Target-distractor correlation does not imply causation of the Stroop effect

Giacomo Spinelli[1] & Stephen J. Lupker[2]

[1] Università degli Studi Milano-Bicocca, Milan (Italy)

[2] The University of Western Ontario, London (Canada)

Running Head: Correlation and the Stroop effect

Corresponding Author:

Giacomo Spinelli

Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milano (MI), Italy

E-mail: giacomospinelli@hotmail.it

Word count: 21762

Abstract

In the Stroop task, the identities of the targets (e.g., colors) and distractors (e.g., words) used are often correlated. For example, in a list in which 4 words and 4 colors are combined to form 16 stimuli, each of the 4 congruent stimuli is typically repeated 3 times as often as each of the 12 incongruent stimuli. Some accounts of the Stroop effect suggest that in this type of list, often considered a baseline because of the matching proportion of congruent and incongruent stimuli (50%), the word dimension actually receives more attention than it does in an uncorrelated list in which words and colors are randomly paired. This increased attention would be an important determinant of the Stroop effect in correlated situations, an idea supported by the observation that higher target-distractor correlation lists are associated with larger Stroop effects. However, because target-distractor correlation tends to be confounded with congruency proportion in common designs, the latter may be the crucial factor, consistent with accounts that propose that attention is adapted to the list's congruency proportion. In 4 experiments, we examined the idea that target-distractor correlation plays a major role in color-word Stroop experiments by contrasting an uncorrelated list with a correlated list matched on relevant variables (e.g., congruency proportion). Both null hypothesis significance testing and Bayesian analyses suggested equivalent Stroop effects in the two lists, challenging accounts based on the idea that target-distractor correlations affect how attention is allocated in the color-word Stroop task.

Keywords: Stroop effect; target-distractor correlation; stimulus informativeness; contingency learning; congruency proportion; adaptive control

Target-distractor correlation does not imply causation of the Stroop effect

One of the most frequently examined effects in cognitive psychology, the Stroop (1935) effect, continues to attract researchers' interest to this day (for a review, see MacLeod, 1991). The Stroop effect refers to the finding that naming the ink color of an incongruent color word (e.g., the word BLUE in the color red) is slower and often less accurate than naming the ink color of a congruent color word (e.g., the word RED in the color red). When a neutral baseline (e.g., the consonant string XXXX in red) is included in the experiment, the Stroop effect can be decomposed into a (typically large) interference effect (with incongruent stimuli producing much slower response times than neutral stimuli) and a (typically small) facilitation effect (with congruent stimuli producing slightly faster response times than neutral stimuli; but note that the relative magnitudes of interference and facilitation depend on the type of baseline used, with facilitation being equivalent or even larger than interference in some cases: Brown, 2011; Melara & Algom, 2003; Sabri et al., 2001).

Part of the reason for the enduring interest in the Stroop effect lies in the fact that, despite being often regarded as a robust effect (e.g., Hedge et al., 2018), more and more processes have been presumed (and, in many cases, found) to modulate it. One of these processes is the degree to which target (i.e., colors) and distractor (i.e., words) components of the stimuli (henceforth, targets and distractors) used in a list are correlated (Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003; see also Sabri et al., 2001). Control accounts of the Stroop effect, that is, accounts that focus on the control processes that are used to deal with target and distractor information, especially when the two pieces of information conflict with one another (e.g., Botvinick et al., 2001; Braver, 2012), make no claim about target-distractor correlations playing a role. However, the same is not true for input-driven accounts, accounts which focus on the "input" itself, that is, on the characteristics of targets and distractors that would determine the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003). According to the latter accounts, in the presence of a non-zero target-distractor correlation, participants may learn to associate each distractor with the targets (including incongruent targets) that that distractor will likely appear with. The distractor would then receive more attention in that type of list compared to an uncorrelated list, that is, a list in which the distractor is not informative at all because the distractors and the targets are randomly paired. As a result of receiving increased attention in a correlated list, the distractor will have a stronger impact on performance than in an uncorrelated list.[1]

Consider, for example, the popular 4 × 4 design in which 4 words and 4 colors are combined to form 16 stimuli (4 congruent and 12 incongruent). This type of design is often used to create a list of stimuli in which congruent and incongruent stimuli are equally frequent overall (e.g., 96 congruent stimuli and 96 incongruent stimuli in a 192-stimuli list) by repeating each of the 4 congruent stimuli 3 times as often as each of the 12 incongruent stimuli (see Table 1). Although this type of list is often considered a baseline in Stroop research because of the matching proportion of congruent and incongruent stimuli (50% each), it is in fact a correlated list because each word in that list is more likely to appear in its congruent color than in any other (incongruent) color. As a result, the Stroop effect in that list might be driven, in whole or in part, by the increased attention that the word would receive in that situation (Algom & Chajut, 2019; Algom et al., 2022; for a similar point, see Mordkoff, 2012). The same would not be true

[1] Note that this process is distinct from the process of learning contingencies between a distractor and the response typically made to the stimuli containing that distractor (Schmidt et al., 2007). The reason is that, first, contingency learning, per se, does not involve a modulation of attention to distractors and/or targets. Second, contingency learning concerns individual distractors in the list (e.g., the word RED as opposed to the word BLUE, etc.) rather than the overall association between distractors and targets in the list as a whole. Finally, whereas, as will be explained, the process of increasing attention to distractors in a correlated list would only lead to an increase of the Stroop effect, the contingency-learning process would either increase or reduce the Stroop effect depending on whether the typical response for the distractor is the congruent one (e.g., the response “red” for the word RED) or an incongruent one (e.g., the response “red” for the word BLUE).

for a less common type of list in which words and colors are randomly paired. In that list, represented in Table 2, there would be no correlation between words and colors. As a result, attention to the word would not be increased by the process of learning color-word correlations and, hence, performance would not be artificially biased toward word processing in that context.

Table 1

Example of a correlated list in the Stroop task

	Word			
Color	RED	BLUE	GREEN	YELLOW
Red	24	8	8	8
Blue	8	24	8	8
Green	8	8	24	8
Yellow	8	8	8	24

Note. The proportion of incongruent stimuli in this list is 50% and *C*, a chi-square based correlation representing the degree to which color and word identities are correlated (see the main text for a definition), is .5.

Table 2

Example of an uncorrelated list in the Stroop task

	Word			
Color	RED	BLUE	GREEN	YELLOW
Red	12	12	12	12
Blue	12	12	12	12
Green	12	12	12	12
Yellow	12	12	12	12

Note. The proportion of incongruent stimuli in this list is 75% and *C*, a chi-square based correlation representing the degree to which color and word identities are correlated (see the main text for a definition), is 0.

The idea that the correlation between targets and distractors plays an important role in the Stroop effect has gained support from a few findings. For example, Dishon-Berkovits and Algom (2000) reported that, in a series of experiments in which participants were presented with a country name (e.g., Austria) and a city name (e.g., Salzburg) simultaneously and were instructed either to respond by saying the country name or by saying the city name, a congruency effect (i.e., faster response times for congruent country-city combinations, e.g., Austria-Salzburg, than for incongruent country-city combinations, e.g., Austria-Bombay) did not emerge when country and city names were randomly paired (i.e., in an uncorrelated list). The lack of a congruency effect in that situation may partially be due to the fact that city names and the corresponding country names are unlikely to be strongly associated (certainly not to the level that colors and the corresponding color names are, for example). Interestingly, however, the effect did emerge (although non-significantly in some cases) when country and city names were correlated, either “positively” (that is, when congruent stimuli were more probable than incongruent stimuli) or “negatively” (that is, when incongruent stimuli were more probable than congruent stimuli). Assuming that the uncorrelated list represents the most appropriate baseline for examining the impact of distractor processing on performance, these results suggest that the congruency effect that was observed in the correlated lists may have resulted entirely from the increased attention that the distractor presumably received in those lists.

Perhaps the most persuasive piece of evidence in support of target-distractor correlation being important in the Stroop effect is Melara and Algom’s (2003) re-analysis of the Stroop effect as a function of target-distractor correlation for a set of 34 previously published experiments (see, e.g., Algom & Chajut, 2019; Algom et al., 2022). For the purpose of that re-analysis, Melara and Algom represented the strength of the target-distractor correlation with *C*, a chi-square based correlation whose absolute value was defined as

$$|C| = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

where χ^2 is the chi-square statistic reflecting the association among targets and distractors in the list, and *N* is the total number of stimuli in the list. *C* takes on positive values when congruent stimuli are more probable than incongruent stimuli, negative values when incongruent stimuli are more probable than congruent stimuli, and a value of zero when targets and distractors are completely uncorrelated. In their re-analysis, Melara and Algom found a striking Pearson's $r = .69$ for the correlation between Stroop effects and *C*, suggesting that *C* explains 48% of the variance in Stroop effects. This result would appear to make a strong case for target-distractor correlation being a crucial determinant of the Stroop effect.

Target-distractor correlation and the Proportion-Congruent effect

Proponents of input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022) have recently given new emphasis to target-distractor correlation in order to explain an effect that has gained popularity in Stroop research in recent years: the Proportion-Congruent (PC) effect (Logan & Zbrodoff, 1979). The PC effect refers to the finding that the Stroop effect is larger in a list in which congruent stimuli are more frequent than incongruent stimuli (i.e., a Mostly Congruent [MC] list) than in a list in which incongruent stimuli are more frequent than congruent stimuli (i.e., a Mostly Incongruent [MI] list).

The PC effect has played, and continues to play, an important role in control accounts, accounts which focus on the ways in which attention is adjusted in response to certain events, particularly events in which conflict is experienced (e.g., Botvinick et al., 2001; Braver, 2012). In general, those accounts interpret the PC effect as the result of a process of adaptation of attention to the list's congruency proportion (for a review, see Bugg & Crump, 2012). More specifically, attention to the target (e.g., the color) and away from the distractor (e.g., the word) would be increased when the distractor frequently conflicts with the target (i.e., when incongruent stimuli are frequent, as in an MI list), resulting in reduced Stroop effects in that situation compared to when the distractor does not frequently conflict with the target (e.g., when congruent stimuli are the most frequent stimuli, as in an MC list). [2]

What is important to note for present purposes, however, is that standard PC manipulations tend to confound congruency proportion and target-distractor correlation systematically, with the target-distractor correlation typically being larger in MC lists than in MI lists. For example, in a 4 × 4 design, an MC list with 25% incongruent stimuli is often created by presenting each of the 4 congruent stimuli 9 times more often than each of the 12 incongruent stimuli, introducing a large positive correlation (in terms of *C*, .76). This type of list is represented on the left side of Table 3. In contrast, in a 4 × 4 design, an MI list with 75% incongruent stimuli is often created by randomly pairing words and colors. That is, an MI list in a 4 × 4 design, represented on the right side of Table 3, is a repetition of the uncorrelated list (*C* = 0) shown in Table 2.

---

[2] There can be several levels at which the target and the distractor can conflict with one another, for example, at the task level (e.g., for any readable colored stimulus, there can be a conflict between naming the color and reading the word), at the semantic level (e.g., for RED in blue, there can be a conflict between the concepts of "red" and "blue") and at the response level (e.g., there can be a conflict between "red" and "blue" responses; Parris et al., 2021). Exactly which level of conflict triggers attentional adjustments in the PC effect still needs to be determined (for an initial discussion, see Spinelli & Lupker, 2022b). For present purposes, however, this determination is not important. What is important, as noted, is the frequency with which the distractor conflicts with the target (at any level).

Table 3

Example of a standard Proportion-Congruent manipulation in the Stroop task

	List type							
	Mostly Congruent				Mostly Incongruent			
Color	RED	BLUE	GREEN	YELLOW	RED	BLUE	GREEN	YELLOW
Red	36	4	4	4	12	12	12	12
Blue	4	36	4	4	12	12	12	12
Green	4	4	36	4	12	12	12	12
Yellow	4	4	4	36	12	12	12	12

Note. In the Mostly-Congruent list, the proportion of incongruent stimuli is 25% and *C* is .76. In the Mostly-Incongruent list, the proportion of incongruent stimuli is 75% and *C* is zero.

As a result, proponents of input-driven accounts have argued that the PC effect obtained in this type of contrast reflects a process of adaptation of attention to the list's target-distractor correlation, rather than to the list's congruency proportion as control accounts propose (Algom & Chajut, 2019; Algom et al., 2022). That is, in input-driven accounts, the relatively large Stroop effect in the MC list would be the result of increased attention to the distractor due to the large target-distractor correlation that is created in that list, rather than being the result of attention being decreased due to the infrequent presentation of incongruent stimuli in that list. In contrast, the smaller Stroop effect in the MI list would be the result of the distractor not attracting attention due to the absence of a target-distractor correlation in that list, rather than due to the frequent presentation of incongruent stimuli in that list.

Although the claims of input-driven accounts are suggestive, they are certainly not conclusive. The reason is that the confound between congruency proportion and target-distractor correlation works both ways. That is, just as differences in congruency proportion often involve a target-distractor correlation confound, the reverse is also true – differences in target-distractor correlation often involve a congruency-proportion confound. For example, because Dishon-Berkovits and Algom (2000) used a set of 7 targets and 7 distractors in their experiments, the uncorrelated list they used was in fact an MI list with 86% incongruent stimuli. Therefore, the fact that no congruency effect emerged in that list may not only be explained by the fact that attention to the distractor was not increased because of the zero target-distractor correlation, but also by the fact that attention to the *target* was *in*creased because of the high frequency of incongruent stimuli. The same is also true in a number of other experiments, that is, that the uncorrelated list was also an MI list with a large proportion of incongruent stimuli (e.g., Arieh & Algom, 2002; Sabri et al., 2001).

Congruency proportion and target-distractor correlation were also confounded in the experiments included in Melara and Algom's (2003) re-analysis, with experiments with higher *C* values tending to have a lower proportion of incongruent stimuli. For example, in one of the experiments included in

Melara and Algom's re-analysis, Glaser and Glaser's (1982) Experiment 2, *C* was .83 (the highest value in the re-analysis) and the proportion of incongruent stimuli was 10% (the lowest value in the re-analysis). Therefore, the large Stroop effect observed in that experiment (371 ms) could be either the result of attention to the distractors being increased because of the strong target-distractor correlation in that experiment or the result of attention to the distractors being decreased because of the infrequent presentation of incongruent stimuli.[3]

Controlling for confounds in Proportion-Congruent manipulations

In general, in the presence of a confound which allows two alternative processes to explain an effect, making a conclusive argument for either process being potentially important would seem to require eliminating the confound and producing evidence that is only consistent with one of those processes. In recent years, the literature on PC effects has taken this route in order to address several potential confounds that would allow explanations for a PC effect in which the list's congruency proportion (or conflict frequency) is not the crucial factor.

One of these confounds is based on the proposal that participants engage a process of binding features of the current stimulus to the features of previously presented stimuli when one or more of the features repeats (i.e., a feature-binding or, more generally, repetition-priming process: Hommel, 1998; Logan, 1988) and/or the distractor-response contingency learning process noted in footnote 1 (Schmidt et al., 2007) – two processes that, according to Schmidt et al. (2020), are essentially the same process acting at different time scales. The reason that these types of processes would explain the PC effect is that, in a typical PC manipulation, any word in the MC list is most frequently combined with the congruent color

[3] The values of *C* and of the Stroop effect were computed directly from Glaser and Glaser (1982). They are slightly different from those reported in the Appendix of Melara and Algom (2003), i.e., a *C* of .82 and a Stroop effect of 375 ms.

and requires a congruent response (e.g., in Table 3, the word RED is presented with the red color 36 times whereas it is presented with each of the incongruent colors only 4 times). Because each congruent stimulus is frequently presented, in a randomized list, there will be many instances in which all of the features of the stimulus (i.e., the color, the word, and the required response) repeat from trial to trial, allowing rapid retrieval of the required response in the current trial and resulting in a large benefit overall. In contrast, because each incongruent stimulus is infrequently presented, there will not be many instances in which all of the features of the stimulus repeat from trial to trial, resulting in little or no benefit overall. At longer time scales, a strong contingency between each word and the congruent response will be learned but not between each word and incongruent responses. Either way, the result will be an increased Stroop effect in the MC list because high-frequency congruent stimuli would speed up whereas low-frequency incongruent stimuli would not.

In contrast, in an MI list constructed as in Table 2, the Stroop effect would not be increased by any repetition-priming/contingency-learning process because each stimulus has the same frequency and each word is equally associated with each of the responses (e.g., the word RED requires each of the 4 responses 12 times). As a result, the Stroop effect would be smaller in that list compared to an MC list in which repetition-priming/contingency-learning processes increase the Stroop effect (for discussions, see Cochrane & Pratt, 2022; Hazeltine & Mordkoff, 2014; Schmidt, 2013a, 2019; Schmidt & Besner, 2008; Tzelgov et al., 1992).

A second type of confound, as noted, is created by the process of adjusting attention to the target-distractor correlation in the list, a process which would produce the PC effect as explained above. An associated notion is that of “stimulus informativeness” (Schmidt, 2014, 2019). Stimulus informativeness, as defined by Schmidt, refers to the degree to which distractors in a list allow learning of distractor-response contingencies. The more a list is informative, that is, the stronger the average contingency between distractors and responses in the list, the more participants will be inclined to allocate attention to distractors. As a result, the impact of distractors on performance in that list will be stronger. For example, an MC list such as that represented in Table 3 would be an informative list because a strong contingency (i.e., 36 to 4) can be learned for each word. Because attention to words would be increased in that list, a large Stroop effect would be produced. In contrast, an MI list constructed as in Table 2 would not be informative because there are no contingencies to learn in that list. As a result, attention to words would not be increased and, yet again, the Stroop effect would be smaller in that list compared to an MC list in which stimulus informativeness increases the Stroop effect.

It is important to note here that although the notions of target-distractor correlation and stimulus informativeness are closely related, they are partially dissociable (for a discussion, see Spinelli & Lupker, 2021). The reason is that the presence of contingencies is not required for a large target-distractor correlation to arise. For example, a situation can be created in which no word in the list is associated with one color response in particular (i.e., a list in which no contingencies can be learned), yet words and colors are correlated because words appear in some of the colors used but not in others. This type of list would be one in which, for example, each of the words RED and BLUE is combined with each of the colors red and blue an equal number of times (e.g., 24 per combination) but never with the colors green and yellow, and vice versa for the words GREEN and YELLOW and the corresponding colors. Although no word-response contingency would exist in that list, the color-word correlation would be quite large (C = .71). On the other hand, an uncorrelated list is always an uninformative list (in Schmidt's sense of the word), and an informative list (also in Schmidt's sense of the word) in which contingencies can be learned always involves a sizeable target-distractor correlation. Because in the present research we will only be dealing with situations in which target-distractor correlation and stimulus informativeness are consistent with one another, the distinction between the two notions is irrelevant (although we will come back to this point in the General Discussion).

In sum, several confounds exist in typical PC manipulations, such as those involving an MC list and an MI list constructed as in Table 3, which could explain the PC effect obtained in those manipulations without invoking congruency proportion as the crucial factor. In recent years, however, alternative PC manipulations have been developed in which the impact of congruency proportion on performance could be examined with the confounds playing little or no role (Blais & Bunge, 2010; Bugg, 2014; Bugg et al., 2008; Hutchison, 2011; Spinelli & Lupker, 2021, 2022b; Spinelli et al., 2019; see also Braem et al., 2019).

In the most common of these alternative manipulations (Bugg et al., 2008), the stimuli used in the two lists are divided into two sets. One set (the inducer or context set, formed, e.g., by the words RED and BLUE and their corresponding colors) involves the same repetition-priming/contingency-learning confounds existing for all items in standard PC manipulations, whereas the other set (the diagnostic or transfer set, formed, e.g., by the words GREEN and YELLOW and their corresponding colors) does not (for a detailed explanation, see Braem et al., 2019). Further, the two lists can be constructed so that they are matched on target-distractor correlation and stimulus informativeness (Spinelli & Lupker, 2022b; see also Spinelli & Lupker, 2021). As a result, all confounds presumed to affect the PC effect are controlled for in the second set of stimuli, the diagnostic set (but for yet another potential confound, see Schmidt, 2013b, and, for further discussion, Cohen-Shikora et al., 2019; Schmidt, 2017, 2021; Spinelli & Lupker, 2022a; Spinelli et al., 2019). Nevertheless, the diagnostic stimuli typically do produce a PC effect (although not always, with the effect typically being smaller than the effect produced by all stimuli in standard PC manipulations; see Spinelli & Lupker, 2022b). The process of adapting attention to the frequency with which the distractor conflicts with the target is, therefore, dissociable from the processes that, in standard PC manipulations, are confounded with it.

The present research

Having established that attentional adaptation to the list's congruency proportion is dissociable from the processes potentially confounded with it in PC manipulations, in the present research we reverse the question – are the processes that have been claimed to confound PC manipulations dissociable from each other and from the process of adapting attention to the list's congruency proportion? With respect to repetition-priming/contingency-learning processes, there is clear evidence for those processes taking place in both non-Stroop tasks, that is, tasks that do not involve a conflict between target and distractor components of the stimuli used (e.g., Hommel, 1998; Lin & MacLeod, 2018; Schmidt et al., 2007), and Stroop tasks (e.g., Schmidt, 2013c; Spinelli & Lupker, 2020; Whitehead et al., 2018). The same is not true for processes of adaptation to the target-distractor correlation and stimulus informativeness, however. In Stroop tasks in particular, to our knowledge, there is no evidence that, as both processes would imply, Stroop effects are larger in lists in which the target-distractor correlation is large (vs. small or zero) and/or in informative (vs. uninformative) lists *in manipulations in which potential confounds are controlled*.

Concerning target-distractor correlation in particular, as noted above, the evidence brought in support of the idea that attention to the distractor would be adapted to the strength of that correlation mainly comes from contrasts in which the list's target-distractor correlation was confounded with the list's congruency proportion. Specifically, uncorrelated lists (lists which are also uninformative in Schmidt's sense of the word) are inevitably MI lists when the set of stimuli used to create those lists is larger than two (see, e.g., Table 2). The same is not true for lists with a non-zero target-distractor correlation (lists that often, but not always, are informative because they contain contingencies). Lists of that type can be MI lists, they can have an equal proportion of congruent and incongruent stimuli (see, e.g., Table 1), or they can be MC lists (see, e.g., the MC list in Table 3). If either equal-proportion or MC lists are contrasted with uncorrelated lists that are also MI lists (as has occurred, for example, in Melara and Algom's (2003) re-analysis), the fact that larger Stroop effects are observed with the former lists might be due to the lower proportion of incongruent stimuli in those lists rather than the strength of the target-distractor correlation (and/or stimulus informativeness). In other words, the contrast between lists varying in target-distractor correlation cannot be used as a conclusive argument for an impact of target-distractor correlation (and/or stimulus informativeness) unless congruency proportion is controlled.

The same reasoning applies for the other confound that we have discussed in relation to the PC effect, i.e., the confound created by repetition-priming/contingency-learning processes. In uncorrelated (uninformative) lists, each congruent stimulus and each incongruent stimulus have the same frequency (e.g., 12, see Table 2) and, therefore, each distractor has the same probability of requiring either a congruent response or incongruent responses. As a result, the Stroop effect in that list will not be modified by repetition-priming/contingency-learning processes. In contrast, in correlated (informative) lists, at least in standard Stroop experiments which do not divide the stimuli used into separate sets, each congruent stimulus typically has a higher frequency than each incongruent stimulus if the list is either a 50:50 congruent/incongruent list (e.g., 24 vs. 8, see Table 1) or if the list is an MC list (e.g., 36 vs. 4, see Table 3). As a result, repetition-priming/contingency-learning processes will increase the Stroop effect in that type of lists compared to uncorrelated lists. Therefore, those processes also need to be controlled when contrasting lists varying in target-distractor correlation (and/or stimulus informativeness) in order to be able to claim that target-distractor correlation does impact performance.

Here, we present four color-word Stroop experiments in which the impact of target-distractor correlation (and/or stimulus informativeness) on the Stroop effect was examined while controlling for both congruency proportion and repetition-priming/contingency-learning processes. In all experiments, an uncorrelated list was contrasted with a list in which a large correlation was introduced by changing the frequencies of color-word combinations but 1) the congruency proportion was the same in the two lists, and 2) the stimuli used in the contrast had equal frequencies in the two lists. In such a situation, any difference between the two lists would have to be attributed to the different strength of the color-word correlation (and/or stimulus informativeness) in the two lists.

Specifically, we tested the hypothesis made by input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003) stating that, because attention to the word would be increased in correlated lists compared to lists with a smaller color-word correlation, the word would have a larger impact on the Stroop effect in the former lists. This larger impact implies a larger Stroop effect (incongruent – congruent) overall but also larger differences in facilitatory (neutral – congruent) and inhibitory (incongruent – neutral) components of the effect. The reason is that, the more attention allocated to a word, the greater the potential for the word to facilitate responding when the word is congruent with the color (e.g., RED in red) and to interfere with responding when the word is incongruent with the color (e.g., RED in blue). We tested these ideas in Experiments 1 and 2, which included (in addition to congruent and incongruent stimuli) a letter-string neutral stimulus (i.e., XXXX) and a word neutral stimulus (i.e., the word SHORT), respectively.

It is important to remember that in standard Stroop experiments, as noted above, facilitation is typically small, i.e., response times for congruent stimuli are only slightly faster than those for neutral stimuli, at least when the neutral stimuli used are letter strings such as XXXX (for a review, see MacLeod, 1991). One of the possible reasons for facilitation being small, as noted by MacLeod, is that it may not be easy to speed up a response that is already quite fast. By the same reasoning, it is possible that the impact of color-word correlation on facilitation may not be particularly large, and most of the action would be in the interference component of the Stroop effect.

It is also worth noting that regular facilitation (i.e., congruent faster than neutral) is not inevitable, with some experiments reporting no difference between congruent and neutral stimuli (e.g., Dalrymple-Alford, 1972; Dunbar & MacLeod, 1984) and some experiments even reporting a reversal of the facilitation pattern (i.e., congruent *slower* than neutral; e.g., Schulz, 1979; Sichel & Chandler, 1969; see also Goldfarb & Henik, 2007), although these patterns are unusual and may depend on the choice of the neutral stimulus (MacLeod, 1991; see also Brown, 2011). However, to foreshadow our results, we did obtain reverse facilitation in Experiments 1 and 2 despite the fact that different neutral stimuli were used in the two experiments. In Experiments 3 and 4, we therefore eliminated neutral stimuli from the design and focused only on congruent and incongruent stimuli.

The present experiments are somewhat similar to a Stroop experiment recently reported by Hasshim and Parris (2021). In their Experiment 2, Hasshim and Parris contrasted a list with a relatively small color-word correlation (C = .28), which they called the "standard" design, with a list with the same congruency proportion and the same type of stimuli (i.e., congruent, incongruent, and word neutral stimuli), but arranged in such a way that a larger color-word correlation was created (C = .5), a list that they called the "alternative" design.[4] The focus of the contrast was the facilitation effect (congruent – neutral). Although Hasshim and Parris did not intend to create a contrast in color-word correlation, the expectation, based on the present discussion, would seem to be for a larger facilitation effect in the "alternative" (relatively-large-correlation) list than in the "standard" (relatively-small-correlation) list. Although facilitation was numerically larger in the former list (19 ms) than in the latter (10 ms), the difference was not statistically significant. The idea that target-distractor correlation has an important role in the Stroop effect, thus, did not gain support from Hasshim and Parris's results.

However, when considering those results, a few points are worth mentioning. First, as noted above, the fact that Stroop facilitation is typically small may have to do with performance being at floor for congruent stimuli (MacLeod, 1991). Observing a speed-up with those stimuli (and, thus, a facilitation

[4] These C values are slightly different from those reported by Hasshim and Parris (2021) for their "standard" and "alternative" design (.31 and .58, respectively).

increase) may be somewhat difficult even if attention to the word *were* increased in correlated lists. What may be easier is observing an *interference* increase in correlated lists, a contrast that Hasshim and Parris's (2021) experiment was not set up to examine.

Second, although the strength of the color-word correlation differed in the two lists contrasted by Hasshim and Parris (2021), it did not differ greatly (*C* = .28 vs. *C* = .5). Therefore, the lack of a list difference for the facilitation effect may not be due only to the difficulty of observing that difference in general but also to the color-word correlation difference not being large enough in that experiment (note, again, that Hasshim and Parris did not intend the contrast between the two lists they used in their Experiment 2 to be a color-word correlation contrast).

Finally, Hasshim and Parris (2021) used a Stroop experiment with manual (i.e., button-press) instead of vocal responses to colors. Manual responses to colors in Stroop experiments are not ideal because they change the nature of the task in potentially important ways. For example, while in a vocal Stroop experiment the word BLUE would produce a strong tendency to say "blue", in a manual Stroop experiment that word would not produce a strong tendency to press the (arbitrary) key designated for the blue response (for a review of supporting evidence, see MacLeod, 1991). This difference is potentially important in relevant manipulations such as PC manipulations, in which PC effects for items for which repetition-priming/contingency-learning processes are controlled are small or absent altogether when manual responses are required in color-word or picture-word Stroop experiments (for a discussion, see Spinelli & Lupker, 2022b). Therefore, in general, the question of whether target-distractor correlation has an impact on the Stroop effect would seem best addressed in the standard, vocal Stroop task.

These concerns do not apply to the present experiments because 1) the experiments were set up to examine the impact of target-distractor correlation on both facilitatory and inhibitory components of the Stroop effect, 2) large differences in color-word correlations were created because one of the lists being contrasted was always an uncorrelated list, and 3) a vocal color-word Stroop task was used. Thus, overall, the present experiments would seem to produce the most direct test to date of the idea, proposed by input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003; see also Schmidt, 2014, 2019), that target-distractor correlation has an important role in that effect.

Experiment 1

In Experiment 1, we contrasted an uncorrelated list with a correlated list, with both lists including congruent, incongruent, and letter-string neutral stimuli (i.e., XXXX), a typical neutral stimulus in Stroop experiments (MacLeod, 1991). In order to control for congruency proportion and repetition priming/contingency learning in the contrast, it was necessary to use a set of at least four colors, the corresponding color names, and the neutral letter string. In both lists, each of the four colors occurred with the same frequency and so did each of the five “words” (i.e. the color names and XXXX). The uncorrelated list was created by randomly pairing each color with each word and is represented on the left side of Table 4. Note that, inevitably, this list is an MI list because it contains 120 incongruent stimuli (60%), 40 congruent stimuli (20%), and 40 neutral stimuli (20%).

Based on the uncorrelated list, another list, represented on the right side of Table 4, was created with the same congruency proportion and containing stimuli with the same frequency as in the uncorrelated list, but with a larger color-word correlation overall. This result was achieved by modifying the uncorrelated list so that 1) each color name was combined with one of the three incongruent colors twice as frequently as in the uncorrelated list (e.g., RED appeared in green 20 times, twice as often as it did in the uncorrelated list), and 2) each color name was only combined with one additional incongruent color (e.g., RED never appeared in yellow).

Despite these modifications, the overall congruency proportion was maintained (i.e., there still were 60% incongruent stimuli, 20% congruent stimuli, and 20% neutral stimuli in this list), and congruent and neutral stimuli were left untouched as was the combination with the additional incongruent color for each color name (e.g., RED in blue, which had a frequency of 10 just as it did in the uncorrelated list). These stimuli (i.e., the stimuli with a frequency of 10) were therefore identical as in the uncorrelated list and were used for the contrast between the two lists.

What did change, though, was the color-word correlation, which, in terms of *C*, went from 0 in the uncorrelated list to -.53 in the correlated list. Note that the minus of the *C* value is due to the MI nature of the list. What is important, however, is not the sign but the absolute value: The higher that value, the stronger the color-word correlation, and, therefore, the more attention should be allocated to the word. The stimulus informativeness of the list also changed. Whereas the uncorrelated list was an uninformative list because no contingencies could be learned in that list, the correlated list was informative because, for each word, a contingency with a particular color response could be learned (e.g., the response “green” for the word RED, see Table 4). Because the correlated list is an informative list, the effect should be the same as that produced by the color-word correlation: Attention to the word should be increased in that list.

In line with these ideas, the hypotheses we tested were that, for the stimuli that were identical in the two lists, the overall Stroop effect (incongruent – congruent), the facilitation effect (congruent – neutral), and the interference effect (incongruent – neutral) would be larger in the correlated list than in the uncorrelated list.

Table 4

Template for the frequency of color-word combinations in Experiment 1

	List type									
	Uncorrelated					Correlated				
Color	RED	BLUE	GREEN	YELLOW	XXXX	RED	BLUE	GREEN	YELLOW	XXXX
Red	10	10	10	10	10	10		20	10	10
Blue	10	10	10	10	10	10	10		20	10
Green	10	10	10	10	10	20	10	10		10
Yellow	10	10	10	10	10		20	10	10	10

Note. In the uncorrelated list, the proportion of incongruent stimuli is 60% and *C* is zero. In the correlated list, the proportion of incongruent stimuli is also 60% and *C* is -.53. The high-frequency color-word combinations (i.e., the color-word combinations occurring 20 times) in the correlated list were not used in the contrast between the two lists.

Method

Participants

An *a priori* power analysis was performed using G*Power 3.1 (Faul et al., 2009) to calculate the sample size needed to have a power of .80 for obtaining an interaction between the list's target-distractor correlation and Stroop congruency types. Reasoning that this effect, examined without the congruency-proportion confound, would be of a similar size as the PC effect examined without the target-distractor correlation confound, we based the analysis on the size of the PC effect produced by diagnostic items in Spinelli and Lupker's (2022b) Experiment 1, the experiment of those reported by Spinelli and Lupker that produced the smallest effect size, η_p^2 = .276. Based on this effect size, we determined that a minimum sample size of 24 participants would be needed. In this and the following experiments, we exceeded that sample size while attempting, where possible, to achieve a sample size comparable to that used in Spinelli and Lupker's (2022b) experiments, i.e., 48 participants.

Forty-nine students at the University of Western Ontario took part in this experiment for course credit. After discarding too-fast, too-slow, and incorrect responses (see below), one participant contributed fewer than 75% of their original observations – our criterion for participant exclusion, determined *a priori* in line with previous work in our laboratory (Spinelli et al., 2020, 2022b). That participant was excluded, leaving 48 participants (29 females and 19 males; age 17-30 years; all native English speakers with normal or corrected-to-normal vision).

Materials

Four color names (RED, BLUE, GREEN, and YELLOW) and a neutral letter string (XXXX) were used as distractors, and the corresponding colors (red [R: 255; G: 0; B: 0], blue [R: 0; G: 0; B: 255], green [R: 0; G: 192; B: 0], and yellow [R: 255; G: 255; B: 0]) were used as targets. "Words" (including the letter string) and colors were combined to form two lists: an uncorrelated list and a correlated list.

In the uncorrelated list, represented on the left side of Table 4, each of the words used (including the letter string) was presented with each of the colors used 10 times (what we will call the "normal frequency" in this experiment). In the correlated list, the letter string was presented with each of the colors 10 times (as in the uncorrelated list) whereas each of the color names was presented with the congruent color 10 times, with one of the three incongruent colors 10 times, with another of the three incongruent colors 20 times (what we will call the "high frequency" in this experiment), and never with the third incongruent color. The frequency of the color-word combinations for the correlated list in one of the counterbalancings in the experiment is presented on the left side of Table 4. The reason that Table 4 represents only one of the counterbalancings for that list is that, for each color name in the correlated list, the specific incongruent colors used to form combinations with a frequency of 10, 20, or zero were counterbalanced across participants. The order with which the correlated and uncorrelated lists were presented to participants was also counterbalanced across participants.[5] Both lists contained 200 stimuli, of which 120 were incongruent (60%), 40 congruent (20%), and 40 neutral (20%). As noted, C for the uncorrelated list was zero whereas it was -.53 for the correlated list.

Procedure

Each trial began with a fixation symbol ("+") displayed for 250 ms in the center of the screen followed by a colored word displayed for 2000 ms or until the participant's response, which was recorded with a microphone connected to the testing computer. Participants were instructed to name the color of the word as quickly and as accurately as possible while ignoring the word itself. Stimuli were presented in uppercase Courier New font, pt. 14, against a medium grey background (R: 128; G: 128; B: 128). No

[5] By mistake, one participant was assigned to a counterbalancing with the uncorrelated-first order instead of the corresponding counterbalancing with the correlated-first order, resulting in 25 participants in the former order and 23 in the latter.

feedback was provided. There was a self-paced pause between the two lists. The order of trials within each list was randomized. Initially, participants performed a practice session including 8 trials in which the neutral letter string (XXXX) was presented in each of the four colors used in the experiment twice (the practice session included neither congruent nor incongruent stimuli). The experiment was run using DMDX (Forster & Forster, 2003) software. For this and the following experiments, participants provided written informed consent at the beginning of the study. This research was conducted according to the Declaration of Helsinki and was approved by the Research Ethics Board of the University of Western Ontario (protocol # 108956).

Results

For this and the following experiments, the waveforms of responses were manually inspected with CheckVocal (Protopapas, 2007) in order to determine the accuracy of the response and the correct placement of timing marks. Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than 2000 ms, the time limit (accounting for 1.5% of the data points), were discarded.

Mean response times for correct responses and mean error rates for the normal-frequency stimuli (i.e., the stimuli appearing with a frequency of 10 in each of the two lists) were analyzed with a 3 × 2 ANOVA with Congruency (congruent vs. neutral vs. incongruent) and List Type (uncorrelated vs. correlated) as within-subject factors.[6] For effects involving the Congruency factor, degrees of freedom were corrected

---

[6] In order to account for potential effects of time on task, for this and the following experiments, the analyses were repeated using Order (correlated list first vs. uncorrelated list first) as an additional between-subject factor. Although fatigue effects (i.e., slower/less accurate performance in the list presented second than in the list presented first) tended to emerge, they affected the interaction between Congruency and List Type in no case but one, noted in footnote 7.

using Greenhouse-Geisser's correction when Mauchly's test indicated that the assumption of sphericity was violated. Further, for both RTs and error rates, we computed mean Stroop, facilitation, and interference effects in the two lists and contrasted them using a one-tailed *t*-test reflecting the alternative hypothesis that those effects would be smaller in the uncorrelated list than in the correlated list, as predicted by input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003; see also Schmidt, 2014, 2019).

Both sets of analyses were conducted with both traditional Null Hypothesis Significance Testing (NHST) analyses and Bayes Factor analyses using JASP version 0.16.4 (JASP Team, 2022). Bayes Factor analyses were used to quantify the evidence supporting the presence vs. the absence of a difference between the effects in the uncorrelated list vs. the correlated list. In the ANOVAs, this quantification was performed by comparing the effect without the interaction between Congruency and List Type (interpreted as the null hypothesis H_0) and the model with that interaction (interpreted as the alternative hypothesis H_1). Note, however, that the H_1 in an ANOVA is not a directional hypothesis. The directional hypotheses H_- (note the minus in the subscript) that Stroop, facilitation, and interference effects would be smaller in the uncorrelated list vs. the correlated list were tested in Bayesian *t*-tests (against the null hypothesis of no difference). The result of these comparisons was BF_{01} (BF_{0-} for the *t*-tests), with $BF_{01} > 1$ suggesting evidence in support of H_0 (i.e., the absence of the interaction), and $BF_{01} < 1$ suggesting evidence in support of H_1 (i.e., the presence of the interaction) ($BF_{01} = 1$ would suggest equal evidence for the two hypotheses). Jeffreys's (1961) classification scheme (as reported in adjusted form by Lee and Wagenmakers, 2013) was used to help interpret the size of the Bayes Factor. JASP's default prior specification was used.

Note again that the analyses described up to this point were based on all stimuli excluding the high-frequency incongruent stimuli in the correlated list (e.g., RED in green). An additional set of *t*-tests comparing those stimuli and normal-frequency incongruent stimuli (e.g., RED in blue) was conducted in order to test the hypothesis that the former would be faster and/or more accurate than the latter, consistent with repetition-priming/contingency-learning processes.

The mean RTs and error rates are presented in Table 5. For this and the following experiments, the study materials, the raw data and the JASP files used for the analyses are publicly available at https://osf.io/buvxy/. The study was not preregistered.

Table 5

Mean RTs and percentage error rates (and corresponding standard errors) for Experiment 1

	RTs		Error rates	
Stimulus type	Uncorrelated list	Correlated list	Uncorrelated list	Correlated list
Congruent	756 (18)	769 (18)	.32 (.12)	.65 (.21)
Neutral	723 (17)	729 (16)	.47 (.16)	.22 (.10)
Incongruent	841 (20)	839 (19)	2.67 (.44)	2.90 (.46)
High-frequency incongruent		828 (19)		2.69 (.40)
Stroop effect	85	70	2.35	2.25
Facilitation effect	-33	-40	.15	-.43
Interference effect	118	110	2.20	2.68
Stimulus frequency effect		11		.21

RTs. There was a main effect of Congruency, $F(2, 94) = 145.23$, $MSE = 2241$, $p < .001$, $\eta_p^2 = .756$. Post-hoc tests with Holm adjustments for multiple comparisons indicated that, overall, incongruent stimuli were slower than both congruent stimuli (a mean difference of 78 ms), $t(47) = -11.35$, $p < .001$, and neutral stimuli (a mean difference of 114 ms), $t(47) = -16.69$, $p < .001$. That is, there were regular Stroop and interference effects, as expected. Unexpectedly, however, congruent stimuli were *slower* than neutral stimuli (a mean difference of 36 ms), $t(47) = 5.34$, $p < .001$, a *reverse* facilitation effect. Importantly, there was no main effect of List Type, $F(1, 47) = .74$, $MSE = 3200$, $p = .394$, $\eta_p^2 = .016$, nor an interaction between Congruency and List Type, $F(2, 94) = 1.48$, $MSE = 819$, $p = .233$, $\eta_p^2 = .031$. The Bayes Factor, $BF_{01} = 4.71 \pm 3.58\%$, suggested "moderate" evidence in support of the absence of the interaction.

The directional hypotheses of smaller effects in the uncorrelated list than in the correlated list were not supported either: for the Stroop effect, $t(47) = 1.68$, $p = .950$, $BF_{0-} = 15.95 \pm .0001\%$; for the facilitation effect, $t(47) = .80$, $p = .787$, $BF_{0-} = 10.75 \pm .0021\%$; for the interference effect, $t(47) = .94$, $p = .824$, $BF_{0-} = 11.54 \pm .0018\%$. Indeed, all three Bayes Factors suggested "strong" evidence for the absence of the hypothesized difference, and numerically, in fact, the differences went in the opposite direction (i.e., the effects were slightly larger, or, in the case of the facilitation effect, more negative, in the correlated list than in the uncorrelated list – see Table 5).

Note, finally, that the high-frequency incongruent stimuli in the correlated list were 11-ms faster than the normal-frequency incongruent stimuli in that list, a difference that was significant, $t(47) = -2.06$, $p = .023$, although the Bayes Factor, $BF_{0-} = .48 \pm .0001\%$, indicated only "anecdotal" evidence for the presence of the effect.

Error rates. There was a main effect of Congruency, $F(1.27, 59.65) = 31.12$, $MSE = .0009$, $p < .001$, $\eta_p^2 = .398$. Post-hoc tests with Holm adjustments for multiple comparisons indicated that, overall, incongruent stimuli produced more errors than both congruent stimuli (a mean difference of 2.30%),

$t(47) = -6.62$, $p < .001$, and neutral stimuli (a mean difference of 2.44%), $t(47) = -7.03$, $p < .001$. That is, there were regular Stroop and interference effects, as expected. There was no (regular or reverse) facilitation effect, however, as congruent stimuli did not differ from neutral stimuli, $t(47) = .40$, $p = .687$. In this case as well, there was no main effect of List Type, $F(1, 47) = .40$, $MSE = .0002$, $p = .529$, $\eta_p^2 = .009$, nor an interaction between Congruency and List Type, $F(1.76, 82.55) = 1.35$, $MSE = .0003$, $p = .263$, $\eta_p^2 = .028$, with the Bayes Factor, $BF_{01} = 4.03 \pm 2.39\%$, suggesting "moderate" evidence in support of the absence of the interaction.

Concerning the directional hypotheses, no significant reduction in the uncorrelated list was observed for either the Stroop effect, $t(47) = .23$, $p = .590$, $BF_{0-} = 7.55 \pm .0056\%$, the facilitation effect, $t(47) = 1.93$, $p = .970$, $BF_{0-} = 17.49 \pm .1239\%$, or the interference effect, $t(47) = -1.20$, $p = .119$, $BF_{0-} = 1.87 \pm .0445\%$, although for the latter effect, the evidence suggested for the absence of the hypothesized difference by the Bayes Factor was only "anecdotal".

Finally, the high-frequency incongruent stimuli in the correlated list produced only slightly fewer errors than the normal-frequency incongruent stimuli in that list, a .21% difference that was not significant, $t(47) = -.52$, $p = .303$, $BF_{0-} = 4.05 \pm 5*10^{-6}\%$.

Discussion

Experiment 1 produced three results worth noting. First and most important, the degree to which colors and words were correlated in the list (and/or the stimulus informativeness was different in the two lists) had no significant impact on performance, and what little numerical impact it did have tended to be in the opposite direction than the hypothesized direction. That is, while input-driven accounts (e.g., Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003) propose that Stroop, facilitation, and interference effects would be smaller in lists in which targets and distractors are uncorrelated (vs. correlated) because attention to the distractor would not be increased in those lists, the general tendency, especially in the response times, was for those effects being slightly *larger* in the uncorrelated list. Consistent with this pattern, the Bayes Factor analyses supported the absence of Stroop differences between uncorrelated and correlated lists in general and certainly so in the hypothesized direction.

Second, the fact that, in the correlated list, high-frequency incongruent stimuli produced faster responses than the other incongruent stimuli in that list suggests that participants were not insensitive to the correlated nature of that list, that is, to the fact that some color-word combinations were more frequent than others. That effect, as noted, might reflect either repetition-priming processes, processes that produce benefits for frequently presented stimuli (e.g., feature binding: Hommel et al., 2004), and/or a process of learning to associate a particular word with its typical color response in the experiment (even if that response happens to be incongruent), a contingency-learning process (Schmidt et al., 2007). Note that the fact that this effect was observed with incongruent stimuli suggests that repetition-priming/contingency-learning processes can be used to reduce the interference that incongruent words produce (see, e.g., Hasshim & Parris, 2021; Hutchison, 2011; Schmidt, 2013c).

A final point concerns the fact that although we obtained regular Stroop and interference effects, the facilitation effect was reversed in the response times, with congruent stimuli being *slower* than neutral stimuli rather than the other way round (the error rates for congruent and neutral stimuli were equivalent). As noted in the Introduction, this pattern is atypical, but not unheard of (e.g., Schulz, 1979; Sichel & Chandler, 1969), at least in normal Stroop experiments (the situation is different in Stroop experiments that require alternating between color-naming and word-reading tasks, for example; for a review of experiments creating these alternative situations, see Kalanthroff et al., 2018).

In order to explain this pattern, it is useful to note that several researchers have suggested that because congruent stimuli (as well as incongruent stimuli) are color names, i.e., readable stimuli, they involve task conflict, that is, a conflict between the instructed color-naming task and the word-reading task, a task that words are strongly associated with (e.g., Goldfarb & Henik, 2007; MacLeod & MacDonald, 2000; Monsell et al., 2001; for an argument that the use of letter-string neutral stimuli in the Stroop task confounds the lexicality of the distractors with their congruency with the targets, see Brown, 2011). Because letter-string neutral stimuli such as XXXX do not involve that conflict, they might be presumed to be faster than congruent stimuli. According to Goldfarb and Henik (2007; see also Kalanthroff et al., 2018), the reason why, normally, this reverse facilitation pattern is not observed is that in most Stroop experiments, color names are more frequent than letter-string neutral stimuli. As a result, task conflict would often be expected and resolved easily when it occurs, preventing reverse facilitation from emerging. The reason why regular facilitation would emerge instead is due to congruent stimuli containing information compatible with the color, which neutral stimuli do not contain.

The situation would be different in a Mostly-Neutral list in which the proportion of (congruent and incongruent) color names is small and the proportion of letter-string neutral stimuli is large. That type of list would induce a relaxation of control over task conflict because that conflict arises only infrequently in that list. As a result, when task conflict does occur (i.e., with the color names), it will have a larger impact. Most notably, response times to congruent stimuli may increase to the point of becoming slower than response times to neutral stimuli, producing reverse facilitation. Goldfarb and Henik (2007; see also Entel et al., 2015) obtained precisely this result in an experiment involving a Mostly-Neutral list of the sort described.

The problem is that the present experiment did not include that type of list (i.e., one involving a large number of neutral stimuli). In fact, both the uncorrelated list and the uncorrelated list in our experiment were lists in which most of the stimuli (80%) were color names and only 20% of the stimuli were letter strings. In that type of list, based on the reasoning explained above, it could be presumed that task conflict would be resolved easily and would have virtually no impact on performance, with the result being either regular facilitation or, potentially, no facilitation. The fact that reverse facilitation emerged instead is somewhat concerning as it seems to suggest that participants experienced significant task conflict in our experiment even though, for the type of lists we used, they were supposed to have that type of conflict well under control.

This fact may potentially undermine the conclusions one can draw from the present experiment concerning the impact of the list's target-distractor correlation on the facilitation effect. Because there was no regular facilitation effect to begin with, a modulation of that effect might not have been possible. We addressed this concern in Experiment 2 by changing the neutral stimuli from the letter string XXXX to a color-unrelated word. It is known that the pattern of facilitation and interference effects may depend on the baseline used, with (regular) facilitation typically being larger when words, rather than letter strings, are used as neutral stimuli (e.g., Brown, 2011; MacLeod, 1991). Further, because a neutral word is just as readable as a color name, there would be no task-conflict difference among the stimuli used (in Brown's (2011) terms, lexicality and congruency would not be confounded). Most notably, because congruent and neutral stimuli would both involve task conflict, the expected pattern of regular facilitation should emerge. That situation would be the most appropriate situation for observing a potential impact of the list's target-distractor correlation on the facilitation component of the Stroop effect.

Experiment 2

Method

Participants

Forty-one students at the University of Western Ontario (33 females and 8 males; age 17-22 years; all native English speakers with normal or corrected-to-normal vision) took part in the experiment for course credit. No participant had participated in the previous experiment, and no participant contributed fewer than 75% of their original observations after discarding too-fast, too-slow, and incorrect responses. Therefore, no participant was excluded.

Materials

The materials were the same as in Experiment 1 with the exception that the neutral letter string was replaced with the word SHORT. We chose this word because it is similar to the color names used (i.e., RED, BLUE, GREEN, and YELLOW) in its dominant part of speech (adjective), length in letters (5 vs. *M* = 4.5 for the color names) and in syllables (1 vs. *M* = 1.25 for the color names), and frequency (87.83 occurrences per million vs. *M* = 91.48 for the color names), but has little orthographic and phonological overlap with the color names (no letters or phonemes in the same positions. This information was extracted from Subtlex-US: Brysbaert & New, 2009; Brysbaert et al., 2012). Other than using a word neutral stimulus rather than a letter-string neutral stimulus, the uncorrelated and correlated lists were similar in all respects to those used in Experiment 1, including the frequencies of color-word combinations, the congruency proportion, the *C* values, and the counterbalancing scheme that was used.

Procedure

The procedure was the same as in Experiment 1. Note that, although the neutral letter string (XXXX) was not included in the experiment proper, it was included in the practice session, which, as in Experiment 1, involved 8 trials in which the letter string was presented in each of the four colors used in the experiment twice.

At the end of this experiment and Experiments 3 and 4, participants were also presented with a questionnaire intended to probe their objective and subjective awareness of the correlation manipulation and of the word-color contingencies in the correlated list. This questionnaire was included for exploratory purposes only and its results have no bearing on the hypotheses being tested here because adjustments of attention in response to a list's target-distractor correlation are not assumed to require participants' awareness of the correlation or of associated phenomena such as contingency learning. Therefore, for the sake of conciseness, we do not include a detailed description of the questionnaire and the results here, which are included in the Supplementary Materials instead. Note, importantly, participants were not told of the existence of the questionnaire at the beginning of the experiment. They received the same instructions as in Experiment 1. Therefore, the existence of an end-of-study questionnaire could not have altered their performance on the Stroop task.

Results

Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than 2000 ms, the time limit (accounting for .7% of the data points), were discarded. The analyses were conducted as in Experiment 1: First, mean response times for correct responses and mean error rates were analyzed with a 3 × 2 ANOVA with Congruency (congruent vs. neutral vs. incongruent) and List Type (uncorrelated vs. correlated) as within-subject factors, and second, one-tailed *t*-tests were conducted to test the hypotheses of smaller Stroop, facilitation, and interference effects in the uncorrelated list vs. the correlated list. Paralleling the procedure in the previous experiments, only data from conditions involving 10 trials (i.e., the normal-frequency stimuli) were included in these analyses. Another set of one-tailed *t*-tests was also conducted to test the hypothesis that the high-frequency incongruent stimuli in the correlated list would be faster and more accurate than the normal-frequency incongruent stimuli in that list. Both NHST and Bayesian analyses were used. The mean RTs and error rates are presented in Table 6.

Table 6

Mean RTs and percentage error rates (and corresponding standard errors) for Experiment 2

	RTs		Error rates	
Stimulus type	Uncorrelated list	Correlated list	Uncorrelated list	Correlated list
Congruent	751 (20)	739 (17)	.61 (.23)	.12 (.09)
Neutral	731 (15)	726 (14)	.37 (.21)	.37 (.14)
Incongruent	817 (16)	810 (16)	2.14 (.43)	2.52 (.56)
High-frequency incongruent		799 (16)		2.13 (.36)
Stroop effect	66	71	1.53	2.40
Facilitation effect	-20	-13	-.24	.25
Interference effect	86	84	1.77	2.15
Stimulus frequency effect		11		.39

RTs. There was a main effect of Congruency, $F(1.69, 67.51) = 69.97$, $MSE = 2817$, $p < .001$, $\eta_p^2 = .636$. Post-hoc tests with Holm adjustments for multiple comparisons indicated that, similar to Experiment 1, overall, incongruent stimuli were slower than both congruent stimuli (a mean difference of 68 ms), $t(40) = -8.98$, $p < .001$, and neutral stimuli (a mean difference of 85 ms), $t(40) = -11.16$, $p < .001$, whereas congruent stimuli were slower than neutral stimuli (a mean difference of 17 ms), $t(40) = 2.17$, $p = .033$. That is, there were regular Stroop and interference effects but a *reverse* facilitation effect, although the size of the latter effect was reduced compared to that in Experiment 1. Importantly, List Type did not have a main effect, $F(1, 40) = 1.53$, $MSE = 2475$, $p = .224$, $\eta_p^2 = .037$, nor did it interact with Congruency, $F(2, 80) = .34$, $MSE = 975$, $p = .716$, $\eta_p^2 = .008$, with the Bayes Factor, $BF_{01} = 8.29 \pm 10.73\%$, suggesting "moderate" evidence in support of the absence of the interaction.

The directional hypotheses of smaller effects in the uncorrelated list than in the correlated list were, again, not supported: for the Stroop effect, $t(40) = -.55$, $p = .294$, $BF_{0-} = 3.68 \pm 5*10^{-6}\%$; for the facilitation effect, $t(40) = -.87$, $p = .196$, $BF_{0-} = 2.61 \pm 5*10^{-5}\%$; for the interference effect, $t(40) = .21$, $p = .583$, $BF_{0-} = 6.92 \pm .0050\%$. The Bayes Factors suggested "moderate" evidence for the absence of the hypothesized difference for the Stroop and interference effects whereas for the facilitation effect, the evidence for the absence of the hypothesized difference was only "anecdotal". Numerically, in fact, the facilitation effect was the effect for which the largest difference in the hypothesized direction was observed (a 7-ms difference). Note, however, that the facilitation effect was reversed in both lists (-20 ms in the uncorrelated list and -13 ms in the correlated list, see Table 6).

Note, finally, that the high-frequency incongruent stimuli in the correlated list were 11-ms faster than the normal-frequency incongruent stimuli in that list. Although this difference is the same size as in the corresponding contrast in Experiment 1, it did not reach significance in this experiment, $t(40) = -1.67$, $p = .051$. The Bayes Factor, however, slightly favored the presence of the effect, $BF_{0-} = .88 \pm 2*10^{-5}\%$.

Error rates. There was a main effect of Congruency, $F(1.17, 46.98) = 23.70$, $MSE = .0008$, $p < .001$, $\eta_p^2 = .372$. Post-hoc tests with Holm adjustments for multiple comparisons indicated that there were regular Stroop and interference effects because, overall, incongruent stimuli produced more errors than both congruent stimuli (a mean difference of 1.96%), $t(40) = -5.97$, $p < .001$, and neutral stimuli (a mean difference of 1.96%), $t(40) = -5.96$, $p < .001$. However, there was no (regular or reverse) facilitation effect, because congruent stimuli produced essentially the same error rate as neutral stimuli, $t(40) = -.01$, $p = .994$. As in the latency data, there was no main effect of List Type, $F(1, 40) = .02$, $MSE = .0004$, $p = .884$, $\eta_p^2 < .001$, nor an interaction between Congruency and List Type, $F(1.32, 52.69) = 1.80$, $MSE = .0003$, $p = .184$, $\eta_p^2 = .043$. The Bayes Factor, $BF_{01} = 2.61 \pm 4.05\%$, suggested only "anecdotal" evidence in support of the absence of the interaction.

Numerically, all the effects did tend to be smaller in the uncorrelated list. The directional hypotheses, however, only supported a significant reduction in the uncorrelated list for the facilitation effect, $t(40) = -1.94$, $p = .030$, $BF_{0-} = .56 \pm 7*10^{-5}\%$ (for the Stroop effect, $t(40) = -1.56$, $p = .064$, $BF_{0-} = 1.05 \pm 4*10^{-5}\%$; for the interference effect, $t(40) = -.76$, $p = .226$, $BF_{0-} = 2.95 \pm 4*10^{-6}\%$). The Bayes Factor values, however, were all "anecdotal", suggesting no strong evidence for either the presence or the absence of the hypothesized differences.

Finally, the high-frequency incongruent stimuli in the correlated list produced only slightly fewer errors than the normal-frequency incongruent stimuli in that list, a .39% difference that was not significant, $t(40) = -.74$, $p = .233$, $BF_{0-} = 3.02 \pm 3*10^{-6}\%$.

Discussion

The results of Experiment 2 were largely in line with those of Experiment 1. First, there was little evidence that the list's target-distractor correlation (and/or the list's stimulus informativeness) had an impact on performance, either at all or in the direction that input-driven accounts hypothesize (i.e., smaller Stroop, facilitation, and interference effects in uncorrelated lists). In this experiment, however, there were numerical tendencies in the hypothesized direction, especially for the facilitation effect, for which a significant difference in the hypothesized direction did emerge in the error rates (i.e., the facilitation effect was smaller in the uncorrelated list than in the correlated list). Note, however, that there was no overall facilitation effect in the error rates, and accuracy for congruent and neutral stimuli was essentially at ceiling, with error rates for those stimuli below 1% in both lists. Further, the Bayes Factor analyses suggested that the evidence in favor of the presence of the hypothesized difference for the facilitation effect in the error rates was weak, as was the evidence in favor of the *absence* of the difference for the other contrasts.

Evidence for repetition-priming/contingency-learning processes was also limited because high-frequency incongruent stimuli were not statistically distinguishable from the other incongruent stimuli in the correlated list. Numerically, however, the former stimuli were 11-ms faster than the latter, the same difference that was observed in Experiment 1 in that contrast. Therefore, it seems likely that participants in this experiment were not insensitive to the nature of the correlated list.

Another similarity with Experiment 1 is the fact that the facilitation effect was reversed in the response times. That is, even though a word rather than a letter string was used to create the neutral stimuli, congruent stimuli still produced slower responses than those stimuli, although the reverse facilitation effect was halved compared to the effect observed in Experiment 1 (17 ms in Experiment 1 vs. 36 ms in Experiment 2). As noted, reverse facilitation is atypical in (normal) Stroop experiments, but when it is observed, it is generally in situations in which nonword neutral stimuli, such as letter strings, are used (for discussions, see Brown, 2011; MacLeod, 1991). When word neutral stimuli are used, the typical result is a facilitation effect which can even be close in size to that of the interference effect (e.g., Brown, 2011).

Regular facilitation when word neutral stimuli are used is, in fact, the pattern that accounts of the reverse facilitation effect would predict (Goldfarb & Henik, 2007; Kalanthroff et al., 2018). According to those accounts, reverse facilitation would represent a special case emerging only when both of the following are true: 1) the neutral stimuli being contrasted with the congruent stimuli are stimuli not strongly associated with the reading task, that is, when stimuli, such as XXXX, which do not involve task conflict are used, and 2) a situation is created in which control over task conflict is reduced, for example, with a Mostly-Neutral list in which the proportion of (congruent and incongruent) color names is small and the proportion of letter-string neutral stimuli is large (although that this type of situation does not need to be created experimentally and may also be observed in individuals with lower control abilities instead, see Kalanthroff et al., 2018, for a review).

Neither of those conditions applied in the present Experiment 2 because 1) word neutral stimuli were used, stimuli that, presumably, involved task conflict just as congruent stimuli do, and 2) *all* of the stimuli in the experiment were readable stimuli, that is, stimuli involving task conflict. Therefore, control over task conflict in our experiment should have been as efficient as it could ever be, and no reverse facilitation effect should have emerged. In the General Discussion, we propose an explanation for why a reverse facilitation effect did emerge in our Experiments 1 and 2. However, because this explanation is not one that is readily available in the literature, in the following experiments we decided to side-step the issue and to focus on the Stroop effect without decomposing it into facilitation and interference, that is, without including neutral stimuli at all.

Experiment 3

In this experiment, because of the difficulties associated with the reverse facilitation effect that we obtained in Experiments 1 and 2, we used the same design as in those experiments but we did not include neutral stimuli (the stimuli that can allow us to compute facilitation as well as interference). That is, both the uncorrelated list and the correlated list only included congruent and incongruent stimuli and the only effect that could computed in the two lists was the Stroop effect. Note that, even though neutral stimuli were not included, the impact of the list's target-distractor correlation on the Stroop effect should be the same as in the experiments that did include those stimuli: The effect should be smaller in the uncorrelated list, the list in which attention to the word should not be increased, than in the correlated list, the list in which attention to the word should be increased. In fact, the elimination of the neutral stimuli from the design had the side effect of making the correlation manipulation slightly stronger because, while *C* in the uncorrelated list remained zero (and the list remained uninformative because there were still no contingencies to learn), its absolute value increased to .58 (from .53 in Experiments 1 and 2) in the correlated list (and the list became more informative because the one distractor for which no contingencies could be learned, i.e., the neutral distractor which was equally associated with all targets in the previous experiments, was eliminated in this experiment). Therefore, according to input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003; see also Schmidt, 2014, 2019), there is even more reason to expect a difference in the magnitude of the Stroop effect between the two lists.

Method

<u>Participants</u>

Forty-eight students at the University of Western Ontario (29 females and 19 males; age 18-31 years; all native English speakers with normal or corrected-to-normal vision) took part in the experiment for course credit. No participant had participated in the previous experiments, and no participant contributed fewer than 75% of their original observations after discarding too-fast, too-slow, and incorrect responses. Therefore, no participant was excluded.

<u>Materials</u>

The materials were the same as in the previous experiments except that no neutral letter string or word was used. That is, the only words used were the four color names RED, BLUE, GREEN, and YELLOW. The frequencies of color-word combinations were adjusted as represented in Table 7. In the uncorrelated list, represented on the left side of Table 7, each of the words used was presented with each of the colors used 12 times (the normal frequency in this experiment). In the correlated list, represented on the right side of Table 7, each of the words was presented with the congruent color 12 times, with one of the three incongruent colors 12 times, with another of the three incongruent colors 24 times (the high frequency in this experiment), and never with the third incongruent color. The assignment of the specific incongruent colors used to form combinations with a frequency of 12, 24, or zero was counterbalanced across participants, as was the order with which the correlated and uncorrelated lists were presented to participants. Both lists contained 192 stimuli, of which 144 were incongruent (75%) and 48 congruent (25%). As noted, *C* for the uncorrelated list was zero whereas it was -.58 for the correlated list.

Table 7

Template for the frequency of color-word combinations in Experiment 3

	List type							
	Uncorrelated				Correlated			
Color	RED	BLUE	GREEN	YELLOW	RED	BLUE	GREEN	YELLOW
Red	12	12	12	12	12		24	12
Blue	12	12	12	12	12	12		24
Green	12	12	12	12	24	12	12	
Yellow	12	12	12	12		24	12	12

Note. In the uncorrelated list, the proportion of incongruent stimuli is 75% and *C* is zero. In the correlated list, the proportion of incongruent stimuli is also 75% and *C* is -.58. The high-frequency color-word combinations (i.e., the color-word combinations occurring 24 times) in the correlated list were not used in the contrast between the two lists.

Procedure

The procedure was the same as in Experiment 2. Note that the practice session, in this case as well, involved 8 trials in which the neutral letter string (XXXX) was presented in each of the colors used in the experiment twice.

Results

Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than 2000 ms, the time limit (accounting for 1.1% of the data points), were discarded. The analyses were conducted as in the previous experiments, except for the fact that the Congruency factor included only two levels in this experiment: congruent vs. incongruent. Thus, mean response times for correct responses and mean error rates were first analyzed with a 2 × 2 ANOVA with Congruency (congruent vs. incongruent) and List Type (uncorrelated vs. correlated) as within-subject factors. Second, one-tailed *t*-tests were conducted to test the hypothesis of smaller Stroop effects in the uncorrelated list vs. the correlated list. Paralleling the procedure in the previous experiments, only data from conditions involving 12 trials (i.e., the normal-frequency stimuli) were included in these analyses. Another set of one-tailed *t*-tests was also conducted to test the hypothesis that the high-frequency incongruent stimuli in the correlated list would be faster and more accurate than the normal-frequency incongruent stimuli in that list. Again, both NHST and Bayesian analyses were used. The mean RTs and error rates are presented in Table 8.

Table 8

Mean RTs and percentage error rates (and corresponding standard errors) for Experiment 3

	RTs		Error rates	
Stimulus type	Uncorrelated list	Correlated list	Uncorrelated list	Correlated list
Congruent	771 (19)	770 (17)	.40 (.17)	.31 (.11)
Incongruent	835 (17)	841 (18)	2.06 (.44)	1.71 (.37)
High-frequency incongruent		826 (15)		1.77 (.37)
Stroop effect	66	71	1.66	1.40
Stimulus frequency effect		15		-.06

RTs. There was a main effect of Congruency, $F(1, 47) = 96.53$, $MSE = 2263$, $p < .001$, $\eta_p^2 = .673$, indicating faster responses to congruent than incongruent stimuli – the regular Stroop effect. There was no main effect of List Type, however, $F(1, 47) = .09$, $MSE = 2880$, $p = .770$, $\eta_p^2 = .002$, and no interaction between Congruency and List Type, $F(1, 47) = 1.22$, $MSE = 553$, $p = .275$, $\eta_p^2 = .025$. With JASP's default settings, the Bayes Factor, $BF_{01} = 3.45 \pm 31.47\%$, suggested "moderate" evidence in support of the absence of the interaction. However, because an error percentage above 20% for a Bayes Factor should not be deemed acceptable according to the JASP guidelines (van Doorn et al., 2020), we re-ran the analysis specifying the number of samples for the computation of the Bayes Factor as 100000 in the "numerical accuracy" option in JASP. The result was $BF_{01} = 2.71 \pm 11.07\%$, with the absence of the interaction still being favored although the evidence for the absence was only "anecdotal".

Further, although the Stroop effect was numerically smaller in the uncorrelated list (66 ms) than in the correlated list (71 ms), the directional hypothesis of a smaller Stroop effect in the correlated list was not supported, $t(47) = -1.10$, $p = .138$, with the Bayes Factor, $BF_{0-} = 2.10 \pm .0215\%$, suggesting "anecdotal" evidence in support of the absence of the hypothesized difference.

Note, finally, that the high-frequency incongruent stimuli in the correlated list were 15-ms faster than the normal-frequency incongruent stimuli in that list, a significant difference, $t(47) = -3.45$, $p < .001$, $BF_{0-} = .02 \pm 6*10^{-5}\%$.

Error rates. There was a main effect of Congruency, $F(1, 47) = 17.58$, $MSE = .0006$, $p < .001$, $\eta_p^2 = .272$, indicating fewer errors to congruent than incongruent stimuli. Neither the main effect of List Type, $F(1, 47) = 1.79$, $MSE = .0001$, $p = .188$, $\eta_p^2 = .037$, nor the interaction between Congruency and List Type, $F(1,$

47) = .49, *MSE* = .0002, *p* = .488, η_p^2 = .010, were significant.[7] The Bayes Factor, BF_{01} = 3.70 ± 4.66%, suggested “moderate” evidence in support of the absence of the interaction.

In fact, the Stroop effect was numerically larger in the uncorrelated list (1.66%) than in the correlated list (1.40%). Not surprisingly, then, the directional hypothesis of a smaller Stroop effect in the correlated list was not supported, *t*(47) = .70, *p* = .756, with the Bayes Factor, BF_{0-} = 10.14 ± .0327%, suggesting “strong” evidence in support of the absence of the hypothesized difference.

Note, finally, that the high-frequency incongruent stimuli in the correlated list produced essentially the same error rate as the normal-frequency incongruent stimuli in that list, *t*(47) = .20, *p* = .580, BF_{0-} = 7.41 ± .0072%.

Discussion

The results of Experiment 3 were straightforward. Once again, we did not find a significant difference in the magnitude of the Stroop effect between the uncorrelated list and the correlated list. Numerically,

[7] In the analysis with Order (correlated list first vs. uncorrelated list first) as an additional factor, the error rates produced a significant three-way interaction between Congruency, List Type, and Order, *F*(1, 46) = 6.69, *MSE* = .0002, *p* = .013, η_p^2 = .127. This interaction reflected the fact that, while the Stroop effect was larger in the uncorrelated list (2.72%) than in the correlated list (1.55%) in the correlated-list-first group, *F*(1, 23) = 10.68, *MSE* = .0001, *p* = .003, η_p^2 = .317, the Stroop effect was smaller in the uncorrelated list (.61%) than in the correlated list (1.26%) in the uncorrelated-list-first group, albeit only numerically, *F*(1, 23) = 1.15, *MSE* = .0002, *p* = .295, η_p^2 = .048. Note, however, that there was also a general tendency for larger Stroop effects in the list that was presented second (1.99%) than in the list that was presented first (1.07%), *F*(1, 46) = 2.89, *MSE* = .0001, *p* = .096, η_p^2 = .059, presumably a fatigue effect. Therefore, we conducted an additional analysis on the data from the list presented first using Congruency as a within-subject factor and List Type as a between-subject factor. The results showed no significant difference between the Stroop effects produced by the two lists, *F*(1, 46) = 1.58, *MSE* = .0003, *p* = .216, η_p^2 = .033.

the Stroop effect was slightly smaller in the uncorrelated (vs. correlated) list in the response times, as hypothesized, but slightly larger in that list in the error rates. In any case, the Bayes Factor analyses supported the absence of a difference, although more weakly so for the response times than for the error rates. Note, further, that in the correlated list, high-frequency incongruent stimuli were significantly faster than the other incongruent stimuli in that list, suggesting that participants engaged the repetition-priming/contingency-learning processes that that list enabled.

The main point is that even when the examination is focused solely on the Stroop effect, there is no evidence for this effect being modulated by the degree to which colors and words are correlated in a list (and/or the stimulus informativeness of the two lists).

Experiment 4

In this experiment, we created a stronger test of the idea that the target-distractor correlation in a list of stimuli has an important role in how attention is adjusted. One argument that could be made with respect to the previous experiments is that the reason that we were not able to detect an impact of the list's target-distractor correlation is that that correlation did not differ enough between the two lists. Note that proposing this argument means being generous to input-driven accounts, accounts that have proposed that even 50:50 congruent/incongruent lists such as that represented in Table 1 would involve a sizeable correlation, sizeable enough that, despite the matching proportion of congruent and incongruent stimuli, the situation created should not be considered “neutral by any means” (Algom et al., 2022, p. 888). In that type of list, the absolute value of *C* is .5. In the correlated lists that we used in our experiments the absolute value of *C* was .53 in Experiments 1 and 2 and .58 in Experiment 3. Because those lists were contrasted with lists with a *C* of zero (the “neutral” situation, according to input-driven accounts), one would presume that a difference would emerge in that contrast.

In this experiment, we made the contrast even larger by increasing the stimulus set size from four colors and the corresponding color names to six (in this experiment as in Experiment 3, neutral stimuli were not included). Doing so allowed us to maintain an uncorrelated list with a *C* of zero, represented on the left side of Table 9, by randomly pairing the six colors and the six words used. However, the target-distractor correlation for the correlated list, represented on the right side of Table 9, was larger than in the previous experiments because each word (e.g., RED) was combined with one incongruent color (e.g., green) *four times* as frequently as the other incongruent color with which it was combined (e.g., blue) whereas it was never combined with the other *three* incongruent colors. In contrast, in the previous experiments, each color name was only combined with one incongruent color *twice* as frequently as the other incongruent color with which it was combined and there was only *one* incongruent color with which it was never combined. Essentially, in the correlated list in the present experiment, participants had a much better chance of guessing the colors with which a word would appear than they did in the correlated lists in the previous experiments. This situation is reflected in a large absolute value of *C*, .82, for the correlated list in the present experiment. Note that this value is close to the largest value for the experiments reported in Melara and Algom's (2003) re-analysis, the value in Glaser and Glaser's (1982) Experiment 2 noted above (*C* = .83). Further, because the contingency strength difference in the correlated list was increased in the present experiment (i.e., 20 to 5) compared to the previous experiments (i.e., 20 to 10 for Experiments 1 and 2 and 24 to 12 in Experiment 3), it could be presumed that the stimulus informativeness in the correlated list was also increased.

Note, finally, that although the proportion of incongruent stimuli in this experiment inevitably increased (due to there being more incongruent stimuli because of the use of larger stimulus sets), congruency proportion was matched across the two lists, as was the frequency of the stimuli being contrasted (i.e., the stimuli in Table 9 with a frequency of 5), as in the previous experiments. By contrasting a zero-correlation list with a list with a very high correlation while controlling for all relevant variables, the present experiment would thus seem to create the strongest test of the impact of target-distractor correlation on performance in the Stroop task thus far.

Method

Participants

Fifty students at the University of Western Ontario (37 females and 13 males; age 18-21 years; all native English speakers with normal or corrected-to-normal vision) took part in the experiment for course credit. No participant had participated in the previous experiments, and no participant contributed fewer than 75% of their original observations after discarding too-fast, too-slow, and incorrect responses. Therefore, no participant was excluded.

Materials

The materials were the same as in Experiment 3 with the addition of two color names (BLACK and WHITE) and the corresponding colors (black [R: 0; G: 0; B: 0] and white [R: 255; G: 255; B: 255]). The frequencies of color-word combinations were also adjusted as represented in Table 9. In the uncorrelated list, represented on the left side of Table 9, each of the words used was presented with each of the colors used 5 times (the normal frequency in this experiment). In the correlated list, represented on the right side of Table 9, each of the words was presented with the congruent color 5 times, with one of the five incongruent colors 5 times, with another of the five incongruent colors 20 times (the high frequency in this experiment), and never with the other three incongruent colors. Similar to the previous experiments, the assignment of the specific incongruent colors used to form combinations with a frequency of 5, 20, or zero was counterbalanced across participants, as was the order with which the correlated and uncorrelated lists were presented to participants. Both lists contained 180 stimuli, of which 150 were incongruent (83.33%) and 30 congruent (16.67%). As noted, C for the uncorrelated list was zero whereas it was -.82 for the correlated list.

Table 9

Template for the frequency of color-word combinations in Experiment 4

	List type											
	Uncorrelated						Correlated					
Color	RED	BLUE	GREEN	YELLOW	BLACK	WHITE	RED	BLUE	GREEN	YELLOW	BLACK	WHITE
Red	5	5	5	5	5	5	5				20	5
Blue	5	5	5	5	5	5	5	5				20
Green	5	5	5	5	5	5	20	5	5			
Yellow	5	5	5	5	5	5		20	5	5		
Black	5	5	5	5	5	5			20	5	5	
White	5	5	5	5	5	5				20	5	5

Note. In the uncorrelated list, the proportion of incongruent stimuli is 83.33% and *C* is zero. In the correlated list, the proportion of incongruent stimuli is also 83.33% and *C* is -.82. The high-frequency color-word combinations (i.e., the color-word combinations occurring 20 times) in the correlated list were not used in the contrast between the two lists.

Procedure

The procedure was the same as in Experiments 2 and 3, except that the practice session included 12 trials, instead of 8, in which the letter string “XXXX” was presented in each of the six colors twice.

Results

Prior to the analyses, invalid trials due to technical failures and responses faster than 300 ms or slower than 2000 ms, the time limit (accounting for 1.6% of the data points), were discarded. The analyses were conducted as in Experiment 3 (involving data from only the normal-frequency stimuli, except for the contrast between high-frequency and normal-frequency incongruent stimuli in the correlated list). The mean RTs and error rates are presented in Table 10.

Table 10

Mean RTs and percentage error rates (and corresponding standard errors) for Experiment 4

	RTs		Error rates	
Stimulus type	Uncorrelated list	Correlated list	Uncorrelated list	Correlated list
Congruent	772 (17)	781 (18)	.47 (.19)	.40 (.16)
Incongruent	818 (16)	827 (17)	2.09 (.31)	2.63 (.42)
High-frequency incongruent		811 (17)		2.28 (.31)
Stroop effect	46	46	1.62	2.23
Stimulus frequency effect		16		.35

RTs. There was a main effect of Congruency, $F(1, 49) = 56.38$, $MSE = 1903$, $p < .001$, $\eta_p^2 = .535$, indicating a regular Stroop effect, with faster responses to congruent than incongruent stimuli. There was no main effect of List Type, however, $F(1, 49) = 1.20$, $MSE = 3094$, $p = .279$, $\eta_p^2 = .024$, and no interaction between Congruency and List Type, $F(1, 49) = 4*10^{-9}$, $MSE = 1112$, $p = 1$, $\eta_p^2 < .001$, with the Bayes Factor, BF_{01} = 5.91 ± 12.43%, suggesting "moderate" evidence in support of the absence of the interaction.

The directional hypothesis of a smaller Stroop effect in the correlated list was not supported either, $t(49) = -6*10^{-5}$, $p = .500$, with the Bayes Factor, BF_{0-} = 6.50 ± .0344%, suggesting "moderate" evidence in support of the absence of the hypothesized difference. In fact, the size of the Stroop effect in the two lists was identical (46 ms).

Note, finally, that the high-frequency incongruent stimuli in the correlated list were 16-ms faster than the normal-frequency incongruent stimuli in that list, a significant difference, $t(49) = -3.02$, $p = .002$, BF_{0-} = .06 ± .0005%.

Error rates. There was a main effect of Congruency, $F(1, 49) = 38.91$, $MSE = .0005$, $p < .001$, $\eta_p^2 = .443$, indicating, once again, a regular Stroop effect, with fewer errors to congruent than incongruent stimuli. Neither the main effect of List Type, $F(1, 49) = 1.38$, $MSE = .0002$, $p = .247$, $\eta_p^2 = .027$, nor the interaction between Congruency and List Type, $F(1, 49) = 1.04$, $MSE = .0004$, $p = .314$, $\eta_p^2 = .021$, were significant, with the Bayes Factor, BF_{01} = 2.23 ± 3.92%, suggesting "anecdotal" evidence in support of the absence of the interaction.

Further, although the Stroop effect was numerically smaller in the uncorrelated list (1.63%) than in the correlated list (2.23%), the directional hypothesis of a smaller Stroop effect in the correlated list was not supported, $t(49) = -1.02$, $p = .157$, with the Bayes Factor, BF_{0-} = 2.38 ± $4*10^{-5}$%, suggesting "anecdotal" evidence in support of the absence of the hypothesized difference.

Note, finally, that the high-frequency incongruent stimuli in the correlated list produced only a slightly lower error rate than the normal-frequency incongruent stimuli in that list, a difference of .35% that was not significant, $t(49) = -.71$, $p = .242$, $BF_{0-} = 3.41 \pm 3*10^{-6}\%$.

Discussion

Despite the strong manipulation of the list's target-distractor correlation (and stimulus informativeness), we once again failed to obtain evidence in support of an impact of that manipulation on the Stroop effect. In fact, the Stroop effect was identical in the two lists in the response times, with Bayes Factor analyses supporting the absence of a difference, either in the hypothesized direction or at all. Also similar to the previous experiments, high-frequency incongruent stimuli in the correlated list were faster than normal-frequency incongruent stimuli in the same list, suggesting that participants were not insensitive to the correlated nature of that list. Nevertheless, inconsistent with input-driven accounts (Algom & Chajut, 2019; Algom et al., 2022; Melara and Algom, 2003; see also Schmidt, 2014, 2019), there was no evidence to suggest that, in that list, attention to the word was increased.

Combined analysis of all experiments

One potential criticism of the experiments reported thus far is that the effect that the list's target-distractor correlation has on performance in the Stroop task may be much smaller than originally hypothesized, and our experiments individually did not have enough power to detect an effect of that size. In order to address this issue, we combined mean response times for correct responses and mean error rates for congruent and (normal-frequency) incongruent stimuli in all our experiments ($N = 187$) and analyzed them in a similar way as was done for Experiments 3 and 4 (note that we focused on congruent and incongruent stimuli because they were the types of stimuli that were common to all of the experiments). That is, the data were first analyzed with a 2 × 2 ANOVA with Congruency (congruent vs. incongruent) and List Type (uncorrelated vs. correlated) as within-subject factors, and second, one-tailed *t*-tests were conducted to test the hypothesis of smaller Stroop effects in uncorrelated lists vs. the correlated lists. Because in the analyses of the previous experiments high-frequency incongruent stimuli in the correlated list were consistently faster than normal-frequency incongruent stimuli in the same list, we did not include that contrast in this analysis. However, we report the mean RTs and error rates for high-frequency incongruent stimuli in the correlated list, as well as for the other conditions, in Table 11.

Table 11

Mean RTs and percentage error rates (and corresponding standard errors) for all experiments combined

	RTs		Error rates	
Stimulus type	Uncorrelated list	Correlated list	Uncorrelated list	Correlated list
Congruent	763 (9)	766 (9)	.44 (.09)	.38 (.08)
Incongruent	828 (9)	830 (9)	2.24 (.20)	2.44 (.23)
High-frequency incongruent		817 (8)		2.22 (.18)
Stroop effect	65	64	1.80	2.06
Stimulus frequency effect		13		.22

RTs. Not surprisingly, the main effect of Congruency was significant, $F(1, 186) = 310.62$, $MSE = 2413$, $p < .001$, $\eta_p^2 = .626$, with faster responses to congruent than incongruent stimuli. More importantly, the main effect of List Type, $F(1, 186) = .36$, $MSE = 2632$, $p = .551$, $\eta_p^2 = .002$, and the interaction between Congruency and List Type, $F(1, 186) = .01$, $MSE = 903$, $p = .912$, $\eta_p^2 < .001$, remained non-significant.[8] With JASP's default settings, the Bayes Factor, $BF_{01} = 12.73 \pm 24.39\%$, suggested "strong" evidence in support of the absence of the interaction. However, because the error percentage was above 20%, we re-ran the analysis specifying a number of samples of 100000 in the "numerical accuracy" option in JASP. The result was $BF_{01} = 8.64 \pm 2.36\%$, with the absence of the interaction still being favored although the evidence for the absence was "moderate".

The directional hypothesis of smaller Stroop effects in the correlated list was not supported either, $t(186) = .11$, $p = .544$, with the Bayes Factor, $BF_{0-} = 13.34 \pm .0656\%$, suggesting "strong" evidence in

[8] In order to account for potential differences across experiments and effects of time on task, this analysis was repeated with Experiment (1 vs. 2 vs. 3 vs. 4) and Order (correlated list first vs. uncorrelated list first) as additional between-subject factors. The results were similar. Most importantly, the interaction between Congruency and List Type was not modified by either Experiment or Order (*p*s > .25 for the two three-way interactions and the four-way interaction). Interestingly, the two-way interaction between Congruency and Experiment was significant, $F(3, 179) = 3.49$, $MSE = 2458$, $p = .017$, $\eta_p^2 = .055$. This interaction was examined with an ANOVA with Experiment and Order as independent variables and mean Stroop effects as the dependent variable, followed up with post-hoc *t*-tests with the Holm correction for multiple comparisons contrasting mean Stroop effects for each pair of experiments. Those tests revealed that the main source of the interaction was that the Stroop effect in Experiment 4 was smaller than in Experiment 1, $t(186) = 3.12$, $p = .013$ (all other *p*s > .15, although note that, numerically, the Stroop effect in Experiment 4 was smaller than in Experiments 2 and 3 as well). A potential reason for this result is that the proportion of incongruent stimuli was higher in Experiment 4 (in which it was 83.33%) than in the other experiments (in which it was 75% or lower).

support of the absence of the hypothesized difference. In fact, overall, the size of the Stroop effect in the two lists was almost identical (65 ms and 64 ms in uncorrelated and correlated lists, respectively).

Error rates. Again, there was a main effect of Congruency, $F(1, 186) = 108.59$, $MSE = .0006$, $p < .001$, $\eta_p^2 = .369$, indicating fewer errors to congruent than incongruent stimuli. Similar to the RTs, the main effect of List Type, $F(1, 186) = .35$, $MSE = .0002$, $p = .554$, $\eta_p^2 = .002$, and the interaction between Congruency and List Type, $F(1, 186) = 1.11$, $MSE = .0003$, $p = .293$, $\eta_p^2 = .006$, were not significant.[9] Further, the Bayes Factor, $BF_{01} = 4.59 \pm 3.50\%$, suggested "moderate" evidence in support of the absence of the interaction.

---

[9] We also repeated the error analysis with Experiment and Order as additional between-subject factors. In this case, there was a three-way interaction between Congruency, List Type, and Order, $F(1, 179) = 4.50$, $MSE = .0003$, $p = .035$, $\eta_p^2 = .025$ (but there was neither a three-way interaction between Congruency, List Type, and Experiment nor the four-way interaction, both *p*s > .30 – note also that, in this case, Experiment did not interact with Congruency, $F(3, 179) = .71$, $MSE = .0006$, $p = .546$, $\eta_p^2 = .012$). The three-way interaction between Congruency, List Type, and Order reflected the fact that, while the Stroop effect was larger in uncorrelated lists (2.26%) than in correlated lists (1.99%) in the correlated-list-first group, albeit only numerically, $F(1, 88) = .61$, $MSE = .0002$, $p = .437$, $\eta_p^2 = .007$, the Stroop effect was smaller in the uncorrelated list (1.34%) than in the correlated list (2.12%) in the uncorrelated-list-first group, $F(1, 91) = 4.60$, $MSE = .0003$, $p = .035$, $\eta_p^2 = .048$. Although the latter result is consistent with the hypothesis of larger Stroop effects in correlated than uncorrelated lists, it may be partially due to the fatigue effects that emerged in our experiments (see footnotes 6 and 7), reflecting a general tendency for larger Stroop effects in the list that was presented second (in the uncorrelated-list-first group, the correlated list) than in the list that was presented first (in the uncorrelated-list-first group, the uncorrelated list), $F(1, 179) = 3.47$, $MSE = .0002$, $p = .064$, $\eta_p^2 = .019$. Therefore, we conducted an additional analysis on the data from the list presented first using Congruency as a within-subject factor and List Type and Experiment as between-subject factors. The results showed, once again, no significant difference between the Stroop effects produced by the two lists, $F(1, 179) = 2.23$, $MSE = .0004$, $p = .137$, $\eta_p^2 = .012$.

Further, although the Stroop effect was numerically smaller in the uncorrelated list (1.80%) than in the correlated list (2.06%), the directional hypothesis of a smaller Stroop effect in the correlated list was not supported, $t(186) = -1.06$, $p = .146$, with the Bayes Factor, $BF_{0-} = 4.16 \pm .0060\%$, suggesting “moderate” evidence in support of the absence of the hypothesized difference.

In sum, even combining the data from all our experiments did not produce evidence in support of an impact of the list’s target-distractor correlation on performance in the Stroop task. In fact, Bayes Factor analyses supported the conclusion that there is an absence of Stroop effect differences between uncorrelated and correlated lists not only in the hypothesized direction but also in general.

General Discussion

According to input-driven accounts of the Stroop effect (Algom & Chajut, 2019; Algom et al., 2022; Melara & Algom, 2003), in lists of stimuli in which targets and distractors are correlated (i.e., “correlated lists”), attention to distractors would be increased, resulting in a larger benefit produced by congruent stimuli compared to neutral stimuli (i.e., larger facilitation) but also in a larger cost produced by incongruent stimuli compared to neutral stimuli (i.e., larger interference), with respect to a list of stimuli in which targets and distractors are not correlated as strongly or are completely uncorrelated (i.e., “uncorrelated lists”). Because the Stroop effect combines facilitation and interference, a larger Stroop effect would also emerge in correlated lists. Proponents of input-driven accounts argued that such a process of attentional adjustment to the list’s target-distractor correlation would be lurking in many Stroop experiments, particularly in Proportion-Congruent (PC) manipulations, manipulations in which the proportion of congruent and incongruent stimuli in a list is manipulated. Indeed, according to Algom et al. (2022), target-distractor “correlation is a major determinant of the magnitude of the Stroop effect across the vast Stroop literature” (p. 888).

The results of the present experiments suggest that this idea and, therefore, the quoted statement, are likely incorrect. As noted in the Introduction, the evidence that has been brought in support of the idea that a larger target-distractor correlation increases Stroop effects comes from experiments (e.g., Dishon-Berkovits, 2000) and re-analyses (Melara & Algom, 2003) in which the list's target-distractor correlation and the list's congruency proportion tended to be confounded, with uncorrelated lists often being lists with a large proportion of incongruent stimuli and correlated lists often being lists with a small proportion of incongruent stimuli.

The reason this fact is relevant is that, in recent years, the list's congruency proportion has been shown to have an impact on the magnitude of the Stroop effect independently from the processes that, in standard PC manipulations, tend to be confounded with congruency proportion, including repetition-priming/contingency-learning processes and attentional adjustments to the target-distractor correlation itself (e.g., Bugg, 2014; Hutchison, 2011; Spinelli & Lupker, 2022b; Spinelli et al., 2019). That is, for stimuli for which those processes are controlled, Stroop effects are smaller in lists in which incongruent stimuli are frequent overall (i.e., Mostly-Incongruent lists) than in lists in which incongruent stimuli are infrequent overall (i.e., Mostly-Congruent lists). Control accounts (e.g., Botvinick et al., 2001; Braver, 2012) interpret this Proportion-Congruent effect as reflecting a process of attentional adjustment to the frequency with which distractors produce conflict in the list.

Similar control explanations have been offered for the closely associated Proportion-*Neutral* effect, an effect based on the contrast between an MI list composed of many incongruent and few non-readable neutral stimuli (i.e., a list in which most of the stimuli produce task conflict) and a Mostly-*Neutral* list composed of few incongruent and many unreadable neutral stimuli (i.e., a list in which most of the stimuli do not produce task conflict), with the latter list producing larger Stroop interference than the former (e.g., Entel et al., 2015; Goldfarb & Henik, 2007; Tzelgov et al., 1992) even when relevant confounds are controlled for (Spinelli & Lupker, 2021). That is, the Proportion-Neutral effect has been argued to reflect, among other things, increased control over task conflict in situations in which that type of conflict occurs more frequently (Goldfarb & Henik, 2007).

The important point, in any case, is that *this* type of control process may be the one lurking in the data that have been presumed to demonstrate an impact of the target-distractor correlation because, as noted, congruency proportion and target-distractor correlation tended to be confounded in those experiments. Therefore, in order to make a credible case for the list's target-distractor correlation being a major determinant of Stroop effects, that factor needs to be dissociated from the list's congruency proportion, as well as from the repetition-priming/contingency-learning processes that have been known to influence performance in the Stroop task. The reason the latter dissociation is also relevant is that target-distractor correlation and processes such as contingency learning, though related, are conceptually distinct notions: The former is an overall characteristic of a list of trials assumed to induce increased attention to distractors (and, hence, increased Stroop effects) when the correlation is larger than zero; the latter is a process of learning not to regulate attention but to associate individual distractors with their typical responses, either congruent or incongruent (see footnote 1).

In the present experiments using the classic color-word Stroop task, we created such dissociations by contrasting uncorrelated lists in which the targets (i.e., the colors) and the distractors (i.e., the words) used were randomly paired with correlated lists in which the targets and the distractors used were not randomly paired. Specifically, for each of the color names used, one color was more frequent than the other colors in which that color name appeared, and some colors never appeared with that color name. The value of *C*, a chi-square based correlation (Melara & Algom, 2003), in these lists varied from .53 to .82 in absolute value, with *C* being zero in the uncorrelated lists. Importantly, both congruency proportion and repetition priming/contingency learning were controlled in the contrast between the two lists because 1) the proportion of congruent, incongruent, and neutral stimuli (in the experiments in which we used neutral stimuli) were the same in the two lists, and 2) the stimuli being used for the contrast had the same frequency in the two lists, meaning that repetition-priming/contingency-learning processes would have the same impact for those stimuli. As a result, any Stroop effect difference between the two lists would have to be attributed to the list's target-distractor correlation, or closely associated notions such as stimulus informativeness (Schmidt, 2014, 2019).

In our experiments, uncorrelated lists were "uninformative" because there were no word-response contingencies that participants could learn in those lists. In contrast, correlated lists were "informative" because participants could learn word-response contingencies in those lists. Therefore, according to Schmidt, attention to words in the latter lists should increase and, as a result, so would Stroop, facilitation, and interference effects, the same pattern of results that would be hypothesized assuming a process of attentional adjustment to the list's target-distractor correlation.

In four experiments, however, virtually no evidence emerged in support of this hypothesis. In Experiments 1 and 2, experiments in which the Stroop effect was decomposed into facilitation and interference, with one exception, no reduction in the uncorrelated list was observed for either Stroop, facilitation, or interference effects. The exception was the facilitation effect in the error rates in Experiment 2, which was found to be smaller in the uncorrelated list than in the correlated list. That effect, however, was observed in the presence of a null facilitation effect overall and very low (i.e., below 1%) error rates for congruent and neutral stimuli. Further, the Bayes Factor analyses that we conducted suggested evidence, albeit weak, for the absence of that difference, and a numerical difference in the opposite direction (i.e., a larger facilitation effect in the uncorrelated list than in the correlated list) was observed for the error rates in Experiment 1. Therefore, that significant result from Experiment 2 does not appear to be particularly reliable and the most likely conclusion is that the target-distractor correlation had no impact on the facilitation effect (a conclusion that would be consistent with the results of Hasshim and Parris's (2021) Experiment 2 discussed in the Introduction). As for the other effects, the Bayes Factor analyses showed consistent support for the absence of the hypothesized difference (or any difference) between the two lists.

Note, however, that Experiments 1 and 2 showed an unexpected pattern concerning the expected facilitation in the response times. Facilitation was reversed, that is, congruent stimuli were slower than neutral stimuli rather than the, more typical, opposite pattern. Interestingly, this pattern emerged both when a letter string was used to create the neutral stimuli (Experiment 1) and when a color-unrelated word was used (Experiment 2). In general, although reverse facilitation effects have been reported under certain circumstances (for a review, see Kalanthroff et al., 2018), those effects would not be expected to occur under the present circumstances. We explain this contradiction and a potential reason for it below, in the section "Reverse facilitation: exception or norm?". Given this situation, the point is that a modulation of facilitation effects in the response times in Experiment 1 and 2 might not have been possible because there was no regular facilitation to begin with.

With this consideration in mind, we did not include neutral stimuli in Experiments 3 and 4 in order to focus only on the Stroop effect. The main pattern of Experiments 1 and 2 was replicated: the Stroop effect was the same size in the two lists, with Bayes Factor analyses again favoring the absence of the hypothesized difference (or any difference) between the two lists. Notably, this pattern emerged even in Experiment 4 despite the fact that the correlation manipulation in that experiment was quite strong. In fact, in that experiment, the Stroop effect in the response times not only was the smallest in our experiments but it was also identical in the two lists. The same pattern of results emerged in a combined analysis of all four experiments, addressing the potential argument that our experiments individually did not have enough power to detect the impact of the target-distractor correlation on the Stroop effect.

Overall, these results clearly make the case that the list's target-distractor correlation is not a "major determinant" of Stroop effects: Whether targets and distractors in a list are correlated or not appears to make no difference once potential confounds are controlled. Therefore, as far as the target-distractor correlation is concerned,[10] there appears to be no empirical basis for "reclaiming" the Stroop effect back from control accounts to input-driven accounts, as Algom and Chajut (2019) contended researchers should do, or to assume that Stroop effects in most experiments in the literature were artificially inflated by the common use of designs with target-distractor correlations different from zero (e.g., Algom & Chajut, 2019; Algom et al., 2022; Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003). While proponents of input-driven accounts are correct in noting that many Stroop experiments do create target-distractor correlations different from zero, they are most likely incorrect that perception of those correlations would have such a large impact that large Stroop effects, or large differences between Stroop effects such as PC effects, would be produced as a result of that process. Note, further, that the exploratory data reported in the Supplementary Materials suggest that people infrequently become aware of the correlated nature of a list, and whether they do so or not does not seem to have a strong relationship with their performance in the experiment. In the following section, we address some of the potential limitations in our experiments that could challenge these conclusions.

Potential limitations

One apparent limitation of our experiments is that, despite our claim that what we have called repetition-priming/contingency-learning processes were controlled for in the stimuli used for the contrast between uncorrelated and correlated lists, contingency learning was not actually controlled according to the traditional definition of that process in the color-word contingency-learning literature. According to that definition, an association is formed between a distractor and a response when the

---

[10] Note that we are making no claim about other factors that, according to input-driven accounts, would be important determinants of Stroop effects, e.g., the relative ease with which targets and distractors are processed.

relative probability of that response conditioned on that distractor is higher than the relative probability of the other responses associated with that distractor. Based on this definition, three types of stimuli have been distinguished (see, in particular, Lin & MacLeod, 2018; see also Schmidt et al., 2007): high-contingency stimuli, e.g., a word appearing in the color that is most typical (i.e., has a higher probability for that word than the other colors used in the experiment); low-contingency stimuli, e.g., a word appearing in a color that is not typical (has a lower probability for that word than the other colors used in the experiment); and no-contingency stimuli, e.g., a word that does not have typical or atypical colors because all colors are equally probable for that word in the experiment.

In experiments in which the stimuli are created using exclusively color-unrelated (i.e., neutral) words (e.g., the word MONTH appearing in yellow, its typical color, more often than in red and green, its atypical colors), responding to the color of high-contingency stimuli is typically faster than for no-contingency ones (i.e., a contingency-learning facilitation effect), and no-contingency stimuli are typically responded to faster than low-contingency ones (i.e., a contingency-learning interference effect). Note, however, that in contrast with facilitation and interference in the Stroop effect, interference tends be smaller (and sometimes, not significant) than facilitation in the contingency-learning effect (Geukes et al., 2019; Lin & MacLeod, 2018).

This traditional contingency-learning definition also applies to stimuli in proper Stroop experiments, that is, experiments including color names such as the present experiments. In the present experiments, all stimuli in the uncorrelated lists, including congruent, incongruent, and (where used) neutral stimuli, were no-contingency stimuli. For example, RED in the uncorrelated list in Experiment 1 (see the left side of Table 4) had the same probability of appearing with red, the congruent color, as with any of the incongruent colors. Similarly, XXXX had the same probability of appearing with any of the four colors used in that experiment.

However, the same is not true for the correlated lists. In those lists, neutral stimuli, in the experiments that included those stimuli (i.e., Experiments 1 and 2), remained no-contingency stimuli because the neutral distractor in the experiment was equally associated with all of the colors used. However, for each of the color names, one of the incongruent colors was made the typical, i.e., high-contingency, color (e.g., the color green for the word RED, see the right side of Table 4). As a result, the other incongruent colors (e.g., RED in blue) and the congruent color (e.g., RED in red) became atypical, i.e., low-contingency, colors for that color name.

Remember that the latter stimuli in the correlated lists were the ones that were contrasted with the congruent and incongruent stimuli in the uncorrelated list. Thus, congruent and incongruent stimuli in the uncorrelated list, stimuli that were no-contingency stimuli, were contrasted with congruent and incongruent stimuli in the correlated list, stimuli that were low-contingency stimuli. From this point of view, our contrast could be presumed to confound list type (uncorrelated vs. correlated) and contingency type (no-contingency vs. low-contingency) because the stimuli being contrasted in the two lists were not matched on contingency (for an example of Stroop experiments in which contingency learning was manipulated/controlled based on this definition, see Hasshim & Parris, 2021).

The traditional contingency-learning definition, however, might be a little outdated because, as noted in the Introduction, recent research suggests that contingency learning may essentially be a repetition-priming phenomenon, just as feature binding is (Schmidt et al., 2020). In fact, Schmidt et al. suggested that statistically controlling for discrete binding events would explain virtually all of contingency-learning effects without the need to assume an additional process. Although the situation is more complicated than described here (for discussions, see Schmidt et al., 2020), the point is that binding effects and contingency-learning effects would be based entirely on stimulus repetitions. Therefore, the implication is that what is being learned in the contingency-learning process is the association between the distractor and the (correct) response in the stimuli that are most frequently presented.

Based on this slightly altered definition, the impact of contingency learning would seem to have been controlled in our contrast because the individual stimulus frequencies of the stimuli involved in that contrast were matched across the two lists. More specifically, despite the fact that, by the traditional definition of contingency learning, our congruent and incongruent stimuli were no-contingency stimuli in the uncorrelated lists whereas they were low-contingency stimuli in the correlated lists, the fact that each of those stimuli was presented with the same frequency in the two lists (e.g., 10 times in Experiment 1) suggests that contingency learning, as has more recently been conceived, would have had a similar impact on the stimuli involved in the two lists.

Further, even if one preferred to stand by the traditional definition of contingency learning, it should be noted that although contingency type was confounded with list type in our experiments, it was not confounded with congruency as far as the stimuli involved in the Stroop effect are concerned. The reason is that *both* congruent and incongruent stimuli were no-contingency stimuli in the uncorrelated list and *both* congruent and incongruent stimuli were low-contingency stimuli in the correlated list. Therefore, the only impact that contingency learning, according to the traditional definition, could have had in that contrast was an overall slowdown in the correlated list for congruent *and* incongruent stimuli due to both being low-contingency stimuli. Most importantly, that type of contingency learning could not create a difference between the two lists with respect to the magnitude of the Stroop effect, which was the focus of the present experiments.

A related potential concern with our experiments is that the magnitude of what we have called the “stimulus frequency effect”, that is, the difference between the “high-frequency” incongruent stimuli in correlated lists (in Lin and MacLeod’s, 2018, terms, the “high-contingency” incongruent stimuli) and the “normal-frequency” incongruent stimuli in the same lists (the “low-contingency” incongruent stimuli according to Lin and MacLeod), was not large. Although the former stimuli were consistently faster than the latter across our experiments, consistent with the ideas that the correlated nature of those lists had some impact on performance and that repetition priming/contingency learning were engaged, the difference ranged from 11 to 16 ms. Contrasts between high-contingency and low-contingency stimuli in color-word contingency-learning experiments typically produce larger effects, e.g., around 50 ms (Lin & MacLeod, 2018; Schmidt et al., 2007, 2010). Therefore, the smaller-than-normal size of the effect in our contrast might be taken to mean that participants in our experiments were not as sensitive to the correlated nature of our correlated lists as they could have been and, by implication, our correlation manipulation was not particularly strong.

Note, however, that that 50-ms figure that we indicated typically comes from experiments in which 1) stronger contingency manipulations were used (e.g., manipulations in which the high-contingency color occurred 8 times more frequently than each of the low-contingency colors whereas in our experiments, high-contingency stimuli occurred only twice as frequently as low-contingency stimuli in Experiments 1-3 and 4 times as frequently in Experiment 4), 2) color-unrelated stimuli were used, and 3) manual responses to colors were used. All of these factors have been shown to increase the magnitude of contingency-learning effects (for evidence that those effects are larger with greater contingency strength, see Forrin & MacLeod, 2018; for evidence that they are larger with color-unrelated stimuli compared to color names, see Whitehead et al., 2018; and finally, for evidence that they are larger with manual than vocal responses to colors, see Forrin & MacLeod, 2017; Spinelli et al., 2020). Therefore, the fact that, in our experiments, we reported effect sizes on the small side would not seem to be particularly odd. Further, as argued by Schmidt (2018), a smaller effect size does not necessarily mean that the underlying process would be different: It may simply mean that the repetition-priming/contingency-learning process has a smaller window of opportunity for influencing behavior.

More importantly, the present experiments were designed to create a strong target-distractor correlation, not a strong difference between high-contingency and low-contingency stimuli. As noted in the Introduction when discussing the difference between target-distractor correlation and stimulus informativeness, a list's target-distractor correlation can be large even if there are no contingencies to learn in that list. What matters is the degree to which targets and distractors are correlated, not whether each word has a very typical color. Because in our experiments we contrasted lists with a *C* of zero (the smallest possible absolute value) with lists with an absolute value of *C* as large as .82 (a value that, as noted above, is close to the largest value examined in Melara and Algom's (2003) re-analysis of Stroop effects as a function of *C* values), it seems unlikely that our correlation manipulation could be deemed weak.

Of course, we do not mean to claim that target-distractor correlations play no role at all in Stroop task performance. For example, it may be possible to observe an effect if the list's distractor-correlation was made even stronger by further increasing the stimulus set beyond a size of six (i.e., the stimulus set size in our Experiment 4). Also, some factors that we did not examine may modulate the impact that the list's target-distractor correlation has on performance. For example, while participants may not feel a need to use that correlation when the task is relatively easy, they may feel that need in situations in which the task is made more difficult. For example, when manual, arbitrary responses to multiple colors are required, it may be more beneficial to use the word to predict the possible colors it may appear in, and, hence, narrow down the potential responses.

What is important to note, however, is that the present experiments were commensurate with those used to motivate the claims made by proponents of input-driven accounts, that is, that a list's target-distractor correlation is a general, pervasive factor in the Stroop literature, so much so that entire research programs that ignore it would be misguided (Algom & Chajut, 2019; Algom et al., 2022; Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003). Because the vocal color-word Stroop task is considered the "gold standard" in that literature (MacLeod, 1992), the fact that we found virtually no evidence for a role of target-distractor correlation in that task appears, at the very least, to destroy the alleged generality of that role. If target-distractor correlation has a role in Stroop experiments, that role may be less likely to be found in the most typical situations than in specific (and yet to be determined) ones.

Reverse facilitation: Exception or norm?

As noted, Experiments 1 and 2 produced the unexpected result of reverse facilitation effects in the response times. As was also noted, those effects appear to be at odds with current theorizing on reverse facilitation (Goldfarb & Henik, 2007; Kalanthroff et al., 2018), according to which reverse facilitation would only emerge in situations in which 1) congruent stimuli are contrasted with neutral stimuli that do not involve task conflict, such as letter strings, and 2) a reduction of control over task conflict is induced, for example, by increasing the proportion of letter-string neutral stimuli in the list. Although our Experiment 1 complied with the first condition because the neutral stimulus used in that experiment was a letter string, it did not comply with the second condition because stimuli involving task conflict represented the majority of stimuli in that experiment. The situation was even more extreme in Experiment 2 which did not comply with either condition because the neutral stimulus used in that experiment was a word, rather than a letter string, and *all* of the stimuli in the experiment were stimuli involving task conflict.

It is important to note that our letter-string neutral stimulus in Experiment 1, XXXX, is the standard stimulus that has been used in many Stroop experiments (MacLeod, 1991), and that our word neutral stimulus in Experiment 2, the word SHORT, was matched with the color names on several relevant variables such as part of speech, length in letters and in syllables, and frequency, without there being much orthographic and phonological overlap with the color names. Therefore, the particular stimuli that we chose to create the neutral stimuli in the two experiments are unlikely to be the cause of the unusual pattern that we obtained. Further, the fact that the pattern emerged in two experiments seems to challenge the idea that the pattern is spurious. What would that cause be, then?

Here, we propose a speculative but somewhat bold answer: In standard Stroop experiments, such as those reported here, reverse facilitation may not be the exception, but something closer to the norm. That is, obtaining reverse facilitation may not require a special experiment setup, as suggested by Kalanthroff et al. (2018) – it might be sufficient that the individual frequency of the congruent and neutral stimuli used be matched. By "individual stimulus frequency" we mean the number of times a particular color-word combination (e.g., RED in red, or XXXX in red) appears in the experiment.

Congruent and neutral stimuli were matched on individual stimulus frequency in the present Experiments 1 and 2, in which, in both correlated and uncorrelated lists, each congruent and each neutral stimulus appeared the same number of times (i.e., 10, see Table 4). That situation occurred also in other Stroop experiments that produced reverse facilitation (e.g., Schulz, 1979; Sichel & Chandler, 1969). In the general Stroop literature, however, the situation is often different. Specifically, it is often the case that the individual stimulus frequency of congruent stimuli is higher than that of neutral stimuli. For example, some researchers have used a small set of congruent target-distractor combinations (e.g., 4 combinations) and a larger set of neutral target-distractor combinations (e.g., 16 combinations), with the result being that, even with equal total proportions of congruent and neutral stimuli, the former would appear with higher individual frequencies than the latter (e.g., Augustinova et al., 2019; Bugg et al., 2011b; Dalrymple-Alford, 1972; Spieler et al., 1996). Some researchers have even presented each individual neutral stimulus only once (i.e., as infrequently as possible) while presenting each individual congruent stimulus repeatedly (e.g., Brown et al., 2002a, 2002b, 2011). Others have included a lower proportion of neutral than congruent stimuli in the experiment, with the result being that the individual frequencies of neutral stimuli in those experiments was also lower than those of congruent stimuli (e.g., Glaser & Glaser, 1982; Kane and Engle, 2003).

The implication is that, in those Stroop experiments, the contrast between congruent and neutral stimuli was confounded with the individual frequency of those stimuli, with congruent stimuli being presented more frequently (for a similar point, see Lorentz et al., 2016). That frequency difference might have contributed to producing the regular facilitation effect that is often observed in those experiments. A similar reasoning would apply to the manipulations that, conversely, presented neutral stimuli with higher individual frequencies than congruent stimuli (e.g., Entel et al., 2015; Goldfarb & Henik, 2007), often leading to the observation of reverse facilitation effects. It is possible that at least part of those effects was produced by the fact that a large advantage was given to each neutral stimulus by presenting it more often than each congruent stimulus.

One question that this speculation would raise, of course, is why reverse facilitation would be the norm – that is, why, when congruent and neutral stimuli are matched on individual stimulus frequency, the former would be slower than the latter rather than the other way round. This question would be easily addressed for the contrast between congruent stimuli and letter-string neutral stimuli (the contrast we examined in Experiment 1) by assuming 1) that the former involve task conflict but the latter do not, a commonly made assumption (e.g., Goldfarb & Henik, 2007; Kalanthroff et al., 2018; Monsell et al., 2001; Parris et al., 2021), and 2) that control over task conflict is not as efficient as is currently thought and would create a cost (i.e., leading to reverse facilitation) even in lists in which most of the stimuli involve task conflict (e.g., lists in which most of the stimuli are color names), although the cost may be smaller in those lists.

Explaining the reason that a reverse facilitation effect, albeit smaller, occurred in the contrast between congruent and word neutral stimuli in Experiment 2 is more of a challenge. It would seem reasonable that task conflict is not an all-or-none type of conflict, such that all readable stimuli involve task conflict and all non-readable stimuli do not. Instead, task conflict likely comes in degrees. In particular, in the context of a color-naming task, task conflict is likely lower for words that are not color names, including words such as SHORT, because those names are not eligible responses for the instructed task (e.g., because SHORT is not a color name, it can never be the case that a stimulus would require "short" as the response). In contrast, task conflict would likely be higher for color names, including congruent stimuli, because color names are eligible responses in the instructed task (e.g., for RED in red, the word RED would be an appropriate color response; for a similar point concerning effects in picture-word interference experiments, see Lupker, 1979). The two tasks would then be confused more easily for congruent stimuli than they would be for word neutral stimuli, leading to slower response times for the former stimuli. Further, this confusion would persist even in situations in which task conflict is frequently experienced, such as a list in which all stimuli involve task conflict to some degree, leading to reverse facilitation in those situations as well.

Again, these ideas are no more than speculation at this point. Corroborating them would require, first, replicating the patterns observed in the present experiments, not only in direct replications but also in slightly different designs (e.g., a design in which the number of neutral distractors used matches that of the non-neutral distractors and the way in which the neutral distractors are combined with the targets parallels the way in which the non-neutral distractors are – see Lorentz et al., 2016). Second, meta-analytic and experimental work would also be required to test these ideas, work that we are currently pursuing. At the very least, however, the unexpected reverse facilitation effects that we reported invite a rethinking of the conditions under which reverse facilitation would occur. Those effects also seem to be relevant to one of the arguments that proponents of input-driven accounts have recently used to dismiss control accounts – the argument that the latter accounts only focus on interference and would be unable to explain regular facilitation effects (Algom & Chajut, 2019; Algom et al., 2022). Here, not only were facilitation effects small, as they typically are – they were reversed. Therefore, there was no facilitation to explain, a pattern that may not be limited to the present experiments.

Increasing attention to distractors: What would be the benefit?

We conclude with a more conceptual point concerning the potential impact of a list's target-distractor correlation. We do so because many of the arguments recently used by critics of control accounts are conceptual in nature. For example, Algom and Chajut (2019) and Algom et al. (2022) frequently refer to those accounts as lacking "parsimony" and to the processes assumed by those accounts as being "gratuitous" (see also Schmidt, 2013a, 2019). Similarly, Schmidt (2023) has recently questioned the "adaptive" nature of control accounts reasoning that effects that are typically attributed to adaptive control, such as PC effects, involve no real performance benefit overall.

We believe that this line of conceptual examination is useful for exposing the weaknesses that undoubtedly affect control accounts, but we suggest that the same types of criticisms can be leveled at some of the alternative processes that critics of control accounts propose. Consider, in particular, the process of increasing attention to distractors in lists with a large target-distractor correlation (Dishon-Berkovits & Algom, 2000; Melara & Algom, 2003) and/or informative lists, that is, lists in which word-response contingencies can be learned (Schmidt, 2014, 2019). What is important to note is that, because that type of process is assumed to be a general process, it is independent from the congruency of the stimuli in the list. Therefore, for example, that process could be applied both in lists in which most of the trials are congruent and in lists in which most of the trials of the trials are incongruent.

Indeed, Algom and Chajut (2019) and Algom et al. (2022) specifically discuss the case of Stroop's (1935) original experiment, Experiment 2, an experiment which involved a list in which each of five color names was presented in each of the four noncorresponding colors, i.e., a 100% incongruent list. Because each color name never occurred in the congruent color in that list, the color-word correlation was not zero (in fact, C = -.45), a fact that, according to Algom and collaborators, would help explain the large interference effect (47 s) that Stroop obtained when contrasting the total time participants took to complete that list with the total time they took to complete a control list including color patches. What is assumed by input-driven accounts would be that participants increased attention to words in the

100% incongruent list because colors and words were correlated – more specifically, the words could be used to predict the color that they would not be presented with (i.e., the congruent color). As a result of receiving increased attention, the words would then produce larger interference effects.

The glaring question, here, is why would participants direct attention to words in order to make a prediction that is not very useful (because it would only eliminate one color, the congruent color, out of five potential colors), only to suffer increased interference from the incongruent words as a result of the increased attention allocated to the words? In general, increasing attention to distractors would not seem to produce real benefits, and in situations of this sort, it would only seem to produce costs. It seems somewhat strange that, while proponents of input-driven accounts question the adaptive nature of control accounts, they do not seem to have a problem with their own assumption that participants would increase attention to distractors even in situations in which distractors are often (or always) incongruent with the target, an assumption which would mean that participants would engage in *maladaptive* behavior.

Similar reasoning could be made concerning the present experiments, for which larger facilitation, but also larger interference was hypothesized in the correlated lists, lists that were also informative in Schmidt's (2017, 2019) sense of the word. If we had observed those patterns (which we did not), they would have mainly reflected increased costs to increasing attention to distractors. Thus, in many situations, it is not clear why a process of increasing attention to distractors in the presence of large target-distractor correlations and/or informative lists would ever be useful.

Indeed, it is not clear why people would not engage, if anything, the opposite process, that is, increasing attention to *targets* when targets and distractors are correlated. When targets and distractors are correlated, the latter can predict the former, but the former can also predict the latter. For example, in the correlated list in our Experiment 1 (see the right side of Table 4), participants could predict that the word RED would occur with the colors red, blue, and green. However, they could also predict that the *color* red would occur with the *words* RED, GREEN, YELLOW, or with XXXX. The latter prediction would seem to be more useful because it would allow participants to prepare to handle the words that a color is most likely to occur with, especially the incongruent ones (e.g., in the case of the color red, the words GREEN and YELLOW). In order to use these types of predictions in correlated lists, potentially, attention to *targets*, rather than to *distractors*, might be increased. The same would not be true in uncorrelated lists, lists in which colors are not at all predictive of the identity of the associated words (e.g., in the uncorrelated list in Experiment 1, the color red occurs with all four color names and XXXX equally often, see the left side of Table 4). Therefore, participants would have no reason to increase attention to the targets in those lists.

Note that target information, as well as distractor information, has been shown to trigger attentional adjustments in the Stroop task (e.g., Bugg et al., 2011a; Bugg & Hutchison, 2013; Spinelli et al., 2022). Therefore, conceptually, there appears to be no reason to assume that, in correlated situations, participants would increase attention to distractors. They may very well increase attention to targets instead, a process that, overall, would seem more useful because, if anything, it would *reduce*, rather than *increase*, the interference produced by incongruent words.

A convincing account of target-distractor correlation effects in Stroop experiments would need to address these concerns. Empirically, in any case, the main point made by the present experiments is that they produced no evidence in support of the presence of those effects, with Bayes Factor analyses favoring, in fact, their absence. Although those effects may emerge in different situations than the situations examined here, situations which future research should attempt to discover, what the present research suggests is that the idea that a process of attentional adjustment to the list's target-distractor correlation would be lurking in many published Stroop experiments, potentially compromising their interpretation, may very well have no solid basis.

Funding acknowledgements

This research was partially supported by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

Data availability statement

The raw data, JASP files, and study materials are available at https://osf.io/buvxy/. The research was not preregistered.

Author note

Correspondence concerning this article may be addressed to: Giacomo Spinelli, Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano (MI), Italy (e-mail: giacomospinelli@hotmail.it) or Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, N6A 5C2, Canada (e-mail: lupker@uwo.ca).

References

Algom, D., & Chajut, E. (2019). Reclaiming the Stroop Effect Back From Control to Input-Driven Attention and Perception. *Frontiers in Psychology*, 10. https://doi.org/10.3389/fpsyg.2019.01683

Algom, D., Fitousi, D., & Chajut, E. (2022). Can the Stroop effect serve as the gold standard of conflict monitoring and control? A conceptual critique. *Memory & Cognition*, *50*, 883-897. https://doi.org/10.3758/s13421-021-01251-5

Arieh, Y., & Algom, D. (2002). Processing picture-word stimuli: The contingent nature of picture and of word superiority. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 221-232. https://doi.org/10.1037/0278-7393.28.1.221

Augustinova, M., Parris, B. A., & Ferrand, L. (2019). The loci of Stroop interference and facilitation effects with manual and vocal responses. *Frontiers in Psychology*, *10*, 1786. https://doi.org/10.3389/fpsyg.2019.01786

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

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652. https://doi.org/10.1037/0033-295X.108.3.624

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, *23*, 769-783. https://doi.org/10.1016/j.tics.2019.07.002

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

Brown, T. L. (2011). The relationship between stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 85-99. https://doi.org/10.1037/a0019252

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition "automatic?". *Journal of Experimental Psychology: General*, *131*, 220-240. https://doi.org/10.1037/0096-3445.131.2.220

Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the Stroop task: Testing the limits of involuntary word processing. *The American Journal of Psychology*, *115*, 515-543. https://doi.org/10.2307/1423526

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*, 991-997. https://doi.org/10.3758/s13428-012-0190-4

Bugg, J. (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. https://doi.org/10.1037/a0035032

Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*, 367. https://doi.org/10.3389/fpsyg.2012.00367

Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 433-449. https://doi.org/10.1037/a0029145

Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011b). Why it is too early to lose control in accounts of item-specific proportion congruency effects. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 844-859. https://doi.org/10.1037/a0019957

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

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

Cochrane, B. A., & Pratt, J. (2022). The item-specific proportion congruency effect can be contaminated by short-term repetition priming. *Attention, Perception, & Psychophysics*, *84*, 1-9. https://doi.org/10.3758/s13414-021-02403-0

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1703-1723. https://doi.org/10.1037/xlm0000670

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

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

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 622-639. https://doi.org/10.1037/0096-1523.10.5.622

Entel, O., Tzelgov, J., Bereby-Meyer, Y., & Shahar, N. (2015). Exploring relations between task conflict and informational conflict in the Stroop task. *Psychological Research*, *79*, 913-927. https://doi.org/10.1007/s00426-014-0630-0

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

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

Forrin, N. D., & MacLeod, C. M. (2018). Contingency proportion systematically influences contingency learning. *Attention, Perception, & Psychophysics*, *80*, 155-165. https://doi.org/10.3758/s13414-017-1424-4

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

Geukes, S., Vorberg, D., & Zwitserlood, P. (2019). Disentangling semantic and response learning effects in color-word contingency learning. *PLOS ONE*, *14*, e0212714. https://doi.org/10.1371/journal.pone.0212714

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 875-894. https://doi.org/10.1037/0096-1523.8.6.875

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1170-1176. https://doi.org/10.1037/0096-1523.33.5.1170

Hasshim, N., & Parris, B. A. (2021). The role of contingency and correlation in the Stroop task. *Quarterly Journal of Experimental Psychology*, *74*, 1657-1668. https://doi.org/10.1177/17470218211032548

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

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*, 183-216. https://doi.org/10.1080/713756773

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

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 851-860. https://doi.org/10.1037/a0023437

JASP Team (2022). *JASP* (Version 0.16.4.0) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review*, *125*, 59-82. https://doi.org/10.1037/rev0000083

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70. https://doi.org/10.1037/0096-3445.132.1.47

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian data analysis for cognitive science: A practical course*. New York, NY: Cambridge University Press.

Lin, O. Y. H., & MacLeod, C. M. (2018). The acquisition of simple associations as observed in color–word contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 99-106. https://doi.org/10.1037/xlm0000436

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527. https://doi.org/10.1037/0033-295X.95.4.492

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166-174. https://doi.org/10.3758/BF03197535

Lorentz, E., McKibben, T., Ekstrand, C., Gould, L., Anton, K., & Borowsky, R. (2016). Disentangling genuine semantic Stroop effects in reading from contingency effects: On the need for two neutral baselines. *Frontiers in Psychology*, *7*, 386. https://doi.org/10.3389/fpsyg.2016.00386

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

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

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, *121*, 12–14. https://doi.org/10.1037/0096-3445.121.1.12

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383-391. https://doi.org/10.1016/S1364-6613(00)01530-8

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

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete?. *Memory & Cognition*, *29*, 137-151. https://doi.org/10.3758/BF03195748

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

Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M., & Ferrand, L. (2021). The loci of Stroop effects: a critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*, *86*, 1029-1053. https://doi.org/10.1007/s00426-021-01554-x

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859-862. https://doi.org/10.3758/BF03192979

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

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

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

Schmidt, J. R. (2013c). The Parallel Episodic Processing (PEP) model: Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychologica*, *142*, 119-126. https://doi.org/10.1016/j.actpsy.2012.11.004

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

Schmidt, J. R. (2017). Time-out for conflict monitoring theory: Preventing rhythmic biases eliminates the list-level proportion congruent effect. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *71*, 52-62. https://doi.org/10.1037/cep0000106

Schmidt, J. R. (2018). Best not to bet on the horserace: A comment on Forrin and MacLeod (2017) and a relevant stimulus-response compatibility view of colour-word contingency learning asymmetries. *Memory & Cognition*, *46*, 326-335. https://doi.org/10.3758/s13421-017-0755-7

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, *26*, 753-771. https://doi.org/10.3758/s13423-018-1520-z

Schmidt, J. R. (2021). When data transformations are appropriate or even necessary: A response to Cohen-Shikora, Suh and Bugg (2019). *Timing & Time Perception*, *9*, 161-197. https://doi.org/10.1163/22134468-bja10019

Schmidt, J. R. (2023). Is conflict adaptation adaptive? An introduction to conflict monitoring theory and the ecological problems it faces. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218231161555

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 514-523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., Crump, M. J., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*, 421-435. https://doi.org/10.1016/j.concog.2006.06.010

Schmidt, J. R., De Houwer, J., & Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, *19*, 235-250. https://doi.org/10.1016/j.concog.2009.12.016

Schmidt, J. R., Giesen, C. G., & Rothermund, K. (2020). Contingency learning as binding? Testing an exemplar view of the colour-word contingency learning effect. *Quarterly Journal of Experimental Psychology*, *73*, 739-761. https://doi.org/10.1177/1747021820906397

Schulz, T. (1979). Components of the reaction time Stroop-task. *Psychological Research*, *40*, 377-395. https://doi.org/10.1007/BF00309418

Sichel, J. L., & Chandler, K. A. (1969). The color-word interference test: The effects of varied color-word combinations upon verbal response latency. *The Journal of Psychology*, *72*, 219-231. https://doi.org/10.1080/00223980.1969.10543502

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461-479. https://doi.org/10.1037/0096-1523.22.2.461

Spinelli, G., Krishna, K., Perry, J. R., & Lupker, S. J. (2020). Working memory load dissociates contingency learning and item-specific proportion-congruent effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 2007–2033. https://doi.org/10.1037/xlm0000934

Spinelli, G., & Lupker, S. J. (2020). 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*, *48,* 426-435. https://doi.org/10.3758/s13421-019-00980-y

Spinelli, G., & Lupker, S. J. (2021). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and* Cognition, *47*, 1550-1562. https://doi.org/10.1037/xlm0000820

Spinelli, G., & Lupker, S. J. (2022a). Conflict-monitoring theory in overtime: Is temporal learning a viable explanation for the congruency sequence effect? *Journal of Experimental Psychology: Human Perception and Performance*, *48*, 497-530. https://doi.org/10.1037/xhp0000996

Spinelli, G., & Lupker, S. J. (2022b). Robust evidence for proactive conflict adaptation in the proportion-congruent paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0001144

Spinelli, G., Morton, J. B., & Lupker, S. J. (2022). Both task-irrelevant and task-relevant information trigger reactive conflict adaptation in the item-specific proportion-congruent paradigm. *Psychonomic Bulletin & Review.* Advance online publication. https://doi.org/10.3758/s13423-022-02138-5

Spinelli, G., Perry, J. R., Lupker, S. J. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture-word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 995-1014. http://dx.doi.org/10.1037/xhp0000656

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

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*, 727-735. https://doi.org/10.3758/BF03202722

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., ... & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813-826. https://doi.org/10.3758/s13423-020-01798-5

Whitehead, P. S., Brewer, G. A., Patwary, N., & Blais, C. (2018). Contingency learning is reduced for high conflict stimuli. *Acta Psychologica*, *189*, 12-18. https://doi.org/10.1016/j.actpsy.2016.09.002