided after every 120 trials, resulting in four blocks. At the end of the experiment, participants were debriefed.

Results

The raw data for this and all subsequent experiments can be accessed at: osf.io/n7ecq/ In this and all subsequent experiments, the alpha level was .05, Bayesian analyses are reported for all theoretically significant null effects in the form of BF_{01} values where a value between 1 and 3 means anecdotal evidence for the null hypothesis and a value between 3 and 10 means substantial evidence for the null hypothesis (Wagenmakers, Wetzels, Borsboom, & Van Der Maas, 2011), and RTs faster than 200 ms or slower than 3,000 ms were excluded consistent with our prior research with this task (e.g., Bugg et al., 2011; Bugg & Dey, 2018). In Experiment 1, this resulted in the removal of < 1% of trials. In addition, only correct responses were included in the analysis of RT. Mean RTs and error rates are presented in Table 3.

Following Hutcheon and Spieler (2017) and Crump et al. (2017), we performed separate 2 (Location PC: MC vs. MI) × 2 (Trial Type: Congruent vs. Incongruent) × 2 (Set Type: Inducer vs. Diagnostic) within-subjects ANOVAs on RT and error rate[6].

Reaction time.—There were main effects of trial type, $F(1, 58) = 258.28$, $p < .001$, $\eta_p^2 = .82$, and set type, $F(1, 58) = 38.79$, $p < .001$, $\eta_p^2 = .40$, and trial type interacted with set type such that the congruency effect was smaller for the diagnostic items ($M = 85$ ms) than the inducer items ($M = 100$ ms), $F(1, 58) = 10.32$, $p = .002$, $\eta_p^2 = .15$. However, and most critically, there was not a Location PC × Trial Type interaction, $F < 1$, $BF_{01} = 7.93$, nor a Location PC × Trial Type × Set Type interaction, $F(1, 58) = 1.79$, $p = .187$, $BF_{01} = 5.56$. The congruency effect ($incongruent_{RT} - congruent_{RT}$) was equivalent in the MC and MI locations for both inducer and diagnostic sets (see Figure 3).

Error rate.—There was a main effect of trial type, $F(1, 58) = 57.14$, $p < .001$, $\eta_p^2 = .50$. However, all other effects were non-significant including the Location PC × Trial Type interaction, $F(1, 58) = 1.18$, $p = .283$, $BF_{01} = 4.60$, and the Location PC × Trial Type × Set Type interaction, $F < 1$, $BF_{01} = 4.71$. Again, this suggests that the congruency effects did not differ between MC and MI locations for either set type.

Discussion

The key finding from Experiment 1 was that there was not a CSPC effect either for inducer or diagnostic items in the picture-word Stroop task despite our attempt to encourage location-PC learning via the crossed sets design, and the Bayesian evidence indicated substantial evidence in favor of the null. This represents a failure to reproduce the findings of Crump and Milliken (2009, Experiment 2), which is the same pattern that was observed in three prior studies that used two different designs[7] in the color-word Stroop task (Crump et al., 2017, Experiment 2; Hutcheon & Spieler, 2017, Experiments 1 and 2; see Experiment

[6]There is mixed evidence that CSPC effects grow with time on task (Crump & Milliken, 2009; Crump et al., 2017; Hutcheon & Spieler, 2017). In none of the present experiments did the factor of half (1st half of trials vs. 2nd half of trials) interact significantly with the theoretically relevant interaction effects of interest (all Fs < 1.23).

[7]Similar to the present design, Crump et al. (2017, Experiment 2) and Hutcheon and Spieler (2017, Experiment 1) had locations that were PC 75 or PC 25, whereas our locations were PC 70 and PC 30. In addition, their experiments comprised 384 trials whereas ours had 480 trials.

3 of Crump et al., 2017 for a partial failure in the flanker task). As such, Experiment 1 supports the consistency hypothesis positing location inconsistency is a boundary condition for CSPC effects.

With respect to the question of what participants did learn under the conditions of location inconsistency in Experiment 1, when comparing the pattern illustrated in Figure 3 to the middle pattern in the lower portion of Figure 1, it appears the present pattern (save for the overall larger congruency effects for inducer items, which was not anticipated by any account) is consistent with the learning of item-PC associations. The item-PC learning hypothesis predicted no CSPC effect for inducer or diagnostic items in the present design. These findings therefore align with the possibility that under conditions of location inconsistency, participants may learn item-PC associations rather than location-PC associations. An alternative possibility, however, is that item-PC learning may be a default learning process in (some) CSPC paradigms, emerging even under conditions of location consistency. One way to adjudicate between these theoretical possibilities is to remove the diagnostic items (cf. Hutcheon & Spieler, Experiment 3).

Experiment 2

Experiment 2 used the same picture-word Stroop task as in Experiment 1 but implemented a CSPC design with only inducer items to create conditions of location consistency (see also Crump et al., 2006). This allowed us to test the prediction of the consistency hypothesis that a CSPC effect should now be observed, and further examine whether the picture-word Stroop task yields similar results to the color-word Stroop task when CSPC is manipulated. Of theoretical import, the design also afforded us the opportunity to further test the role of item-PC learning, including whether item-PC learning additionally may occur under conditions of location consistency. Because item-PC learning was a plausible account of the Experiment 1 findings, suggesting that participants may have binned their experiences according to item despite our partially crossed sets design, we elected to include a location counting manipulation in Experiment 2. A random half of the participants performed the task while engaging in a secondary counting task. Such counting tasks have been effective in biasing participants to process contextual cues that otherwise may not be sufficiently attended, resulting in CSPC effects in conditions that do not typically yield them (e.g., when shapes serve as contextual cues; Crump, Vaquero, & Milliken, 2008). To our knowledge no prior study has had to use this manipulation to bias participants' attention toward a location cue but considering the findings of Experiment 1 (i.e., suggesting item-PC learning), we elected to employ it. Regardless of whether participants count stimuli in a certain location or not, the consistency hypothesis anticipates a CSPC effect (see Crump et al., 2006). A strict form of the item-PC learning hypothesis (as detailed in the *(At Least) Three Opportunities for Learning in the CSPC Paradigm* subsection of Introduction) does not; however, it is possible that the results would support the item-PC learning hypothesis in the no counting condition but support the location-PC learning hypothesis in the counting condition where participants' attention is drawn to location.

Method

Participants.—64 undergraduates (50 females; Age M = 19.64, SD = 1.31) from Washington University in Saint Louis participated in this study. Participants earned class credit for participation. All participants were native English speakers and had normal or corrected-to-normal vision.

Design and Stimuli.—A 2 × 2 × 2 mixed-subjects design was used, with within-subjects factors of trial type (congruent or incongruent) and location PC (MC or MI) and a between subjects factor of counting (counting [n = 32] or not counting [n = 32]). Participants were randomly assigned to the counting condition. The primary difference from Experiment 1 was the absence of a diagnostic set. Only an inducer set was used in Experiment 2 meaning that all four items (animals) were MC (80% congruent) in one location and MI (20% congruent) in the other location (see Figure 2; cf. Crump et al., 2006). Assignment of PC to location was counterbalanced across subjects (see Table 4 for frequency of trial counts). The design and stimuli were otherwise identical to Experiment 1.

Procedure.—The procedure was identical to Experiment 1 with the following exception. In the counting condition, participants were instructed to keep a running count in their head of how many times any stimulus appeared in one of the two locations (e.g., in the upper location). We balanced whether participants were asked to count stimuli in the upper or lower location and whether that location was MC or MI across subjects. Participants reported the number of trials they counted in the designated location to the experimenter at the end of each block.

Results

The RT trim eliminated < 1% of all trials. Mean RTs and error rates from the picture-word Stroop task are presented in Table 5. Counting estimates indicated that participants were reasonably accurate in performing the counting task (actual number of times stimuli appeared in counted location = 240, mean reported number of times stimuli appeared = 206 [mean accuracy = 86%]). 2 (Location PC: MC vs. MI) × 2 (Trial Type: Congruent vs. Incongruent) × 2 (Counting: Counting or No Counting) mixed-subjects ANOVAs were conducted separately for RT and error rate on the inducer items.

Reaction time.—As expected given the secondary task demand, participants in the counting condition (M = 792) were slower than those in the no counting condition (M = 705), $F(1, 62) = 7.34$, $p = .009$, $\eta_p^2 = .11$. In addition, there was a significant main effect of trial type, $F(1, 62) = 424.48$, $p < .001$, $\eta_p^2 = .87$. Most critically, the Location PC × Trial Type interaction was significant indicating a CSPC effect, $F(1, 62) = 17.18$, $p < .001$, $\eta_p^2 = .22$. The congruency effect was reduced in the MI location (M = 96 ms) compared to the MC location (M = 120 ms). This effect did not vary as a function of the counting manipulation ($F < 1$ and $BF_{01} = 3$ for three-way interaction; see Figure 4).

Error rate.—For error rate, participants in the counting condition performed worse (M = .025, SE = .003) compared to those in the no counting condition (M = .019, SE = .003),

though this difference was marginal, $F(1, 62) = 3.28$, $p = .075$. In addition, there were main effects of Location PC, $F(1, 62) = 14.38$, $p < .001$, $\eta_p^2 = .19$, and trial type, $F(1, 62) = 109.31$, $p < .001$, $\eta_p^2 = .64$. There was also a Counting × Location PC interaction, $F(1, 62) = 4.93$, $p = .03$, $\eta_p^2 = .07$ (i.e., MI location had .003 fewer errors than MC for no counting condition but .010 fewer errors for the counting condition), and a Counting × Trial Type interaction, $F(1, 62) = 6.69$, $p = .012$, $\eta_p^2 = .10$ (i.e., the congruency effect in error rate was .023 for the no counting condition but .037 for the counting condition). Of most relevance, the Location PC × Trial Type interaction was significant, $F(1, 62) = 9.69$, $p = .003$, $\eta_p^2 = .14$, such that the congruency effect was reduced in the MI location ($M = .023$) compared to the MC location ($M = .037$). The three-way interaction was marginal, $F(1, 62) = 3.95$, $p = .051$, as a result of the counting condition having a nominally larger overall CSPC effect (difference between the congruency effects in error rate when comparing MC and MI locations = .024) compared to the no counting condition (difference = .005).

Discussion

In Experiment 2, a CSPC effect was found for an inducer set indicating a smaller congruency effect in the MI location compared to the MC location. For RT, the CSPC effect did not differ based on the counting condition suggesting the counting manipulation was not necessary for producing the CSPC effect; for error rate, there was a trend for a larger CSPC effect in the counting condition. Like the findings of Experiment 1, the findings of Experiment 2 reproduced a pattern observed previously in the color-word Stroop task, in this case the CSPC effect Crump and colleagues (2006) first observed in an inducer-set only version of the color-word Stroop task. This indicates that our picture-word Stroop task is sensitive to the location-based PC manipulation (cf. Crump et al., 2006).

The fact that a CSPC effect was evidenced in the present experiment is theoretically in line with the consistency hypothesis since locations were consistent signals of PC in Experiment 2 (Hutcheon & Spieler, 2017). As anticipated by this hypothesis, the CSPC effect was found for the inducer set in the absence of a diagnostic set (Experiment 2) but was not found when both an inducer set and a diagnostic set were included thereby creating location inconsistency (Experiment 1).[8] This converges with the findings observed for the inducer set in Hutcheon and Spieler's (2017) four-item set designs (i.e., CSPC effect in training blocks of Experiment 3 that included only an inducer set, but no CSPC effect in Experiment 2 when design included a diagnostic set). Furthermore, this suggests that in the face of location consistency, participants may be more apt to learn about location in the form of location-PC associations as Hutcheon and Spieler (2017) have forwarded. However, because there was no diagnostic set, one cannot rule out that participants may have alternatively learned about location-item conjunctions. Either way, what is clear is that in the face of

[8]One might suggest, however, that the differing results for the inducer set across Experiments 1 and 2 may instead be due to the slightly more extreme PCs used in Experiment 2 (e.g., 75% in Experiment 1 vs. 80% in Experiment 2). However, this explanation does not fit with several prior studies that failed to find a CSPC effect for the inducer set even though the set had much more extreme PCs than used in our experiments (92% vs. 8% congruent; Hutcheon & Spieler 2017, Experiments 1 and 2; Crump et al., 2017, Experiment 2). Like Experiment 1, these studies included diagnostic items and thus, too, can be explained by the consistency hypothesis.

location consistency, participants did *not* learn item-PC associations, which brings us to the next theoretical point.

The findings of Experiment 2 have important implications for the item-PC learning hypothesis—if participants had learned item-PC associations in Experiment 2, then there should not have been a CSPC effect. However, a CSPC effect was found in both the counting and no counting conditions. Considered alongside the findings of Experiment 1, which were consistent with the item-PC learning hypothesis, the present findings lend support to the theoretical possibility that learning of item-PC associations (i.e., binning of experiences into episodes based on items) may be more apt to occur when it is difficult to learn location-PC associations. In other words, particularly in the face of location inconsistency (i.e., the condition under which prior failed replication/reproduction attempts have occurred), participants may learn other associations that are present in the experimental context such as item-PC associations. Hutcheon and Spieler (2017) assumed this may be possible (see Figure 2 from their study). Experiment 3 provides the first direct test for the learning of item-PC associations in the CSPC paradigm, using again the paradigm which combines inducer and diagnostic items.

Experiment 3

Experiments 1 and 2 demonstrated the boundary condition set forth by the consistency hypothesis—CSPC effects were not found in the presence of location inconsistency (Experiment 1) but were found in the presence of consistency (Experiment 2; see Hutcheon & Spieler, 2017). At the same time, the findings of Experiments 1 and 2 motivate further exploration of the item-PC learning hypothesis, and namely the possibility that item-PC learning may be particularly operative in the face of location inconsistency, as in Experiment 1. That is, participants may have binned their conflict experiences based on items rather than locations in the combined inducer/diagnostic design in Experiment 1 resulting in no CSPC effect for either set (see Figure 1, Panel B). But the absence of CSPC effects in Experiment 1 is not compelling support for the item-PC learning hypothesis because it is based on a set of null effects.[9]

To directly test the item-PC learning hypothesis, we conducted Experiment 3. Experiment 3 used the combined inducer/diagnostic design just as in Experiment 1. Critically, however, unlike the standard design (and that of Experiment 1), the PC of the locations was determined by *different* inducer items (see Figure 2). For example, birds/cats were MC in the upper location whereas fish/dogs were MI in the lower location. In addition, while the diagnostic items in each location were 50% congruent and differed from the inducer items in each location (as is standard), the diagnostic items presented in one location differed from the diagnostic items presented in the opposite location (see Figure 2). Expanding the above example, fish/dogs were 50% congruent in the upper location and birds/cats were 50%

[9]It could also be argued that it is insufficient evidence because an alternative account is possible. That is, participants could have modulated attention based on the list PC of 50% (i.e., binned every trial during the task into one bin, the “list” bin, that was 50% congruent). This is unlikely a strong alternative given prior reports of learning of item-PC associations in 50% congruent lists (e.g., Bugg et al., 2011; Bugg & Dey, 2018; Hutchison, 2013) but the design of Experiment 1, unlike the design of Experiment 3, cannot exclude it.

congruent in the lower location. As in Experiment 1, this resulted in locations that differed in PC (e.g., upper MC = 70% congruent, lower MI = 30% congruent, using above examples). However, unlike Experiment 1, this also resulted in items that differed in PC (e.g., birds/cats = 70% congruent; fish/dogs = 30% congruent, using above examples). Critically, the latter enabled us to seek support for the item-PC learning hypothesis that would not be dependent on null effects.

The predicted performance patterns and hypothetical bins for Experiment 3 are illustrated in Figure 5: First, for items in the inducer set, using the above example, the item-PC learning hypothesis predicts that birds/cats in the upper (MC) location should produce larger congruency effects than fish/dogs in the lower (MI) location because birds/cats are MC and fish/dogs are MI (see Panel B). However, this effect alone would not uniquely support the item-PC learning hypothesis. Indeed, the effect can also be conceived of as the typical CSPC pattern whereby congruency effects are larger in the upper (MC) location than the lower (MI) location, consistent with the location-PC learning hypothesis (see Panel A). Second, for items in the diagnostic set (again, using the example above), the item-PC learning hypothesis predicts that the direction of the typical CSPC effect should be *reversed*. Birds/cats in the lower (MI) location should produce *larger* congruency effects than the fish/dogs in the upper (MC) location because although items in the diagnostic set are all 50% congruent, birds/cats overall occurred more frequently as congruent items (70% congruent) than did fish/dogs overall (30% congruent). In other words, if participants bin their experiences during the task based on the item (i.e., whether it is a bird, cat, fish, or dog), regardless of location, as the item-PC learning hypothesis contends, then the bird/cats bin will be MC in this experiment and associated with a more relaxed attentional setting than the fish/dog bin which is MI (see Panel B). Critically, a location-PC learning hypothesis would not predict this second pattern--instead, it predicts the typical CSPC pattern whereby items in the lower (MI) location should produce smaller congruency effects than items in the upper (MC) location, including the diagnostic items (i.e., birds/cats in the lower [MI] location should also lead to smaller congruency effects than fish/dogs in the upper [MC] location; see Panel A). In other words, the location-PC learning hypothesis predicts that there should be a difference in congruency effects between locations for both sets of items, but that difference should be in the same direction for the inducer and diagnostic items. Similarly, a conjunctive learning hypothesis would not predict the second pattern of a reversed CSPC effect for diagnostic items (see Panel C). Only the item-PC learning hypothesis predicts both patterns, resulting in the specific three way interaction pattern between PC, trial type, and set type detailed above (i.e., a difference in congruency effects between locations for both inducer and diagnostic items with the direction of the typical CSPC difference [MC location > MI location] observed for the inducer set but reversed for the diagnostic set [MI location > MC location]). Regarding the consistency hypothesis, because locations are inconsistent (the PC of the inducer and diagnostic items in each location differs), the prediction is that a CSPC effect should not be observed even for the inducer set.

Method

Participants.—68 undergraduates (51 females; Age $M = 19.82$, $SD = 1.25$) from Washington University in Saint Louis participated in this study[10]. Participants earned class

credit for participation. All participants were native English speakers and had normal or corrected-to-normal vision.

Design and Stimuli.—The design of Experiment 3 was a 2 × 2 × 2 within-subjects design with factors of trial type (congruent or incongruent), location PC (MC or MI), and set type (inducer or diagnostic). As in Experiment 1, there were inducer items and diagnostic items. Two animals (e.g., cats and birds) served as the MC (75% congruent) inducer items in one location (e.g., upper) and the other two animals (e.g., fish and dogs) served as the MI (25% congruent) inducer items in the other location (e.g., lower; see Figure 2). The diagnostic items were 50% congruent, but the specific items serving this role differed across locations (like the inducer items in this experiment) with the items (animals) assigned to each location not matching the inducer items in a given location. For example, if cats and birds played the role of MC inducer items in the upper location, then cats and birds played the role of 50% congruent diagnostic items in the lower location. The reverse was true for fish and dogs (e.g., MI inducer items in lower location and 50% congruent diagnostic items in upper location; see Figure 2). This was counterbalanced across participants as was the assignment of PC to location.

Combining inducer and diagnostic sets within a location resulted in locations that were 70% (MC location) or 30% congruent (MI location), as in Experiment 1. Collapsing across all instances regardless of location, one set of animals (e.g., birds/cats in above example) was 70% congruent (75% congruent in inducer set and 50% congruent in diagnostic set) and the other (e.g., fish/dogs in above example) was 30% congruent (25% congruent in inducer set and 50% congruent in diagnostic set). Table 6 presents the stimulus frequencies in Experiment 3; the design and stimuli were otherwise identical to Experiment 1.

Procedure.—The procedure was identical to Experiment 1.

Results

The RT trim eliminated < 1% of all trials. Mean RTs and error rates from the picture-word Stroop task are presented in Table 7. As in Experiment 1 (see also Crump et al., 2017; Hutcheon & Spieler, 2017), we performed separate 2 (Location PC: MC vs. MI) × 2 (Trial Type: Congruent vs. Incongruent) × 2 (Set Type: Inducer vs. Diagnostic) within-subjects ANOVAs on RT and error rate.

Reaction time.—There were main effects of trial type, $F(1, 67) = 421.82$, $p < .001$, $\eta_p^2 = .86$, and set type, $F(1, 67) = 62.17$, $p < .001$, $\eta_p^2 = .48$, and a Set Type × Location PC interaction, $F(1, 67) = 16.09$, $p < .001$, $\eta_p^2 = .19$. The Location PC × Trial Type interaction was not significant, $F < 1$, $BF_{01} = 6.04$, but most critically, the three way Location PC × Trial Type × Set Type interaction was significant, $F(1, 67) = 15.90$, $p < .001$, $\eta_p^2 = .19$ (see Figure 6). Decomposing the interaction, for inducer items, the congruency effect was larger in the MC location ($M = 113$ ms) than the MI location ($M = 83$ ms), $F(1, 67) = 16.46$, $p <$

[10]Data from nine subjects were collected as a pilot but one fell asleep resulting in eight subjects. No inferential tests were performed on these data. We combined these data with the data collected thereafter from 60 additional subjects

.001, $\eta_p^2 = .20$. However, for diagnostic items, the difference was reversed. The congruency effect was larger in the MI location (M = 119 ms) than the MC location (M = 99 ms), $F(1, 67) = 4.85$, $p = .031$, $\eta_p^2 = .07$.[11]

Error rate.—Similar to RT, there was a main effect of trial type, $F(1, 67) = 64.14$, $p < .001$, $\eta_p^2 = .49$, and a Set Type × Location PC interaction, $F(1, 67) = 4.16$, $p = .045$, $\eta_p^2 = .06$. The Location PC × Trial Type interaction was not significant, $F(1, 67) = 2.31$, $p = .133$, $BF_{01} = 3.32$, but the three way Location PC × Trial Type × Set Type interaction was significant, $F(1, 67) = 4.05$, $p = .048$, $\eta_p^2 = .06$. Decomposing the interaction, for inducer items, the congruency effect was larger in the MC location (M = .049) than the MI location (M = .026), $F(1, 67) = 9.00$, $p = .004$, $\eta_p^2 = .12$. For diagnostic items, the congruency effect was nominally larger in the MI location (M = .039) than the MC location (M = .035), $F < 1$, $BF_{01} = 4.90$.

Discussion

Experiment 3 directly tested the item-PC learning hypothesis in a CSPC paradigm in which participants could have instead learned associations between location and PC. The findings provided the first evidence for item-PC learning in the CSPC paradigm. That evidence comprised a specific three-way interaction pattern whereby a CSPC effect in the standard direction was observed for items in the inducer set (i.e., MC inducer items in the MC location had a larger congruency effect than MI inducer items in the MI location) but a reversed CSPC effect was observed for items in the diagnostic set (i.e., 50% diagnostic items in the MI location had a larger congruency effect than 50% diagnostic items in the MC location). This pattern is uniquely consistent with the predictions of the item-PC learning hypothesis, and at the same time counters the predictions of alternative hypotheses. The customary location-PC learning hypothesis anticipated that CSPC effects would be in the standard direction for both sets of items, and the conjunctive learning hypothesis anticipated that a standard CSPC effect would be found for the inducer set and no difference would be found for the diagnostic set.

The findings of Experiment 3 are also relevant to the consistency hypothesis (Hutcheon & Spieler, 2017). This hypothesis predicts that a CSPC effect should not be observed when the combined inducer/diagnostic design is used. However, a CSPC effect was found for the inducer items. Of course, the collective findings indicated that this effect was not based on location-PC learning processes, which the consistency hypothesis assumes. Rather, as

[11]An exploratory analysis was performed following Thomas Hutcheon's suggestion to compare a) the congruency effects between the birds/cats when they played the role of mostly congruent inducer items to the birds/cats when they played the role of 50% congruent diagnostic items in the mostly incongruent location (i.e., the outermost bars in Figure6) , and b) the fish/dogs when they played the role of mostly incongruent inducer items to the fish/dogs when they placed the role of 50% congruent diagnostic items in the mostly congruent location (i.e., the innermost bars in Figure 6). The means for the two conditions in the first comparison were not different, $t(67) = .75$, $p = .46$, but the means for the two conditions in the second comparison were different, $t(67) = 2.09$, $p = .04$. The *idealized* performance predictions in Figure 5 anticipate that congruency effects may be similar for the items in each comparison based on item-PC learning. However, these comparisons are not optimal tests of these predictions because they represent comparisons of inducer to diagnostic items, which pose an interpretational challenge because the two types of items differ dramatically in frequency of presentation (192 presentations vs. 48 presentations for inducer and diagnostic items, respectively, in each comparison). In both cases the congruency effect was larger for the diagnostic items. This may have worked in favor of the item-PC learning hypothesis in the first comparison and against it in the second comparison.

detailed above, the interaction pattern strongly supports the item-PC learning hypothesis. In so doing, Experiment 3 provides further evidence favoring the possibility that item-PC learning in the CSPC paradigm may be more likely to prevail under conditions of location inconsistency (i.e., in the combined inducer/diagnostic design), a speculation that emerged when comparing the results of Experiments 1 and 2. We reserve discussion of this possibility for the General Discussion.

Regarding the overarching aim of the present study, the findings of Experiment 3 have important implications regarding the reproducibility of the CSPC effect. In particular, the effect of item-PC learning processes may be mistook for a failure to reproduce the CSPC effect in the standard CSPC paradigm because both item-PC learning and failures to reproduce the CSPC effect are characterized by the same signature in the standard paradigm (see Panel B of Figure 1)—an absence of CSPC effects for inducer and diagnostic items. Before further discussing implications for reproducibility, we next examine the potential for conjunctive learning processes to influence performance. As noted in the introduction, conjunctive learning processes could also masquerade as failures to reproduce the CSPC effect particularly when such failures take the form of a CSPC effect being observed for the inducer items but not for the diagnostic items (see Panel C of Figure 1). However, to the best of our knowledge, there has not been a direct test of the conjunctive learning hypothesis.

Experiment 4

The final experiment in this series tested the conjunctive learning hypothesis which proposes that participants may learn a conjunction between location-item combinations and PC. That is, in the standard CSPC paradigm, participants may learn that inducer items are MC in one location (e.g., upper) and MI in the other (e.g., lower), and simultaneously learn that diagnostic items are 50% congruent in the upper and lower locations. If such conjunctive learning occurs, the pattern that results should be a CSPC effect for inducer items but no CSPC effect for diagnostic items (see Panel C of Figure 1). However, presently the potential for conjunctive learning (i.e., associating location-item with PC) is largely speculation, and the anticipated performance pattern in the standard paradigm includes a null effect that could just as easily be explained by an alternative account, namely that the CSPC effect is not generalizable beyond the inducer set to the diagnostic set.

Therefore, to directly test for conjunctive learning, Experiment 4 (like Experiment 3) required a departure from the standard CSPC design. Rather than having both an inducer and diagnostic set, each participant was exposed to two different inducer sets (referred to hereafter as Set 1 and Set 2) each comprising two items [animals]). Set 1 and Set 2 appeared in both locations but with opposing PCs (see Figure 2). For example, Set 1 (e.g., birds and cats) was MC in the upper location but MI in the lower location; in contrast, Set 2 (e.g., fish and dogs) was MI in the upper location but MC in the lower location (assignment of items to sets was counterbalanced across participants). If participants learn the conjunctive associations between location-item combinations and their PCs, then the key prediction for Experiment 4 is as follows: For Set 1 and Set 2, the congruency effect should be smaller when the set is MI compared to when the set is MC.

Note that in this design variant, because the items in each location (upper and lower) average to 50% congruent, the location-PC learning hypothesis predicts no difference in the congruency effect between locations (no standard CSPC effect). Similarly, because each item (birds, cats, fish, and dogs) averages to 50% congruent when collapsed across locations, the item-PC learning hypothesis predicts no difference in the congruency effect between items (between any cells). Consequently, the experiment is informative for our goal of evaluating the conjunctive learning hypothesis but in the presence of null effects, it could not adjudicate between the location-PC or item-PC learning hypotheses.

Method

Participants.—Sixty undergraduates (33 females, Age $M = 19.32$, $SD = 1.30$) from Washington University in St. Louis participated in the study to fulfill a credit as a partial requirement of a course. All participants were native English speakers with normal or corrected-to-normal vision.

Design and stimuli.—The stimuli were identical to those used in Experiments 1 and 2. Unlike the prior experiments, however, the design was a 2 (trial type: congruent vs. incongruent) × 2 (item set: Set 1 vs. Set 2) × 2 (item PC: MC vs. MI) within-subject design. Two items were allocated to each of two sets (e.g., Set 1 = birds and cats; Set 2 = fish and dogs). For half of the participants pictures of birds and cats were MC (PC-75) in the upper location and MI (PC-25) in the lower location, whereas pictures of fish and dogs were MI (PC-25) in the upper location and MC (PC-75) in the lower location (see Figure 2). For the remaining half, the reverse assignment was used. As shown in Table 8, each location (lower and upper) was 50% congruent collapsing across sets and each set of animals (bird/cat vs. dog/fish) was 50% congruent when collapsing across locations.

Procedure.—The procedure was identical to that of Experiment 1 save that participants completed 16 practice trials followed by four blocks of 192 trials (768 trials total). The number of trials was increased in part because we anticipated that participants may take more time to learn a conjunctive association. A brief break was given between the blocks.

Results

The RT trim eliminated < 1% of all trials. Mean RTs and error rates are presented in Table 9. A 2 (trial type: congruent vs. incongruent) × 2 (item set: set 1 vs. set 2) × 2 (item PC: MC vs. MI) within-subjects ANOVA was conducted on RT and error rate.

Reaction time.—There was a main effect of trial type, $F(1,59) = 372.33$, $p < .001$, $\eta_p^2 = .863$, revealing slower responses for incongruent than congruent trials. There was a significant Item Set PC × Trial Type interaction, $F(1,59) = 5.67$, $p = .021$, $\eta_p^2 = .09$, indicating that the congruency effect for Set 1 (M =103) was larger than that of Set 2 (M = 98). More critically, the three-way interaction between item set PC, item PC, and trial type was significant, $F(1,59) = 4.25$, $p = .04$, $\eta_p^2 = .07$ (see Figure 7). Therefore, we decomposed the interaction to examine the effects within each item set. For Set 1, the congruency effect was smaller when the items in the set were MI (M = 93) than when they were MC (M =

113), $F(1,59) = 7.58$, $p = .010$, $\eta_p^2 = .09$. However, for Set 2, there was no difference in the congruency effect when items in the set were MC ($M = 98$) or MI ($M = 98$), $F < 1$, $BF_{01} = 5.39$.

Error rate.—The significant main effect of trial type, $F(1,59) = 68.35$, $p < .001$, $\eta_p^2 = .54$, revealed that participants made more errors on incongruent than congruent trials. A significant Item Set × Trial Type interaction, $F(1,59) = 4.60$ $p = .04$, $\eta_p^2 = .07$, revealed that the congruency effect was larger in Set 1 ($M = .038$) than Set 2 ($M = .027$). The Item Set × Item PC interaction was also significant, $F(1,59) = 7.68$ $p = .01$, $\eta_p^2 = .12$. More importantly, the three-way interaction was significant, $F(1,59) = 5.48$, $p = .02$, $\eta_p^2 = .09$. This three-way interaction was decomposed. For Set 1, there was not a difference in the congruency effect when the items in the set were MC ($M = .040$) or MI ($M = .036$), $F < 1$, $BF_{01} = 4.49$. For Set 2, there was a difference but it was not in the predicted direction—the congruency effect was larger when the items in the set were MI ($M = .034$) than when they were MC ($M = .020$), $F(1,59) = 6.65$, $p = .01$, $\eta_p^2 = .10$.

Discussion

Experiment 4 directly tested the conjunctive learning hypothesis. The evidence in support of the conjunctive learning hypothesis was neither strong nor consistent. For RT, participants did demonstrate the predicted difference between the congruency effects of MC and MI items for inducer Set 1 but not for inducer Set 2, suggesting participants did not learn the full conjunction (location-item associations for both sets). For error rate, the predicted difference between MC and MI was also apparent for only one of the two item sets and, in the set where it was not observed, the difference was in the opposite direction than predicted by the hypothesis. All things considered the findings do not provide strong support for the conjunctive learning hypothesis.

General Discussion

The overarching and primary aim of the present study was to address the question of how to interpret failures to replicate or reproduce the theoretically important CSPC effect first observed by Crump and Milliken (2009, Experiment 2) in the combined inducer/diagnostic design that has recently yielded several such failures (Crump et al., 2017; Hutcheon & Spieler, 2017). Some researchers have argued that the combined design is plagued by a weaker CSPC manipulation, low trial counts, and underpowered statistical tests (Crump et al., 2017). Others have attributed these failures to the inconsistency with which location signals proportion congruence for all sets within a location for this design (i.e., the consistency hypothesis; Hutcheon & Spieler, 2017). Using sample sizes that approximately doubled those used by Hutcheon and Spieler (2017), we further tested the consistency hypothesis. More importantly, we tested two novel hypotheses—the item-PC learning hypothesis and the conjunctive learning hypothesis—aimed at tackling the key question of what it is that participants are learning based on experience with the inducer items in the combined inducer/diagnostic design. As noted in the introduction, the rationale behind adopting this approach was our view that to successfully anticipate when researchers will

or will not observe CSPC effects not just for inducer items but also for diagnostic items, we must understand what participants learn about the inducer items and the conditions under which such learning occurs. Indeed, it is the inducer items that are assumed to induce location-specific control that affects performance on the diagnostic items (Crump & Milliken, 2009). However, the failed replication/reproduction attempts indicate this may be a faulty assumption (i.e., participants are not learning location-PC associations), thereby begging the question of what is being learned about the inducer items. The two learning hypotheses we directly tested were inspired by our theorizing about additional learning opportunities in the CSPC paradigm besides location-PC learning, opportunities that, if exploited, could masquerade as failures to reproduce the CSPC effect. Table 10 summarizes our key findings.

In Experiment 1, we reported another failure to reproduce the CSPC effect for both sets of items in the combined inducer/diagnostic design, here using a picture-word Stroop task. In Experiment 2, using the same task, we demonstrated that a CSPC effect can be found for the inducer set when the diagnostic set is excluded from the design. The findings of Experiment 1 and 2 were therefore consistent with the consistency hypothesis (Hutcheon & Spieler, 2017). With regard to the key question of what participants do learn under conditions of location inconsistency, Experiment 1 provided preliminary support in favor of the item-PC learning hypothesis, which predicted no CSPC effect for either the inducer or diagnostic set in the standard design (i.e., a set of null effects). Collectively, Experiments 1 and 2 raised the theoretical possibility that item-PC learning may operate particularly in the face of location inconsistency.

Experiment 3 provided the first direct test of the item-PC learning hypothesis using a modified CSPC design where evidence for item-PC learning would not depend on a set of null effects. The key finding from Experiment 3 was a three-way interaction that was anticipated by the item-PC learning hypothesis and could not be explained by alternative accounts (e.g., location-PC learning; conjunctive learning). In short, the interaction pattern of a standard CSPC effect for the inducer set but a reversed CSPC effect for the diagnostic set is precisely what was expected if participants learned associations between the items (pictures) and PC, as illustrated in Figure 5. What this finding implies is that in studies that manipulate CSPC based on the location in which stimuli are presented, the primary association that participants learn based (primarily) on experiences with inducer items is not necessarily the location-PC association. Instead, Experiment 3 suggests participants instead may learn associations between items and PC. In the present study, items were represented by pictures of different animals. Accordingly, from the perspective of the episodic retrieval account, the evidence for item-PC learning suggests either that participants exclusively binned (organize and store) experiences during the task based on the item (animal picture) shown on each trial or they weighted this dimension (picture) more than location, for example (see Panel B of Figure 1 for an illustration). The implication of such learning for the diagnostic set in the standard combined inducer/diagnostic design (and context-specific control in designs in which diagnostic items share no overlap with the inducer items) is that a CSPC effect should *not* be observed. That is, generalization of location-specific control to diagnostic items cannot occur if there is item-PC learning and not location-PC learning (i.e., if there is no location-specific control in the first place) because location is the feature that is

shared between inducer and diagnostic items, and thus the feature on which generalization is dependent.

In contrast to the strong evidence for item-PC learning in Experiment 3 (i.e., robust, hypothesis-driven three-way interaction), Experiment 4 did not yield strong evidence for conjunctive learning, the idea that participants learn associations between location-item conjunctions and their PC. To our knowledge, Experiment 4 represented the first direct test of the role of this type of learning in the CSPC paradigm (but see Schmidt & Lemercier, 2019, for evidence of compound S-R contingency learning). To provide a strong test for conjunctive learning that had the potential to yield results that could not be explained by alternative accounts (e.g., CSPC effects for inducer items do not transfer to diagnostic items), we used a design in which stimuli appeared in two locations but location PC did not vary. The collective pattern of results was at best partially consistent with the conjunctive learning hypothesis. Thus, it may be unlikely that conjunctive learning has been masquerading as failures to reproduce the CSPC effect. This view seems plausible given that, of the five failures to reproduce the CSPC effect reported to date (including Experiment 1 in the present study), only one reported evidence that was suggestive of conjunctive learning (i.e., a CSPC effect for the inducer set but not the diagnostic set; see Table 1).

Implications of Observing Item-PC Learning in the CSPC Paradigm

An important contribution of the present study is the novel observation of item-PC learning in the CSPC paradigm in Experiment 3. Before addressing the implications of this finding, it is first important to address what is novel about this finding. There are several reports of item-specific proportion congruence effects in the literature, whereby congruency effects are larger for MC items than MI items including in paradigms where “item” was defined as the animal picture like in the present study (Bugg et al., 2011; Bugg & Dey, 2018). What is novel about the present observation is that item-PC learning was observed under conditions in which participants alternatively could have learned associations between location and PC, which if learned, would have led to a different pattern of results than that produced by item-PC learning. Learning of location-PC associations was not possible in prior studies (Bugg et al., 2011; Bugg & Dey, 2018). That is, unique to the present experiment, item-PC learning was placed in competition with location-PC learning and item-PC learning won out.

One major implication of this novel finding is that item-PC learning may preclude the observation of location based CSPC effects in the standard combined inducer/diagnostic CSPC paradigm, including for the inducer set that is intended to promote learning of location-PC associations and the diagnostic set that can assess context-specific control. In other words, failures to observe (reproduce) the CSPC effect in this paradigm are not necessarily a reflection of the instability of the effect; rather, such failures are theoretically anticipated by the item-PC learning hypothesis. That is, under conditions in which item-PC learning dominates, a CSPC effect necessarily will not be observed for either the inducer or diagnostic set in the standard paradigm (see Figure 1), which is the most common pattern (four out of five) reported in experiments that have failed to replicate or reproduce the CSPC effect to date using this paradigm (Experiment 1; Crump et al., 2017; Hutcheon & Spieler, 2017).

Theoretically, the dominant account of CSPC effects is the episodic retrieval account. A critical question prompted by the observation of item-PC learning in the CSPC paradigm is what factors lead participants to differentially weight item and location features within the episodic representations formed on each trial (or to bin experiences during a conflict task based on items [learn item-PC associations] as opposed to locations [learn location-PC associations], or vice versa). Given the present findings, understanding these factors seems essential for predicting whether future attempts to reproduce the CSPC effect will be successful. One potential factor is efficiency, or how many bins are needed to organize and store experiences during the task. Efficiency may explain why evidence for conjunctive learning was weak, given that it required eight bins to organize and store experiences (see Figure 1), making it the least efficient of the three forms of learning considered herein. However only two bins were needed for location-PC learning but in Experiment 3, item-PC learning, which required four bins, prevailed. This suggests there are other factors besides efficiency that dictate whether participants learn item-PC or location-PC associations.

The present evidence highlights two other factors that appear to be influential in determining the type of learning that occurs in the CSPC paradigm, although additional evidence is needed. One factor is location inconsistency, which was a feature of Experiments 1 and 3. The findings of Experiment 1 were accommodated by the item-PC learning hypothesis but not the location-PC learning hypothesis while the findings of Experiment 3 were uniquely explained by item-PC learning. In contrast, location inconsistency was not a feature of Experiment 2 because Experiment 2 included only inducer items. The strict form of the item-PC learning hypothesis (i.e., the hypothesis we began this investigation with that anticipates item-PC learning may occur in CSPC paradigms irrespective of consistency) anticipated no CSPC effect in Experiment 2 but a CSPC effect was found, as anticipated by the consistency hypothesis. Considering the collective pattern of results across Experiments 1 and 2, we suggest that item-PC learning may be particularly likely to occur in the face of location inconsistency (in contrast to the strict form of the item-PC learning hypothesis; cf. Hutcheon & Spieler, 2017). Hutcheon and Spieler (2017) stated: “It is the breaking of consistency through the inclusion of a transfer [diagnostic] set that appears to prevent the implementation of context-driven control” (p. 1303). Our findings gel with this assertion and expand upon the consistency hypothesis by specifying a potential mechanism via which location inconsistency may prevent the emergence of CSPC effects for inducer items, and consequently location-specific control (i.e., generalization to diagnostic items in the design of Crump & Milliken, 2009). Specifically, location-specific control may be prevented by the exploitation of item-PC associations.[12] In this sense, the two hypotheses may be seen as complementary in that the consistency hypothesis specifies a boundary condition for observing CSPC effects and the item-PC learning hypothesis specifies why location inconsistency results in a boundary condition. In other words, in the face of location inconsistency, participants appear to bin experiences based on the item and not

[12]One finding that does not align perfectly with this view is that of Experiment 4. Location inconsistency was a feature of Experiment 4, but the results were not fully consistent with item-PC learning because for both RT and error rate, a difference was found between MC and MI inducer items for one of the sets. However, as noted in the introduction to Experiment 4, the design was devised to directly test only the conjunctive learning hypothesis and not the item-PC learning hypothesis because both it and the location-PC learning hypothesis predicted null results.

the feature (location) that is critical for observing CSPC effects (including their observation on diagnostic trials).

The second determinant may be the relative salience of items compared to locations in the CSPC paradigm. In the present task, the goal was to name the animal in the picture while ignoring the word. Thus, to perform the task, participants had to attend to the animal in the picture. In contrast, location was nominally irrelevant to the task goal. Although attention had to be directed to the stimulus in order to name it, the task goal was not contingent on location. Whether a cat was in the upper or lower location, the correct response was "cat". A salience account of the Experiment 3 findings, which is admittedly post-hoc, posits that item-PC learning prevailed over location-PC learning because items were more salient than locations during performance of the task (see Bugg & Dey, 2018, for discussion of various task features that can affect salience). Conversely, it is possible that in the original study of Crump and Milliken (2009, Experiment 2), some component of their methodology may have made location more salient than items. Along these lines, it is interesting that the one partial success/failure to reproduce the findings of that experiment came from a letter flanker task requiring participants to select the location of the target dimension and ignore the flanking locations (Crump et al., 2017, Experiment 4). Possibly, such a task primes location to be more salient, thereby encouraging participants to organize and store episodic representations into location-based bins. In line with this possibility, it has been shown that transfer of CSPC effects to novel locations on screen occurs only in tasks that entail location-based conflict (Pickel, Pratt, & Weidler, 2019; Weidler & Bugg, 2016). Taking a broader perspective, looking back on failed attempts to produce a CSPC effect in designs that included only an inducer set using contexts other than location (e.g., shape-based without a counting manipulation; Crump et al., 2006; Crump et al., 2008; color-based without additional blocked training; Lehle & Hübner, 2008), the present findings raise the possibility that in these cases item-PC learning dominated. Future research might combine the location-counting manipulation from Experiment 2 with the design of Experiment 3 to see if one can shift salience to favor locations over items, thereby promoting location-PC learning in that design.

One puzzling observation for the consistency hypothesis (Hutcheon & Spieler, 2017) and salience-based accounts is that a CSPC effect has reliably been found for diagnostic items in the design variant in which one cannot calculate a CSPC effect for the inducer set (i.e., version where the inducer set is 100% or 0% congruent across locations; Crump & Milliken, 2009, Experiment 1; see Table 1). This design also involves location inconsistency (i.e., the inducer and diagnostic sets have different PCs) and a similar task goal to the present study (for which location is nominally irrelevant). Of the three learning hypotheses, only the location-PC learning hypothesis accommodates this pattern (because neither of the others predicts a CSPC effect for the diagnostic set), but it is not immediately obvious why 100%/0% inducer items would yield replicable CSPC effects but 92%/8% inducer items would not. Although the CSPC manipulation in this variant is stronger (Crump et al., 2017), the difference in location PC across the two designs amounts to a small difference of 75/25% vs. 71/29% congruent, which seems an unlikely source of the striking difference in the CSPC effects' stability across designs (cf. Blais, Harris, Guerrero, & Bunge, 2012). While

outside the scope of the present study, solving this puzzle is an important goal for future research.

Practically speaking, a question prompted by the present findings is what design can be used to reliably observe CSPC effects in conflict tasks based on learning of location-PC associations; in other words, what design might discourage learning of item-PC associations? A secondary aim of the present study was to determine whether the partially overlapping sets design may be fruitful in this respect, but the present findings demonstrating item-PC learning in the context of this design suggest it is not. Some researchers may be satisfied with the design where the inducer sets are 100% or 0% congruent (Crump & Milliken, 2009; Crump et al., 2017). However, others may find it dissatisfying not to be able to calculate a CSPC effect for the inducer set given it is this set that is responsible for promoting learning of location-PC associations. Future research should investigate two alternatives. One alternative is to use a CSPC paradigm that presents unique stimuli that are not easily categorized on every trial such that each item is presented only once in either a congruent or an incongruent format (see face-based flanker findings of King, Korb, & Egner, 2012; cf. Spinelli, Perry, & Lupker, 2019). In this case, there is no value in binning experiences with items as there are no summary statistics (item-based PCs) to be gained through binning (not to mention this may be inefficient assuming 100s of trials and consequently 100s of unique item bins in such an experiment). Along these lines, although we have thus far restricted our discussion to traditional conflict tasks, a finding that appears to be difficult to reconcile with the item-PC learning hypothesis comes from a location based CSPC manipulation in a dual-task paradigm. CSPC effects were found in a transfer phase that comprised number stimuli judged as odd or even following a training phase in which CSPC was manipulated for letter stimuli judged as vowels or consonants (Surrey, Dreisbach, & Fischer, 2017). Interestingly, this task involved 16 unique items (Stimulus-Stimulus combinations) in each phase; this may support the above suggestion that CSPC paradigms with entirely unique or relatively large stimulus sets may be more likely to encourage location-PC learning (discourage item-PC learning). A second alternative is to implement the counting manipulation in the combined inducer/diagnostic design for which the CSPC effect has not been reproducible (see Table 1; see also present Experiment 1) to bias attention toward the location of the stimulus (Crump et al., 2008 see also present Experiment 2). This could shift the weighting of dimensions and correspondingly, the binning process, toward locations and away from items.

Limitations and Future Directions

One limitation of the present study is that we investigated CSPC effects in one conflict paradigm. This is a standard approach in the literature (see also Crump & Milliken, 2009; Hutcheon & Spieler, 2017; but see Crump et al., 2017). However, this raises the question of whether our findings are specific to the picture-word Stroop variant used herein. We think this is unlikely for two reasons. First, in Experiment 1, using this variant, we produced the same pattern observed previously in the color-word Stroop task, the task most often used in the relevant CSPC literature. Second, a CSPC effect was observed in Experiment 2 demonstrating sensitivity of performance to the location PC manipulation, again replicating a key pattern from the color-word Stroop task (Crump et al., 2006). Third, item-specific PC

effects, which are indicative of item-PC learning outside of the CSPC paradigm, have also been observed in color-word Stroop (e.g., Bugg & Hutchison, 2013) and in flanker (Bugg, 2015) tasks, and in designs where the sets are partially crossed (Bugg et al., 2011; Bugg & Dey, 2018) and when they are separate (Jacoby et al., 2003; Spinelli & Lupker, 2019). Thus, there is not a strong reason to believe the item-PC learning hypothesis or the novel evidence supporting this hypothesis is specific to the picture-word Stroop task or design we used, or that item-PC learning may compete with location-PC learning selectively in this variant. Nonetheless, converging evidence from different conflict tasks and different designs in future studies would strengthen the present conclusions.

Future research also should further test the conjunctive learning hypothesis. The strong test provided in Experiment 4 required participants to learn that the two different sets of items switched PCs across locations. In contrast, in the standard CSPC paradigm, participants must learn that one item set switches PC across locations (i.e., the inducer items) while the other item set maintains its PC across locations (i.e., the diagnostic items). This type of conjunction may be easier to learn. Although not a direct test of this possibility, the results of Experiment 2 could be interpreted as evidence participants learned that birds, cats, fish, and dogs were MC in the upper location but MI in the lower location (i.e., that the inducer items switched PC across locations), a simpler conjunction. While this pattern suggests participants may be able to learn location-item PC associations for one (inducer) set, without a second (diagnostic) set, it is not possible to disentangle the contributions of conjunctive learning and location-PC learning to this pattern. Nonetheless, it reinforces the possibility that stronger evidence for conjunctive learning of location-item PC associations could be found in such cases (e.g., as another example, in the case that fewer potential bins are needed to represent all location-item pairings; or, if the design was used where the sets do not cross).

An important message that is reinforced by the present findings is that no PC paradigm is learning pure. Schmidt and his colleagues previously have emphasized that stimulus-response (contingency) learning processes can contribute to PC paradigms (Schmidt & Besner, 2008), which stimulated theorists to devise a design that could reveal context-specific control independent of S-R learning (i.e., the combined inducer/diagnostic design of Crump & Milliken, 2009). And more recently Schmidt and Lemercier (2019) demonstrated the role of complex (compound-cue) contingency learning processes in a CSPC paradigm. The present study extends this message by demonstrating that PC paradigms are also not learning pure with respect to the various associations *with PC* that can be learned. In addition to the customary location-PC association, we demonstrated that the CSPC paradigm also enables learning of item-PC associations. While one type of learning may dominate in a given experiment (e.g., location-PC learning in Experiment 2; item-PC learning in Experiment 3), it is likely the case that the results of any experiment in which CSPC is manipulated reflect a mixture of different learners (possibly including conjunctive learners although we did not find strong evidence for group level conjunctive learning in the present study). Consequently, another important goal for future studies is to develop and test alternative analytic methods that could enable researchers to characterize individual participants as location-PC vs. item-PC learners. Such methods may be able to predict based

on inducer set performance whether a CSPC effect will be observed for the diagnostic set (cf. McDaniel, Dimperio, Griego, & Busemeyer, 2009).

Conclusion

Crump and Milliken (2009) observed a CSPC effect for both an inducer set and diagnostic set, a pattern that has represented the strongest evidence to date for context-specific control but proven difficult to replicate or reproduce. We proposed that the potential for neither CSPC effect to be observed (for the inducer or diagnostic set; as was the case for the three previous replication and reproduction failures and the present Experiment 1) or for a CSPC effect to be found only for the inducer set (as has been observed in one case; Hutcheon & Spieler, 2017, Experiment 3) is high if participants learn associations between items and PC or between location-item conjunctions and PC, as opposed to the customary location-PC association. While the current findings suggest conjunctive learning may be unlikely (Experiment 4), the findings demonstrated that item-PC learning is a viable alternative to location-PC learning in the CSPC paradigm (Experiment 3). Thus, a key take home message is that failures to reproduce the CSPC effect do not necessarily indicate a Type 1 error or instability. Instead, consistent with the item-PC learning hypothesis, such failures may indicate that episodic representations of each trial were stored and organized (i.e., binned) according to the item rather than the location, as illustrated in Panel B of Figure 1. When inducer items induce this particular type of learning, the pattern that has been observed in all but one prior attempt to reproduce the CSPC effect for inducer and diagnostic sets (i.e., neither a CSPC effect for the inducer set or the diagnostic set used to infer context-specific control; Crump & Milliken, 2009) is precisely the expected pattern (i.e., not surprising). Predicting whether future attempts at reproducibility are successful will require an understanding of what factors promote learning of item-PC associations versus learning of location-PC associations.

Acknowledgments

We thank Isaac Bindert, Eva Jeliazkova, and Eric Lu for assistance with data collection, and the Cognitive Control and Aging Lab for their helpful comments on this manuscript.

References

Abrahamse E, Braem S, Notebaert W, & Verguts T (2016). Grounding cognitive control in associative learning. Psychological Bulletin, 142, 693 – 728. [PubMed: 27148628]

Blais C, Harris MB, Guerrero JV, & Bunge SA (2012). Rethinking the role of automaticity in cognitive control. Quarterly Journal of Experimental Psychology, 65, 268 – 276.

Braem S, Bugg JM, Schmidt JR, Crump MJC, Weissman DH, Notebaert W, & Egner T (2019). Measuring adaptive control in conflict tasks. Trends in Cognitive Sciences, 23, 769–783. [PubMed: 31331794]

Braver TS, Gray JR, & Burgess GC (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In Conway ARA, Jarrold C, Kane MJ, Miyake A, & Towse JN (Eds.), Variation in working memory (pp. 76–106). Oxford University Press.

Bugg JM (2012) Dissociating levels of cognitive control: The case of Stroop interference. Current Directions in Psychological Science, 21, 302 – 309.

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

Bugg JM (2017). Context, conflict, and control. In Egner T (Ed.) The Wiley handbook of cognitive control. (pp. 79 – 96). Chichester, West Sussex, UK: John Wiley & Sons.

Bugg JM, & Chanani S (2011). List-wide control is not entirely elusive: Evidence from picture-word Stroop. Psychonomic Bulletin & Review, 18, 930 – 936. [PubMed: 21638107]

Bugg JM, & Crump MJC (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. Frontiers in Psychology: Cognition, 3, 1 – 16.

Bugg JM, & Dey A (2018). When stimulus-driven control settings compete: On the dominance of categories as cues for control. Journal of Experimental Psychology: Human Perception and Performance, 44, 1905 – 1932. [PubMed: 30211592]

Bugg JM, & Hutchison KA (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. [PubMed: 22845037]

Bugg JM, Jacoby LL, & 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. [PubMed: 20718569]

Chiu YC, Jiang J, & Egner T (2017). The caudate nucleus mediates learning of stimulus- control state associations. Journal of Neuroscience, 37, 1028 – 1038. [PubMed: 28123033]

Corballis PM, & Gratton G (2003). Independent control of processing strategies for different locations in the visual field. Biological Psychology, 64, 191 – 209. [PubMed: 14602362]

Crump MJC (2016). Learning to selectively attend from context-specific attentional histories: A demonstration and some constraints. Canadian Journal of Experimental Psychology, 70, 59 – 77. [PubMed: 26372055]

Crump MJC, Brosowsky NP, & Milliken B (2017). Reproducing the location-based context-specific proportion congruent effect for frequency unbiased items: A reply to Hutcheon and Spieler (2016). The Quarterly Journal of Experimental Psychology, 70, 1792 – 1807. [PubMed: 27340758]

Crump MJC, Gong Z, & Milliken B (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. Psychonomic Bulletin & Review, 13, 316 – 321. [PubMed: 16893001]

Crump MJC, & Milliken B (2009). The flexibility of context-specific control: Evidence for context-driven generalization of item-specific control. The Quarterly Journal of Experimental Psychology, 62, 1523 – 1532. [PubMed: 19370487]

Crump MJC, Vaquero JMM, & Milliken B (2008). Context- specific learning and control: The roles of awareness, task relevance, and relative salience. Consciousness and Cognition, 17, 22–36. doi: 10.1016/j.concog.2007.01.004 [PubMed: 17349805]

Diede NT, & Bugg JM (2017). Cognitive effort is modulated outside of the explicit awareness of conflict frequency: Evidence from pupillometry. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43, 824 – 835.

Diede NT, & Bugg JM (2019). Boundary conditions for the influence of spatial proximity on context-specific attentional settings . Attention, Perception, &Psychophysics, 81, 1386 – 1404.

Egner T (2014). Creatures of habit (and control): A multi-level learning perspective on the modulation of congruency effects. Frontiers in Psychology, 5: 10.3389/fpsyg.2014.01108

Gonthier C, Braver TS, & Bugg JM (2016). Doubly dissociating proactive and reactive control in the Stroop task. Memory & Cognition, 44, 778 – 788. [PubMed: 26861210]

Gottschalk C, & Fischer R (2017). Activation of context-specific attentional control sets by exogenous allocation of visual attention to the context? Psychological Research, 81, 378 – 391. [PubMed: 26850627]

Hutcheon TG, & Spieler DH (2017). Limits on the generalizability of context-driven control. The Quarterly Journal of Experimental Psychology, 70, 1292 – 1304. [PubMed: 27109465]

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

Jacoby LL, Lindsay DS, & Hessels S (2003). Item-specific control of automatic processes: Stroop process dissociations. Psychonomic Bulletin & Review, 10(3), 638–644. [PubMed: 14620358]

King JA, Korb FM, & Egner T (2012). Priming of control: Implicit contextual cueing of top-down attentional set. The Journal of Neuroscience, 32, 8192 – 8200. [PubMed: 22699900]

Lehle C, & Hübner R (2008). On-the-fly adaptation of selectivity in the flanker task. Psychonomic Bulletin & Review, 15(4), 814–818. [PubMed: 18792509]

Logan GD (1988). Toward an instance theory of automatization. Psychological Review, 95, 492 – 527.

McDaniel MA, Dimperio E, Griego JA, & Busemeyer JR (2009). Predicting transfer performance: A comparison of competing function learning models. Journal of Experimental Psychology: Learning, Memory, & Cognition, 35, 173 – 195.

Pickel L, Pratt J, & Weidler BJ (2019). The transfer of location-based control requires location-based conflict. Attention, Perception, & Psychophysics, 81, 2788 – 2797.

Schmidt JR, & Besner D (2008). The Stroop effect: Why proportion congruence has nothing to do with congruency and everything to do with contingency. Journal of Experimental Psychology: Learning, Memory, & Cognition, 34, 514–523.

Schmidt JR, & Lemercier C (2019). Context-specific proportion congruence effects: Compound cue contingency learning in disguise. Quarterly Journal of Experimental Psychology, 72, 1119 – 1130.

Spinelli G, & Lupker SJ (2019). Item-specific control of attention in the Stroop task: Contingency learning is not the whole story in the item-specific proportion congruent effect. Memory & Cognition, Advanced Online Publication, 10.3758/s13421-019-00980-y

Spinelli G, Perry JR, & Lupker SJ (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, 44, 995 – 1014.

Stroop JR (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643 – 662.

Surrey C, Dreisbach G, & Fischer R (2016). Context-specific adjustment of cognitive control: Transfer of adaptive control sets. The Quarterly Journal of Experimental Psychology, 70, 2386 – 2401. [PubMed: 27696936]

Wagenmakers EJ, Wetzels R, Borsboom D, & Van Der Maas HL (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). Journal of Personality and Social Psychology, 100(3), 426– 432. [PubMed: 21280965]

Weidler BJ, & Bugg JM (2016). Transfer of location-specific control to untrained locations. Quarterly Journal of Experimental Psychology, 69, 2202 – 2217.

Public Significance Statement:

Prior research found that susceptibility to distraction is controlled by contextual cues such as the location in which distraction occurs. Locations with a more frequent history of distraction enable better attentional performance both for stimuli that contributed to that history and for other stimuli. However, later research suggested this finding may not be reproducible. The present findings present an alternative view by demonstrating that other cues are sometimes preferentially exploited instead of location to guide attention, and when this occurs, the above finding should not in fact be reproduced. At a general level, our findings suggest humans' attentional performance in the same situation can be guided by a variety of cues, reflecting the variety of opportunities a situation presents for learning.

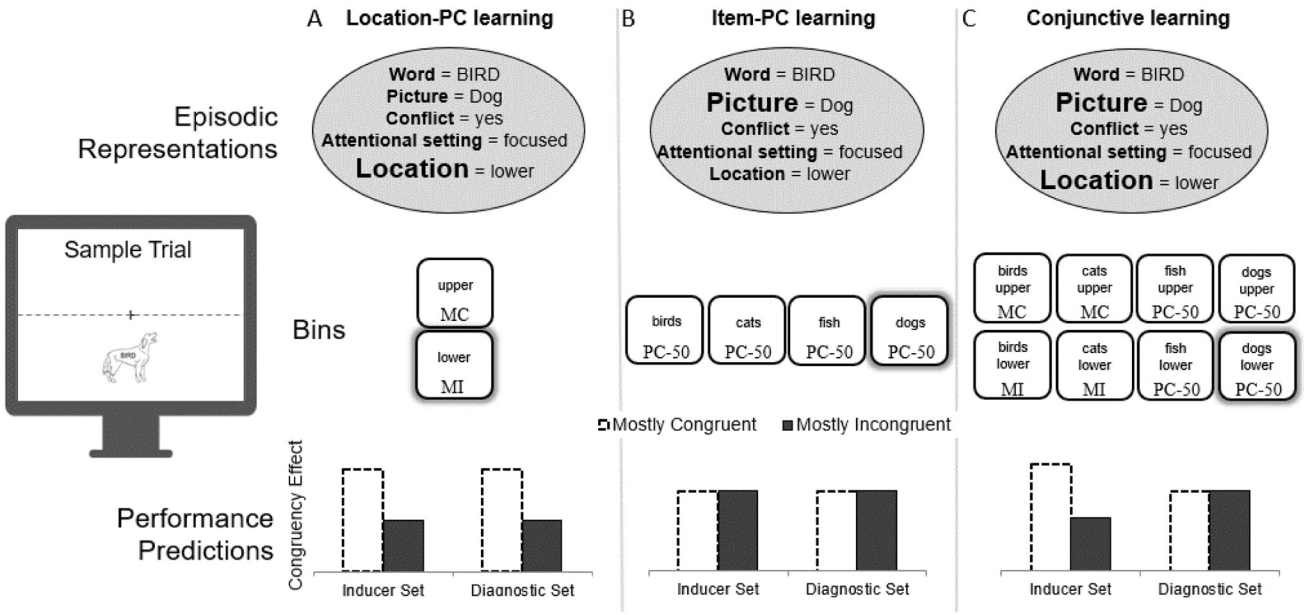

Figure 1.

Three opportunities for learning in the CSPC paradigm. The figure is based on one counterbalance from a picture-word Stroop task in which the upper location is MC and the lower location is MI, and birds and cats play the role of the inducer set while fish and dogs play the role of the diagnostic set. This task is represented in the figure because it was used in the present experiments; however, the episodic representations, bins, and predictions can just as easily be generalized to other tasks such as color-word Stroop. On the left, a sample trial is shown (not to scale). The gray ovals in the upper portion of the figure depict hypothesized differences in the nature of the episodic representation that would be formed and stored on this trial depending on whether learning is based on location (Panel A) , item (Panel B), or a combination of location and item (Panel C). The font size within each episodic representation illustrates that dimensions would be differentially weighted depending on the type of learning at play (bigger font size = heavier weighting). For example, in Panel A, the font size associated with location is larger because Panel A represents location-PC learning. The middle portion of the figure depicts the bins that would be used assuming a given type of learning occurs, with the highlighted (glowing) bin in each panel indicating in which bin the episodic representation corresponding to the sample trial would be stored. Importantly, the PC indicated within each bin (MC, MI, or PC-50) corresponds to the *overall PC of the bin* collapsed across all episodic representations that would be stored in a bin during the task. For example, in Panel B, the bin corresponding to each item is PC-50 because during the task each animal is presented with a congruent distracting word on half of the trials and an incongruent distractor word on half of the trials. The lower portion of the figure depicts *idealized* performance predictions corresponding to the location-PC learning hypothesis, item-PC learning hypothesis, and conjunctive learning hypothesis, respectively, for the standard CSPC paradigm in which inducer and diagnostic sets are combined and CSPC effects can be calculated for each set. The legend in the lower portion refers to location PC. Note that the consistency hypothesis (Hutcheon & Spieler,

2017) assumes that under conditions of location consistency but not location inconsistency, location-PC learning (Panel A) will occur.

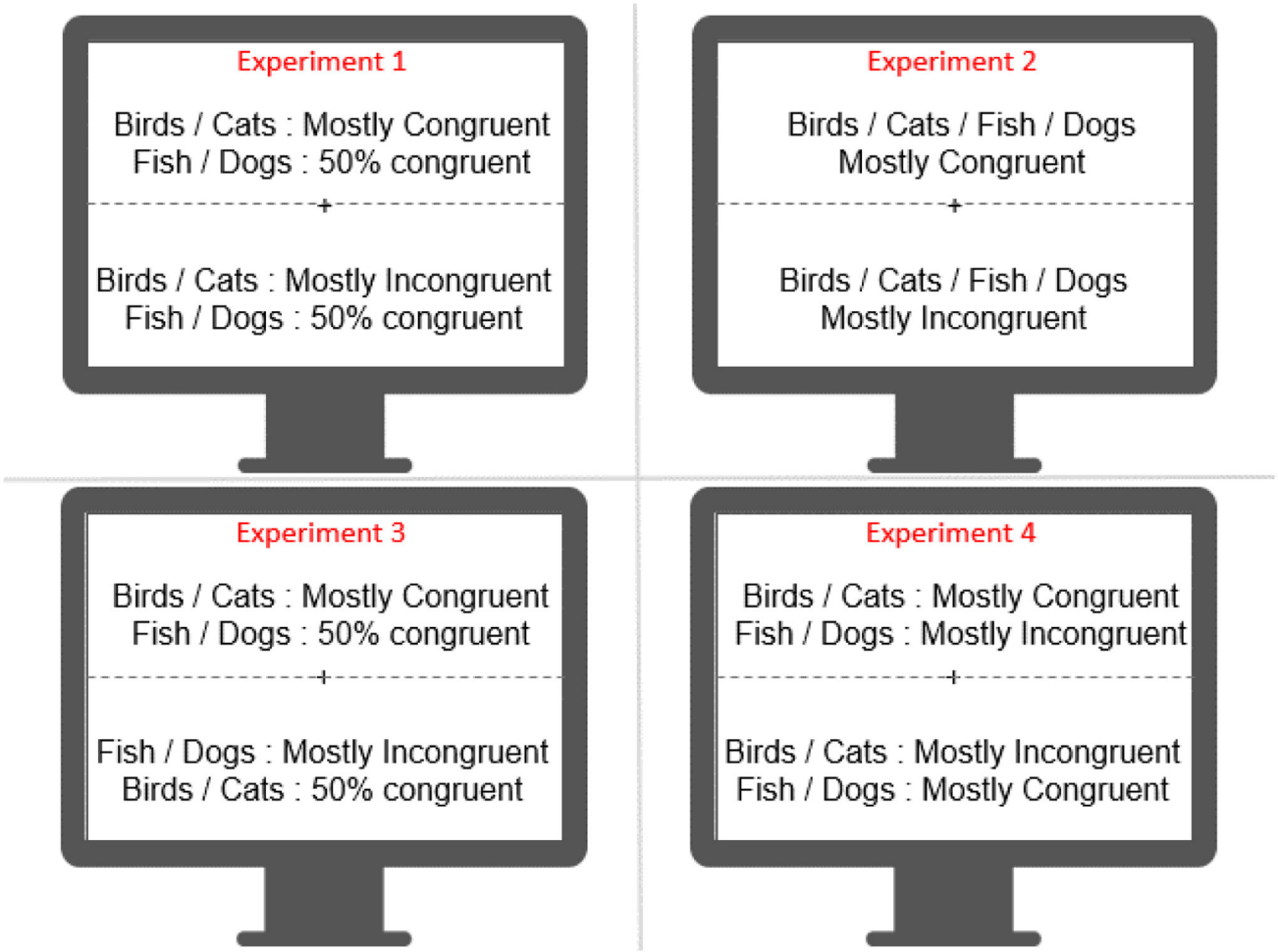

Figure 2.

Overview of the designs of Experiments 1 through 4, namely the assignment of items (animal pictures) to locations and to the roles of inducer and diagnostic sets. Experiment 2 included only an inducer set. Note that only one of the possible counterbalances in each experiment is represented in the figure.

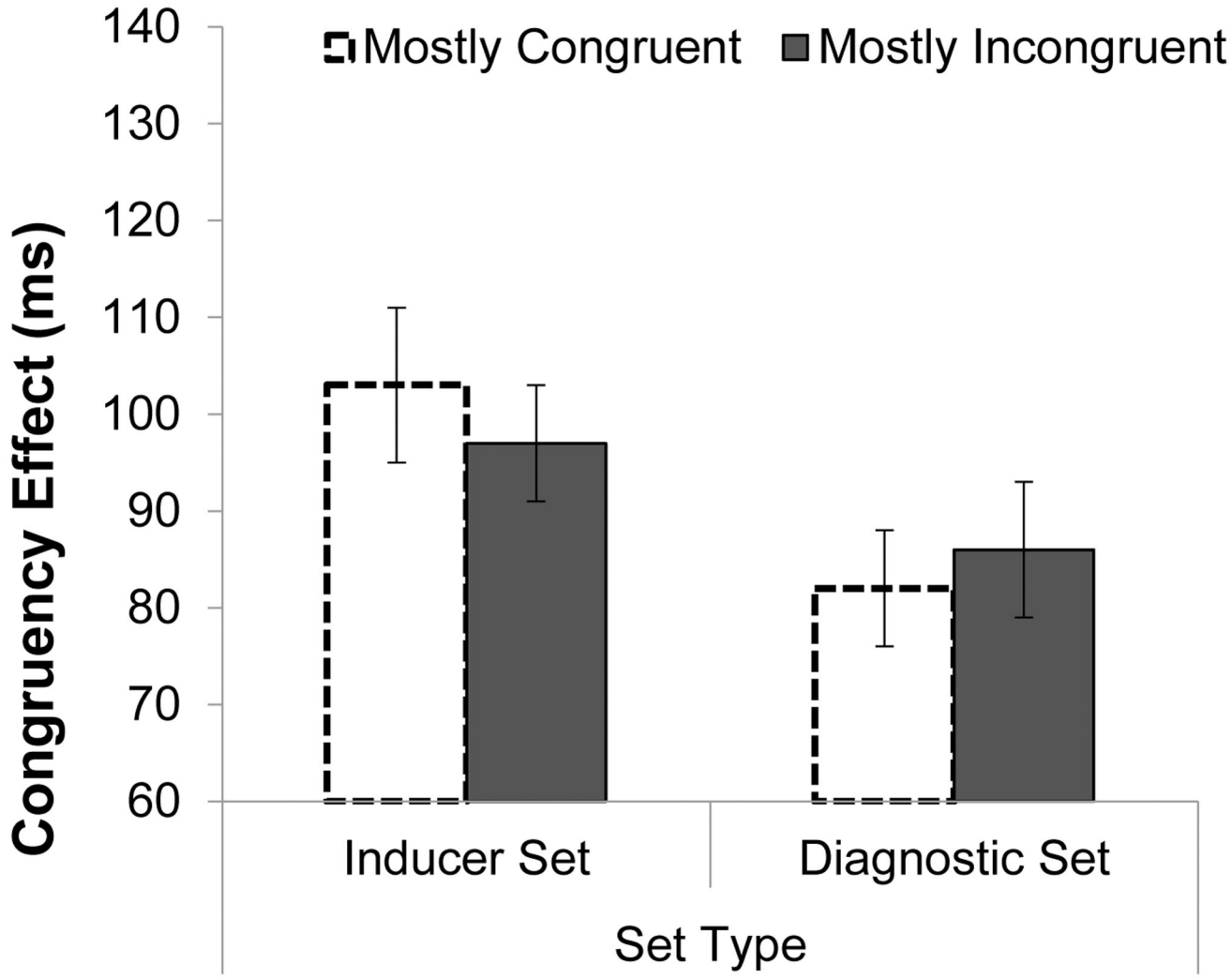

Figure 3.

Mean congruency (Stroop) effects as a function of location-specific proportion congruence for the inducer set and diagnostic set in Experiment 1. Note that items in the diagnostic set were 50% congruent in each location.

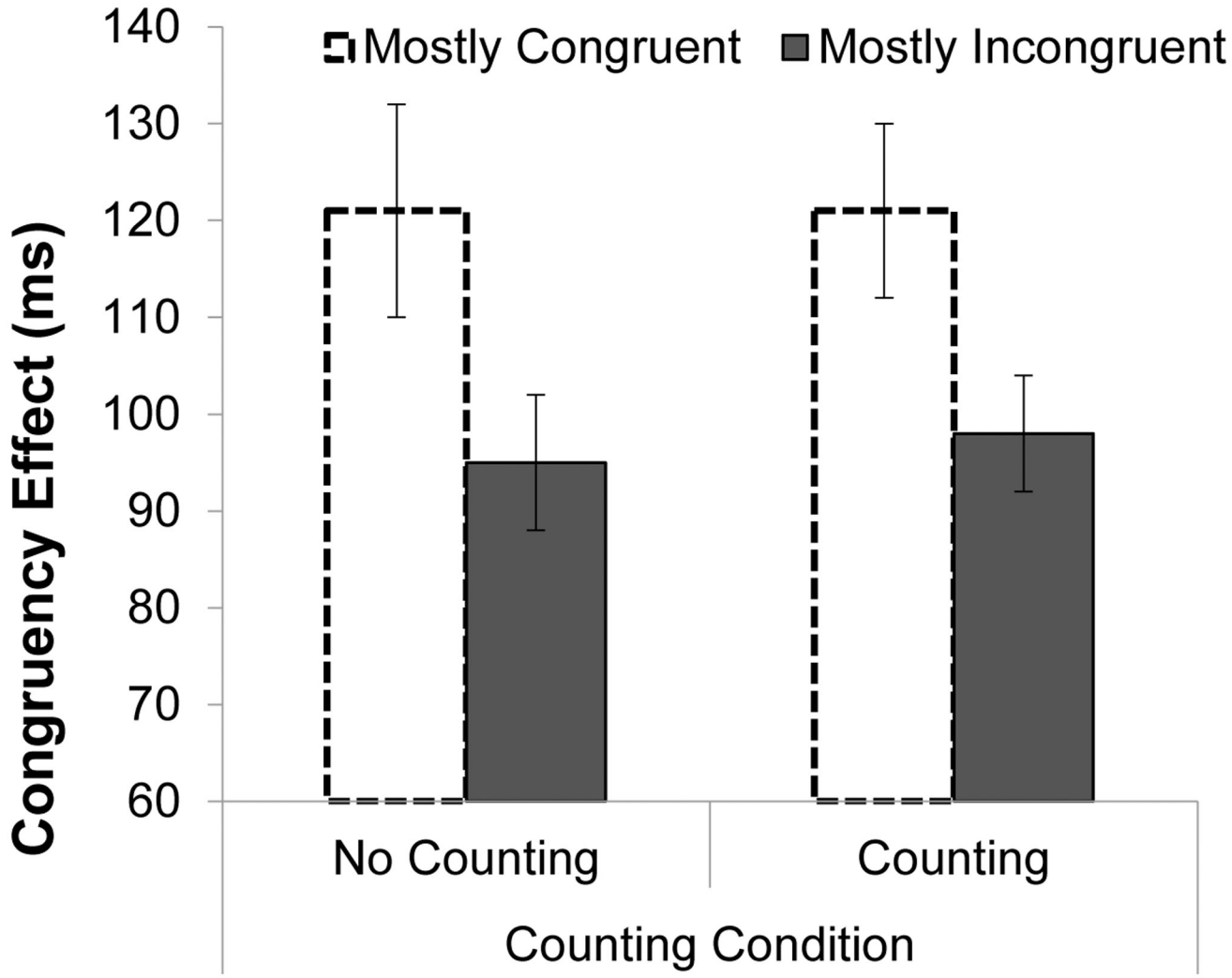

Figure 4.

Mean congruency (Stroop) effects as a function of location-specific proportion congruence for the inducer set for participants that did not count (left) and those that did count (right) the number of stimuli in one location in Experiment 2. Note that a diagnostic set was not included in this experiment.

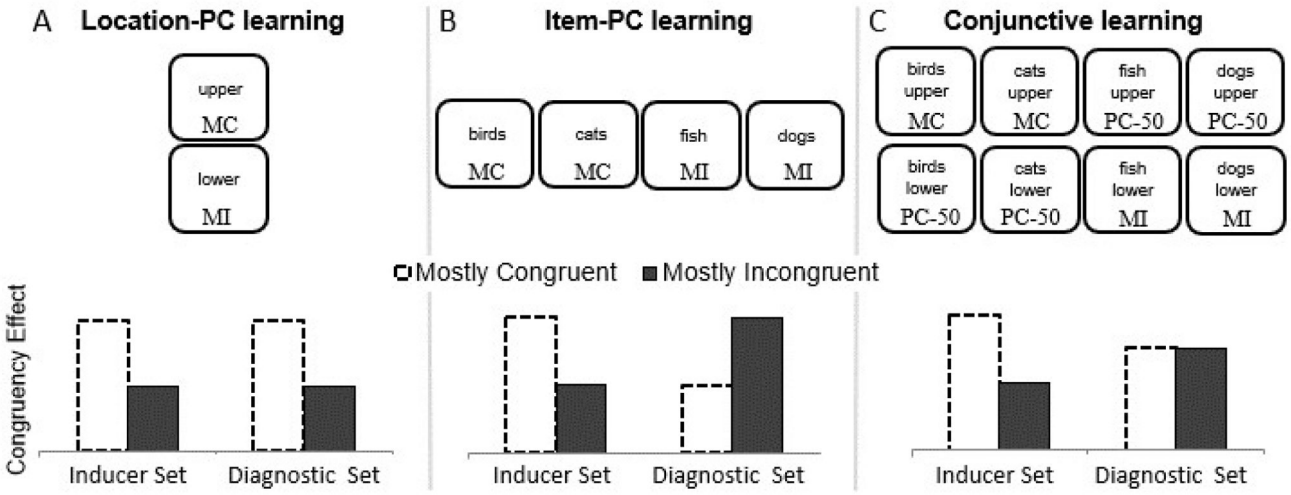

Figure 5.

Hypothesized bins and performance patterns for Experiment 3 for each of the three learning hypotheses (Panels A, B, and C). The figure is based on one counterbalance in which the upper location is MC and the lower location is MI, birds and cats are inducer items in the upper MC location and fish and dogs are inducer items in the lower MI location, and birds and cats are PC-50 diagnostic items in the lower MI location and fish and dogs are PC-50 diagnostic items in the upper MC location. The upper portion of the figure depicts the bins that would be used assuming a given type of learning occurs in Experiment 3. Importantly, the PC indicated within each bin (MC, MI, or PC-50) corresponds to the *overall PC of the bin* collapsed across all episodic representations that would be stored in a bin during the task. For example, in Panel B, the bins for birds and cats are MC because collapsing across all trials on which birds or cats, respectively, are presented as inducer or diagnostic items (across locations), birds and cats are 70% congruent. The lower portion of the figure depicts *idealized* performance predictions corresponding to the location-PC learning hypothesis, item-PC learning hypothesis, and conjunctive learning hypothesis, respectively. The legend in the lower portion refers to location PC.

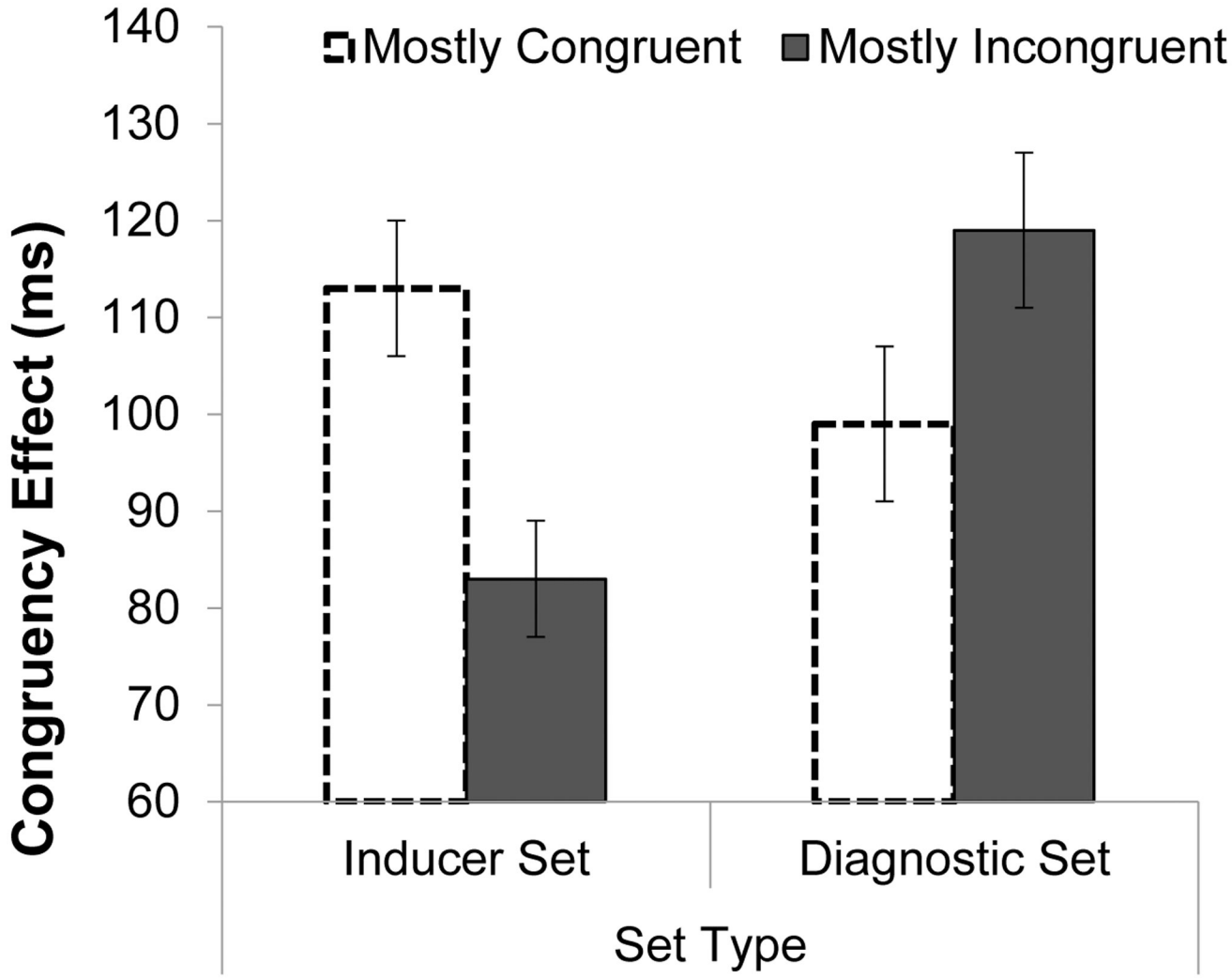

Figure 6.

Mean congruency (Stroop) effects as a function of location-specific proportion congruence for the inducer and diagnostic sets in Experiment 3. Note that the same items (e.g., birds and cats) played the role of the mostly congruent inducer items (leftmost bar) and the role of the 50% congruent diagnostic items in the mostly incongruent location (rightmost bar). Conversely, a different set of items (e.g., fish and dogs) played the role of the mostly incongruent inducer items (left middle bar) and the role of the 50% congruent diagnostic items in the mostly congruent location (right middle bar).

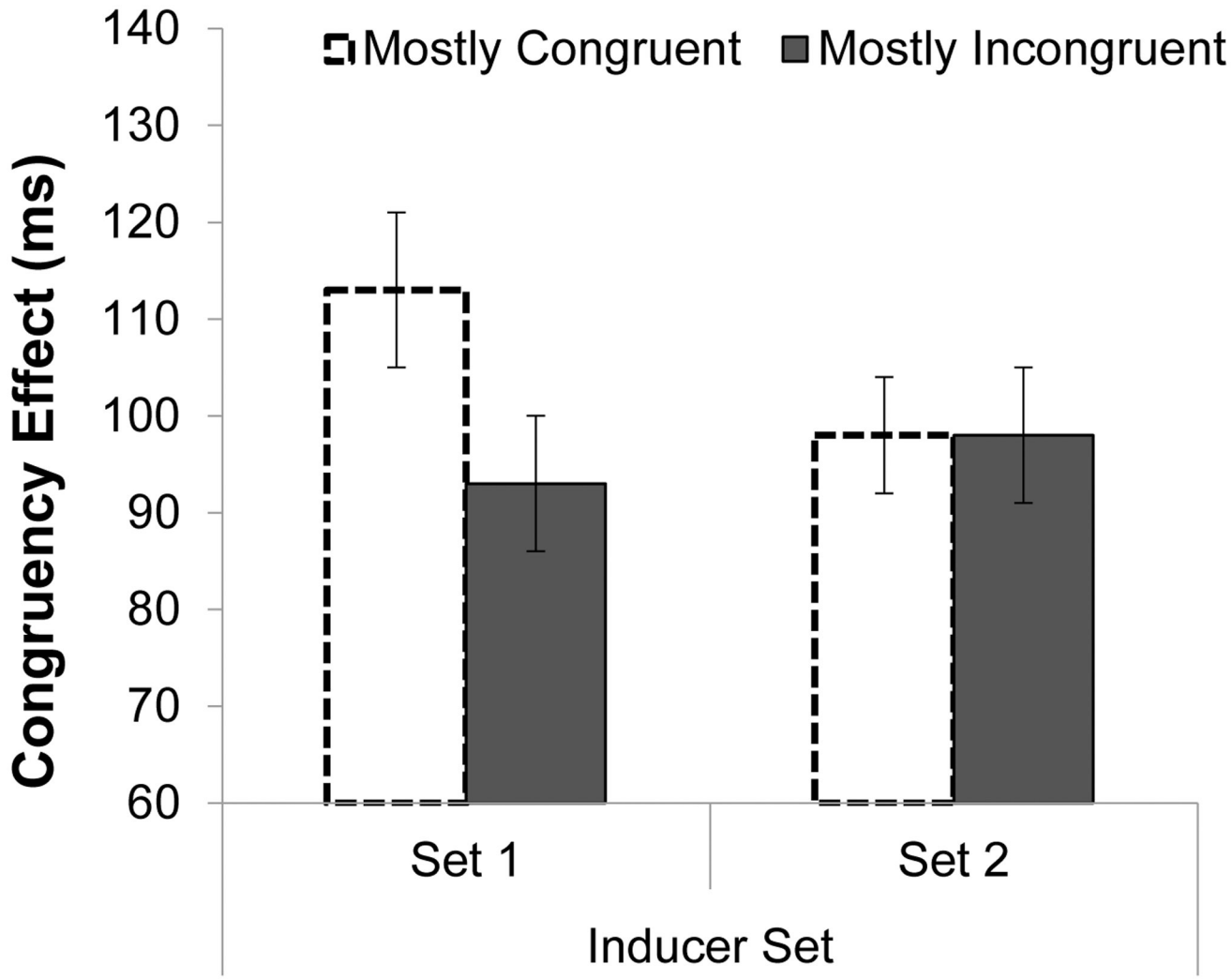

Figure 7.

Mean congruency effects as a function of inducer set (Set 1 vs. Set 2) in Experiment 4 and Item PC (mostly congruent or mostly incongruent). For example, Set 1 of the inducer items (e.g., birds and cats) was mostly congruent in the upper location but mostly incongruent in the lower location, while Set 2 of the inducer items (e.g., fish and dogs) was mostly incongruent in the upper location and mostly congruent in the lower location.

Table 1

Summary of Findings from Past Experiments using the Combined Inducer/Diagnostic CSPC Design in Conflict Tasks with Location as the Contextual Cue and the Correspondence of these Findings with the Consistency Hypothesis and the Three Learning Hypotheses

					Learning Hypothesis		
Study	**Exp**	**Inducer CSPC?**	**Diagnostic CSPC?**	**Consistency Hypothesis**	**Location-PC**	**Item-PC**	**Conjunctive**
Crump & Milliken (2009)	1	--	yes		■		
	2	yes	yes		■		
Hutcheon & Spieler (2017)	1	no	no	■		■	
	2	no	no	■		■	
	3*	yes	no	■			■
Crump, Brosowsky, & Milliken (2017)	1	--	yes		■		
	2	no	no	■		■	
	3	--	yes		■		
	4	no	yes				

Note: Exp = experiment; CSPC = context-specific proportion congruence effect; -- = could not be calculated because the inducer set was 100% congruent or 100% incongruent; yes = effect was observed; no = effect was not observed; black shading = results are consistent with the hypothesis; gray = results are inconsistent;

* = in Experiment 3, a different design was used compared to the other experiments in this table.

The first four blocks comprised only the inducer set and a CSPC effect was found; a diagnostic set was added in later blocks and no CSPC effect was found (a CSPC effect could not be calculated for the inducer set in the later blocks because the inducer set was then 100% congruent or 100% incongruent). For the Consistency Hypothesis column, the relevant finding is the presence/absence of a CSPC effect for just the inducer set or both the inducer and diagnostic sets, where applicable.

Table 2

Stimulus Frequencies in Experiment 1

Location PC	Set Type	Picture	Word			
			BIRD	CAT	FISH	DOG
MC	Inducer	Bird	72	8	8	8
		Cat	8	72	8	8
	Diagnostic	Fish	4	4	12	4
		Dog	4	4	4	12
MI	Inducer	Bird	24	24	24	24
		Cat	24	24	24	24
	Diagnostic	Fish	4	4	12	4
		Dog	4	4	4	12

Note: Numbers represent frequencies of stimulus presentation in Experiment 1 collapsed across all four pictures of birds, cats, fish, and dogs, which were presented equally frequently (e.g., for the cell representing a MC Inducer bird with the word BIRD, there were 18 presentations of each of the four bird pictures). Underlined frequencies represent congruent trials. In this example, birds and cats represent the inducer set and fish and dogs represent the diagnostic set. However, whether a location was MC or MI, or an animal was inducer or diagnostic was counterbalanced across individuals.

Table 3

Mean Reaction Time (ms) and Error Rates in Experiment 1

Set Type	Location PC		Trial type		
			Congruent	Incongruent	Congruency Effect
Inducer	MC	RT	646 (15)	750 (19)	103 (8)
		Error rate	.005 (.001)	.044 (.007)	.040 (.007)
	MI	RT	648 (14)	745 (18)	97 (6)
		Error rate	.003 (.001)	.033 (.004)	.031 (.004)
Diagnostic	MC	RT	686 (13)	768 (16)	82 (6)
		Error rate	.007 (.002)	.043 (.007)	.036 (.007)
	MI	RT	683 (13)	769 (17)	86 (7)
		Error rate	.004 (.002)	.038 (.007)	.034 (.007)

Note. Values in parentheses indicate standard error of the mean. MC = mostly congruent; MI = mostly incongruent. Note that items in the diagnostic set were 50% congruent in each location.

Table 4

Stimulus Frequencies in Experiment 2

		Word			
Location PC	**Picture**	**BIRD**	**CAT**	**FISH**	**DOG**
MC	Bird	48	4	4	4
	Cat	4	48	4	4
	Fish	4	4	48	4
	Dog	4	4	4	48
MI	Bird	12	16	16	16
	Cat	16	12	16	16
	Fish	16	16	12	16
	Dog	16	16	16	12

Note: Numbers represent frequencies of stimulus presentation in Experiment 2 collapsed across all four pictures of birds, cats, fish, and dogs, which were presented equally frequently (e.g., for the cell representing a MC bird with the word BIRD, there were 12 presentations of each of the four bird pictures). Underlined frequencies represent congruent trials. Whether a location was MC or MI was counterbalanced across individuals.

Table 5

Mean Reaction Time (ms) and Error Rates in Experiment 2

Counting	Location PC		Trial type		
			Congruent	Incongruent	Congruency Effect
No	MC	RT	665 (17)	787 (21)	121 (11)
		Error rate	.007 (.001)	.032 (.006)	.025 (.005)
	MI	RT	677 (16)	772 (19)	95 (7)
		Error rate	.007 (.002)	.027 (.003)	.020 (.003)
Yes	MC	RT	732 (17)	852 (21)	121 (9)
		Error rate	.006 (.001)	.055 (.006)	.049 (.008)
	MI	RT	744 (16)	842 (19)	98 (6)
		Error rate	.007 (.002)	.032 (.003)	.025 (.004)

Note. Values in parentheses indicate standard error of the mean. MC = mostly congruent; MI = mostly incongruent.

Table 6

Stimulus Frequencies in Experiment 3

			Word			
Location PC	**Set Type**	**Picture**	**BIRD**	**CAT**	**FISH**	**DOG**
MC	Inducer	Bird	72	8	8	8
		Cat	8	72	8	8
	Diagnostic	Fish	4	4	12	4
		Dog	4	4	4	12
MI	Inducer	Fish	24	24	24	24
		Dog	24	24	24	24
	Diagnostic	Bird	12	4	4	4
		Cat	4	12	4	4

Note: Numbers represent frequencies of stimulus presentation in Experiment 3 collapsed across all four pictures of birds, cats, fish, and dogs, which were presented equally frequently (e.g., for the cell representing a MC inducer bird with the word BIRD, there were 18 presentations of each of the four bird pictures). Underlined frequencies represent congruent trials. In this example, birds and cats represent the inducer set in the MC location and fish and dogs represent the inducer set in the MI location, whereas fish and dogs represent the diagnostic set in the MC location and birds and cats represent the diagnostic sets in the MI location. This was counterbalanced across participants, as was assignment of location to PC.

Table 7

Mean Reaction Time (ms) and Error Rates in Experiment 3

Set Type	Location PC		Trial Type		
			Congruent	Incongruent	Congruency Effect
Inducer	MC	RT	664 (10)	777 (14)	113 (7)
		Error rate	.025 (.004)	.074 (.008)	.049 (.008)
	MI	RT	661 (12)	743 (14)	83 (6)
		Error rate	.027 (.004)	.053 (.004)	.026 (.005)
Diagnostic	MC	RT	670 (12)	769 (15)	99 (8)
		Error rate	.021 (.005)	.056 (.008)	.035 (.008)
	MI	RT	677 (12)	796 (14)	119 (8)
		Error rate	.023 (.004)	.062 (.008)	.039 (.008)

Note. Values in parentheses indicate standard error of the mean. MC = mostly congruent; MI = mostly incongruent. Note that items in the diagnostic set were 50% congruent in each location.

Table 8

Stimulus Frequencies in Experiment 4

Location	Item PC	Item Set	Picture	Word			
				BIRD	CAT	FISH	DOG
Upper	MC	Inducer Set 1	Bird	72	8	8	8
			Cat	8	72	8	8
	MI	Inducer Set 2	Fish	24	24	24	24
			Dog	24	24	24	24
Lower	MC	Inducer Set 2	Fish	8	8	72	8
			Dog	8	8	8	72
	MI	Inducer Set 1	Bird	24	24	24	24
			Cat	24	24	24	24

Note: Numbers represent frequencies of stimulus presentation in Experiment 4 collapsed across all four pictures of birds, cats, fish, and dogs, which were presented equally frequently (e.g., for the cell representing an inducer Set 1 bird with the word BIRD, there were 18 presentations of each of the four bird pictures). Underlined frequencies represent congruent trials. In this example, birds and cats play the role of inducer Set 1 and are MC in the upper location and MI in the lower location. Fish and dogs play the role of inducer Set 2 and are MI in the upper location and MC in the lower location. Assignment of pictures to inducer sets was counterbalanced across individuals.

Table 9

Mean Reaction Time and Error Rate in Experiment 4

Location	Item PC	Item Set		Trial Type		
				Congruent	Incongruent	Congruency Effect
Upper	MC	Inducer Set1	RT	670(11)	784 (15)	113 (8)
			Error rate	.018 (.002)	.059 (.007)	.040 (.006)
	MI	Inducer Set2	RT	678(12)	776 (13)	98 (7)
			Error rate	.020 (.004)	.053 (.005)	.034 (.006)
Lower	MC	Inducer Set2	RT	680 (12)	778 (13)	98 (6)
			Error rate	.020 (.002)	.040 (.005)	.020 (.004)
	MI	Inducer Set1	RT	682 (11)	775 (14)	93 (7)
			Error rate	.017 (.003)	.053 (.006)	.036 (.006)

Note. Values in parentheses indicate standard error of the mean. MC = mostly congruent; MI = mostly incongruent. Note that an inducer set of items (e.g., birds and cats) that was MC in one location (e.g., upper), was MI in the other location (e.g., lower).

Table 10

Correspondence of Present Findings with the Consistency Hypothesis and the Three Learning Hypotheses

		Learning Hypothesis		
Experiment	**Consistency Hypothesis**	**Location-PC**	**Item-PC**	**Conjunctive**
1				
2				
3				
4				

Note: Black shading = results are consistent with the hypothesis; medium gray = results are consistent with more than one hypothesis; light gray = results are inconsistent with the hypothesis; white = this experiment was not designed to directly test the consistency hypothesis; for the Consistency Hypothesis column, the relevant finding is the presence/absence of a CSPC effect for just the inducer set or both the inducer and diagnostic sets, where applicable.