Trial type mixing substantially reduces the Response Set Effect in the Stroop task

Nabil Hasshim and Benjamin A. Parris

Author Note

Nabil Hasshim and Benjamin A. Parris, Department of Psychology, Faculty of Science and Technology, Bournemouth University.

This work was funded through a Bournemouth University PhD studentship awarded to the first author who was supervised by the second author. There are no conflicts of interest.

Correspondence concerning this article should be addressed to Nabil Hasshim or Dr. Ben Parris, Department of Psychology, Faculty of Science and Technology, Bournemouth University, Fern Barrow, Poole, BH12 5BB, United Kingdom

Email: nhasshim@bournemouth.ac.uk or bparris@bournemouth.ac.uk

Abstract

The *response set effect* refers to the finding that an irrelevant incongruent colour-word produces greater interference when it is one of the response options (referred to as a *response set trial*), compared to when it is not (a non-response set trial). Despite being a key effect for models of selective attention, the magnitude of the effect varies considerably across studies. We report two within-subjects experiments that tested the hypothesis that presentation format modulates the magnitude of the response set effect. Trial types (e.g. response set, non-response set, neutral) were either presented in separate blocks (pure) or in blocks containing trials from all conditions presented randomly (mixed)). In the first experiment we show that the response set effect is substantially reduced in the mixed block context as a result of a decrease in RTs to response set trials. By demonstrating the modulation of the response set effect under conditions of trial type mixing we present evidence that is difficult for models of the effect based on strategic, top-down biasing of attention to explain. In a second experiment we tested a stimulus-driven account of the response set effect by manipulating the number of colour-words that make up the non-response set of distractors. The results show that the greater the number of non-response set colour concepts, the smaller the response set effect. Alternative accounts of the data and its implications for research debating the automaticity of reading are discussed.

Trial type mixing substantially reduces the Response Set Effect in the Stroop task

Selective attention refers to the process of selecting only relevant and important parts of the perceptual landscape at the cost of less relevant or irrelevant parts. An experimental analogue of the selective attention challenges we face in everyday life comes in the form of the Stroop task (Stroop, 1935). The Stroop task requires participants to name the colour of the font in which a word is printed while ignoring the meaning of the word itself. The Stroop effect refers to the finding that naming the colour that a word is printed in takes longer when the word spells out a different colour (e.g. the word ‘red’ displayed in blue ink; an *incongruent* trial) compared to when the word spells out the same colour (e.g. the word ‘red’ displayed in red ink; a *congruent* trial) or when the word spells out a neutral word (one that is not associated with any colour, e.g. ‘table’) (see MacLeod, 1991; 2005 for comprehensive reviews of the Stroop task).

Selective attention makes it possible to overcome behaviours that are innate or have become automatic through continued practice, and instead perform behaviours that are in line with current goals (Diamond, 2013). To facilitate goal-oriented behaviour, mechanisms of selective attention appear to increase activation of goal-salient (relevant) concepts. This is demonstrated by the *response set effect,* which refers to the well-established finding that items (e.g. colours) that make up the set of possible responses (task-relevant items) are rendered more salient and, as a consequence, are harder to ignore when in an irrelevant, interfering dimension. In the context of the Stroop task the response set effect refers to the finding that greater interference occurs when the incongruent irrelevant word spells out a possible response option (it is part of the response set) compared to when it is not (i.e. the word spells out a colour that is not part of the response set). Such incongruent trials are referred to as *non-response set trials* (e.g. the word ‘orange’ in blue, when the colour orange is not a possible response colour). Hence, the response set effect is defined as the difference in response time (RT) between response set (incongruent trials where the irrelevant word spells out a colour that is part of the response set) and non-response set trials, whereas the standard term ‘Stroop interference’ generally means the difference between response set trials and a neutral (or congruent) baseline. The response set effect is therefore a component of the larger Stroop interference effect and is often employed as a pure measure of response competition/conflict (see Klein, 1964; Milham et al., 2001; Risko, Schmidt & Besner, 2006; Sharma & McKenna, 1998; also see MacLeod, 1991, for a review).

On the other hand, non-response set incongruent trials have been shown to produce interference compared to a neutral non-colour related word or a congruent trial (e.g. Klein, 1964; and Sharma & McKenna, 1998), which when added to the response set effect would make up the rest of Stroop interference. The difference between non-response set and neutral trials has been attributed to the irrelevant non-response set word belonging to the same semantic category as the eligible response colours and is thus interpreted as indexing semantic conflict. This concurs with evidence showing that interference can occur independently at different levels of processing such as earlier stimulus encoding and lexico-semantic processing stages (e.g. Goldfarb & Henik, 2007; Hock & Egeth, 1970; Luo, 1999; Parris, 2014). Thus, the response set effect is not only important in highlighting a key mechanism of selective attention, it is also consequential for those wanting to dissociate different types of conflict in the Stroop task (e.g. Augustinova & Ferrand, 2014a; Schmidt & Cheesman, 2005; Sharma & McKenna, 1998).

In his review of the Stroop effect, MacLeod (1991) identified the response set effect as one of 18 well-established findings for which models of the effect need to account. Two prominent models of the Stroop task (Cohen, Dunbar, & McClelland, 1990 and WEAVER++; Roelofs, 2000), have accounted for response set effects by proposing that attention is selectively, and thus strategically, allocated to the restricted set of eligible colours in a top-down manner. This ensures that their activation levels to the colours are greater than those to colours not in the response set. In the Cohen et al. (1990) model and its later incarnations (Botvinick et al., 2001), response colour concepts are identified by task demand units where a bias is set such that those particular colours are more likely to guide attention. However, there is no description of the specific processes involved in establishing the colours as response set colours beyond attributing the process to the top-down task demand unit. This is the same process that establishes colour naming as the task goal.

In the WEAVER++ model, the nodes of response set colours are flagged as goal concepts, which allows for subsequent selection and processing of information gleaned from a stimulus. Colours that are not part of the response set are not flagged and thus are less likely to be processed as a potential response or interfere with response selection (although see Caramazza & Costa, 2000; 2001 for evidence against the flagging component). Non-response set trials interfere only through their connections to the flagged response set nodes in the conceptual network. Given this connection, any manipulation that affects performance on incongruent trials would indirectly affect the performance of non-response set trials in tandem, but likely to a smaller degree since second-order activations would be smaller due to being further along the activation pathway. Similar to the Cohen et al. model, there is no description of the development of this process although Roelofs (2001) stated that simple repetition over a few trials would be required to achieve response-level salience, indicating a learning mechanism. Notably, since in both models, greater bias is given to only the response colours (and not for example to the non-response set colours) they describe a strategic, top-down, goal-driven mechanism.

Potential contextual modulation of the response set effect

A review of the literature on studies reporting the use of response set trials indicates that the magnitude of the response set effect varies considerably from study to study, independent of response mode (see Table 1 detailing these studies, their presentation type and measured response set effects). It is the contention of the present work that an experiment's presentation format is a possible moderator of the size of the response set effect. As can be seen in Table 1, studies that present trials in a mixed order (e.g. Proctor, 1978; Stirling, 1979; Hasshim & Parris, 2014) seem to show much smaller response set effects compared to studies that present trials in pure blocks containing only one type of trial in each block[1]. This is of theoretical importance because prominent models of the Stroop task have heavily drawn from the results of early classic studies (see MacLeod, 1991) that favoured pure block presentation due to technological limitations (stimuli were presented on cards and RTs for each block were recorded using a stopwatch; a practice still common in clinical and neuropsychological settings where millisecond precision might not be necessary), while presenting trials in random, mixed order has now become standard in laboratory research. The models described earlier are unable to account

[1] It should be noted that the list in Table 1 includes studies that utilised variants of the Stroop task, namely the picture naming task (Camarazza & Costa, 2000; La Heij, 1998) and digit counting task, (West et al., 2004), which involve processes that do not fully overlap with the regular Stroop paradigm. Although not within the scope of the current research, this may indicate that the phenomenon is applicable to a more general effect on selective attention and not specific to the regular colour naming Stroop paradigm.

for this apparent pattern because even though they account for the response set effect in different ways, they both assume that identifying specific colours as being in the response set occurs early via a top-down strategy to establish that they have goal-level salience. Since colour concepts[2] are either response colours or not the establishment of response set colours should occur in the first few trials regardless of whether presentation is mixed or pure. Essentially, for the models, there is no difference between the mixed and pure block presentations. As such, demonstrating an effect of presentation format on the response set effect would challenge existing top-down accounts of this important mechanism of selective attention.

We report two within-subjects experiments that tested the prediction that trial type mixing reduces the response set effect in the Stroop task. In the first experiment we compared the response set effect in mixed vs. pure blocks and show that the response set effect is indeed substantially reduced in the mixed block context as a result of a decrease in RTs to response set trials. In a second experiment we tested a stimulus-driven account of the response set effect.

Experiment 1

The goal of Experiment 1 was to determine whether the response set effect was smaller when trials were presented in mixed blocks compared to when presented in pure blocks as suggested by the observation in Table 1.

Method

Participants

[2] The colours indicated in the irrelevant dimension are not actually presented, but instead their concepts are indicated by the words describing them i.e. the concept of ‘blueness’ is indicated by the word *blue;* hence we occasionally use the term ‘colour concepts’.

40 participants (9 male) were recruited from the student population of Bournemouth University in exchange for course credit or £5. They had a mean age of 21.7 (*SD* = 4.38).

Apparatus and Materials

Stimuli were presented on a PC using Experiment Builder software (SR Research Ltd.) with responses recorded via pressing one of the assigned keys on a Cedrus response pad (RB 740, Cedrus Corporation). Three response keys were used with each key assigned one of the three possible colour responses.

Design

The experiment consisted of a 3 (trial type: neutral, vs non-response, vs response set) x 2 (presentation format: mixed blocks vs pure blocks) within subjects design.

Stimuli

To control for possible effects of different colours being in the response and non-response set, participants went through one of two versions of the experiment where the non-response colours of one version served as the response-set colours of the other. The colours used were yellow (RGB: 255; 255; 0), pink (255; 20; 147) and green (0; 255; 0) in one version, and blue (0; 112; 192), purple (204; 0; 255) and orange (255; 127; 0) in the other. The words *wall, marvel* and *story* were used for the neutral trials and had been matched for frequency and length using the English Lexicon Project (Balota et al., 2007). All the stimuli were presented in Arial font on a black background and the screen was approximately 60cm away from the participants (participant head position was not restricted), which resulted in each word having a vertical viewing angle of 0.95° and horizontal viewing angle of between 1.91°- 3.82°, depending on word length.

Table 1: Response-set effects from studies that have used non-response set trials

Study	Response-set Effect (ms)	Presentation Type	Response Type	Notes
Caramazza and Costa. (2000)	-1*	Mixed	Vocal	Picture-word naming task. Each block mixed neutral (unrelated) and either response set or non-response set trials
Hasshim and Parris (2014)	-13.65 (non-sig.)	Mixed	Manual	Two-to-one response Stroop task.
Klein (1964)	241**	Pure	Vocal	List method*** (not computerized)
La Heij (1988)	24	Mixed	Vocal	Used picture-word naming task
	12	Mixed	Vocal	Used picture-word naming task
Lamers et al. (2010)	11 19	Mixed	Vocal	Response membership established trial-by-trial
Milham et al.(2001)	6*	Mixed	Manual	Each block mixed neutral and either response set or non-response set trials
Proctor (1978)	111	Pure	Vocal	Experiment 1 - List method (not computerized)
	29.0	Mixed	Vocal	Experiment 2
	23.7	Mixed	Vocal	Experiment 3
Risko et al. (2006)	8	Mixed	Vocal	Used colour associates
	6	Mixed	Manual	Used colour associates
Scheibe et al. (1967)	205	Pure	Vocal	List method (not computerized)
Sharma and McKenna (1998)	96.7	Pure	Manual	
	63.6	Pure	Vocal	
Stirling (1979)	17 (non-sig.)	Mixed	Vocal	
	11 (non-sig.)	Mixed	Vocal	
West et al. (2004)	34	Mixed	Manual	Digit counting task
	12 (non-sig.)	Mixed	Manual	Digit counting task,

* Response set effect was calculated by the difference between the interference effects of the incongruent block and the non-response set block and the statistical non-significance of each comparison is noted only when reported by the study. Note that in Milham et al. the RTs to response set trials were slower than non-response set trials, the RTs of neutral trials in the latter block was faster as well.
** Response set effect was calculated by subtracting RTs of non-response set trials from incongruent trials. In cases where different types of non-response set trials were used, we chose the trial type that resembled standard non-response set trials the most.
*** The RTs for the list method experiments were calculated by dividing the overall time taken to go through the list, by the number of words in the list.

Because only two of the three response options are possible correct responses for response set trials (the third response button would correspond to a congruent trial, which are not involved in the experiment), the same limitation was imposed on each colour stimulus to ensure that regardless of trial type, each word stimulus had the same probability (50%) of its correct response being one of two response options. This was done by never pairing each word stimulus (neutral and colour word) to one specific colour each. The specific colour omitted was counterbalanced across the words in each trial type (e.g. the word *wall* never appeared in blue while *story* never appeared in green).

Procedure

At the start of each trial, participants were presented with a grey fixation cross in the centre of the screen for 500ms. This was followed by the Stroop stimulus, which remained at the centre of the screen until a response was executed. Participants were instructed to press the assigned key corresponding to the colour of the text as quickly and accurately as possible while ignoring the word's meaning. Upon committing an error, an additional auditory tone and a visual error message were presented. The error message lasted for 1500ms followed by a blank screen of 100ms.

Before the experimental trials participants went through a practice block of 60 trials made up of hash symbols (#) of three to six characters in length. For the experimental blocks, participants went through a total of 576 trials, made up of 96 trials of each trial type (neutral, non-response set and response set), presented in the two presentation formats (mixed and pure; i.e. 96 trials x 3 trial types x 2 presentation formats). Thus the proportion of neutral, non-response set and response set trials were equal throughout each version and presentation format.

During the experiment, trials were presented in blocks of 96 trials and the order of presentation format presented was randomised (i.e. participants either did all the pure or all mixed blocks first), as were the trial types within the pure blocks presentation (i.e. the order of the pure blocks were randomised)[3]. At the end of each block of 96 trials participants initiated a keypress to move on to the next block.

Results

Only correct responses within 200ms and 2500ms were included in the analyses. The proportions of valid responses for the mixed and pure blocks were .967 (*SD* = .027) and .965 (*SD* = .021) respectively. Table 2 lists the descriptive statistics for all four trial types.

A 3x2 repeated measures ANOVA revealed a significant main effect of trial type ($F(2,78) = 20.84$, $p < .001$, $r = .459$) while the main effect for presentation format was non-significant ($F(1,78) = 0.976$, $p = .329$, $r = .111$. Finally, the trial type (neutral, non-response set or response set) by presentation format (mixed or pure) interaction was significant, $F(2,78) = 3.56$, $p = .033$, $r = .209$.

A follow-up repeated measures one-way ANOVA measuring differences across the neutral, non-response set, and response set trial types

[3] Although presentation order was controlled for, an omnibus analysis with presentation order (pure or mixed blocks first) as a between subjects factor was conducted to investigate the possibility of order effects. This analysis was found to be statistically non-significant $F(2,76) = 1.15$, $p = .322$, $r = .122$) indicating that there was no order effects.

was significant for both mixed and pure blocks (F(2,78) = 20.589, p < .001, r = .457, and F(2,78) = 13.119, p < .001, r = .379, respectively).

Planned comparisons between the response set and non-response set trials for each presentation format showed that the response set trials were slower than non-response trials in both mixed (17.49ms, t(39) = 2.80, p = .008, r = .409), and pure block presentations (46.22ms, t(39) = 4.15, p < .001, r = .553) which meant that the response set effects for both presentation formats were statistically significant. Follow up comparisons of the size of the effect showed that the response set effect was larger in pure blocks compared to the mixed blocks (28.73ms, t(39) = 2.76, p =.009, r = .553).

Analyses of the non-response set effect (difference between non-response set and neutral trials), revealed the difference to be statistically significant in mixed blocks (23.39ms, t(39) = 4.16, p < .001, r = .554) but not in pure blocks (7.53ms, t(39) = 0.772, p = .445, r = .123). The follow up comparison of the size of non-response set effects in the two presentation yielded a non-significant result (-15.86ms, t(39) = -1.67, p =.104, r = .258).

These analyses on response set and non-response set effects suggest that presentation format mainly affects the former. To conclude that this is the case, a 2 (response set effect vs. non-response set effect) x 2 (mixed vs. pure presentation) analysis was conducted, which showed a statistically significant interaction (F(1,39) = 8.03, p = .007, r =.413.

To further determine the locus of the effect, the three trial types were compared across the presentation formats. Non-response set and neutral trials were non-significantly different across presentation format (-8.49ms, t(39) = -0.92, p = .365, r = .146; and 7.37ms, t(39) = 0.902, p = .373, r =

.143, respectively) but response set trials were slower in pure blocks (20.24ms, $t(39) = 2.15$, $p = .038$, $r = .326$). Given that there was no change in the neutral trials across conditions, we can take the change in the response set trials across conditions to be meaningful.

Accuracy analysis

The 3x2 repeated measures ANOVA on the accuracy rates revealed a non-significant interaction $F(2,78) = 1.55$, $p = .218$, $r = .140$. The main effect of trial type was also non-significant ($F(2,78) = 0.62$, $p = .543$, $r = .089$) as was the main effect of presentation format ($F(1,39) = 1.13$, $p = .295$, $r = .168$)

Table 2: Mean RTs,, and accuracy rates (and SEs) of all trial types and mean response set effect of Experiment 1

	Mixed		Pure	
	RT(ms)	Accuracy(%)	RT(ms)	Accuracy(%)
Neutral	586.62 (13.12)	92.95 (0.446)	593.99 (13.47)	92.42 (0.352)
Non-response set	610.01 (13.41)	93.20 (0.391)	601.52 (13.75)	92.68 (0.291)
Response set	627.50 (15.29)	92.43 (0.394)	647.74 (17.88)	92.70 (0.334)
Response set effect (Response set – Non-response set)	17.49 ms* (11.13)		46.22 ms* (6.24)	

*$p < .05$

Discussion

Consistent with our predictions, the results showed that presentation format modulated the size of the response set effect in the Stroop task; with larger effects observed when trials were presented in pure blocks compared to when presented in mixed blocks. The mixing effect was driven by slower RTs to response set trials in the pure block condition compared to in the mixed block condition. While RTs to non-response set trials were numerically larger in the mixed block presentation, this difference was statistically non-significant.

The mixing effect observed here suggests that it is more difficult to establish response-level salience in the mixed block context, which goes against the predictions of models such as WEAVER++ (Roelofs, 2003) and the PDP model of Cohen et al. (1990), where concepts salient to goals (i.e. response set colours) are identified via top-down processes of flagging or selective biasing of attention. Under such accounts the identification of such concepts should not be affected by experimental design since colours are either response colours or they are not.

Experiment 2

Experiment 1 established that there is an effect of presentation format on the response set effect. The finding that the response set effect is affected by presentation format seems counter intuitive given that only a restricted set of colours are response colours in both pure and mixed contexts. In both the pure and mixed blocks, participants would encounter the same restricted number of response colours and these colours would remain response colours throughout. It is the contention of the extant models, and to some extent common sense, that response level salience is established by effectively noting what the response colours are and strategically biasing their representations to ensure effective goal oriented behaviour (responding to the colour and ignoring the word). Showing an effect of presentation format on the response set effect questions this notion. The aim of the present experiment was to provide and test a stimulus-driven account of the mixing effect and, as a consequence, of the response set effect itself.

Lamers, Roelofs, and Rabeling-Keus (2010) tested competing accounts of response set effects, with one account, held by Roelofs (2003) and Cohen et al. (1990), arguing that response set effects arise due to the selective allocation of attention to eligible responses. They contrast this account with one based on greater inhibition of non-response set colours. This alternative inhibition account is the flip-side of the top-down facilitation of the task relevant colours account formalised in the Roleofs and Cohen and colleagues models. In one experiment, they manipulated response set membership on a trial-by-trial basis by cuing the possible responses before each trial. They also manipulated response set size, reasoning that doing so would make it more difficult to inhibit individual responses under the inhibition account. The results showed that response set effects were independent of response set size which was an additive effect. In their second experiment, the distractor colour was cued before each trial, which resulted in facilitation on both incongruent and congruent trials (they did not use non-response set trials in their second experiment). They concluded that the facilitation on congruent trials was evidence that pre-exposure to the distractor does not result in greater inhibition. These findings were argued to be consistent with the selective allocation of attention account. In contrast to Lamers et al.

approach, in the present experiment we proffer and test an account of the response set effect based on concurrent processing, not inhibition, of colours presented in the irrelevant dimension. More precisely, our account is based on the notion that any colour concept that is encountered, whether in the relevant or irrelevant dimension, contributes to establishing colour concept salience. By salience we mean that the activation level of the representation of a colour concept increases. Activation levels increase upon encountering a colour concept and the more it is encountered the greater the activation level or salience of that colour concept, independent of whether or not that colour concept is encountered in the relevant or irrelevant dimension. This account contrasts with previous accounts that are based solely on establishing salience of the restricted set of colours in the relevant dimension.

Participants are exposed to the same trials in both presentation formats. However, when trials are presented in pure blocks, they are exposed to a restricted set of colour concepts within each block of neutral and response set trials (i.e. they are exposed to only the response set colour concepts). In contrast, participants encounter all experimental colour concepts in the non-response set pure block. The absence of exposure to the non-response colour words in the neutral and response set blocks will likely result in the increased activation of the response set colour concepts (even more so in the response set trial block since the distractor words and font colour activate response colour concepts). When the restricted set of colours is repeatedly presented without any intervening non-response set colours or non-colour words, it is likely that this smaller set of colour concepts would become more highly activated, making them more accessible and thus more likely to interfere when they are presented in the irrelevant dimension.

In mixed blocks, however, the presence of non-response set colour words would result in a greater number of colour concepts being activated in the task at any one time. In the example of Experiment 1, twice the number of colours was activated in the mixed blocks than in a pure response set trials block. With more active colour concepts, it would be harder to establish a special status or salience for any particular colours, which would result in rendering the response set colours relatively easier to inhibit when activated as the irrelevant word dimension (i.e. resulting in better performance to response set trials).

Response set effects were observed in the mixed block condition of Experiment 1 suggesting that salience is still established in the mixed block, but less so when compared to the pure block. This residual response set effect could be due to a top-down biasing or to the fact that the response set colour concepts are still the most encountered in the mixed blocks. However, we believe the response set effect was diluted by the increased number of colours that were presented in the mixed block. As noted, our account does not assume only a strategic top-down mechanism is responsible for establishing certain colour concepts as more salient. Rather, saliency is also established through exposure to concepts in the irrelevant dimension through a presumably implicitly learned, stimulus-driven process.

To test this hypothesis we manipulated the number of non-response set colours participants were exposed to, and consequently the proportion of response set to non-response set colours. Participants completed three sets of blocks: Set 1) Four pure blocks in which each trial type was presented by itself for the entire block. The four blocks comprised of either neutral trials, response set trials, non-response set trials with only two non-response set colours or non-response set trials with six non-response set colours; Set 2) Three mixed blocks where all blocks comprised of neutral trials, response set trials and non-response set trials with only two non-response set colours; Set 3) Three mixed blocks where all block comprised of neutral trials, response set trials and non-response set trials with six non-response set colours (see Appendix 1 for a full list of stimuli and conditions).

Based on the results of Experiment 1 it was predicted that the response set effect would be smaller in the mixed blocks than in the pure blocks. Based on our stimulus-driven account it was also predicted that the response set effect would be smallest in the mixed block with the larger number of non-response set colours because it would be harder to establish colour concept salience. In Experiment 1 a 17.5ms response set effect remained in the mixed block condition when there were three non-response set colour concepts presented. We expected a similar effect size in the present experiment when there were two non-response set colours, but a much reduced and even eliminated response set effect when six non-response set colours were presented. Preventing the establishment of salience of the response set colour concepts would mean that compared to the response set trials in the pure block, the response set trials in the mixed block should decrease because the response set colours would interfere less (as seen in Experiment 1). Moreover, this decrease should be more apparent in the condition with more non-response set colours. In contrast to the stimulus-driven account, an account based on a strategic biasing or flagging of colours in the relevant dimension only would predict no difference between conditions because they all contain the same number of response set colours with equal opportunity to apply that bias.

Method

Participants

40 different students (4 male, age: *M* = 19.03, *SD* = 1.12) from the same population as in Experiment 1 participated in exchange for course credit.

Apparatus and Materials

The apparatus and materials used were the same as those in the previous experiment with the only difference being an additional mixed block condition in which the number of non-response set colours was larger (6 colour-words) than in the other (2 colour-words; referred to here as *6NR* and *2NR* respectively).

Stimuli

As with Experiment 1, two versions of the experiment were administered. The response set colours were purple (204; 0; 255), yellow (255; 255; 0) and green (0; 255; 0); in one version, and white (255; 255; 255), blue (0; 112; 192), and orange (255; 127; 0) in the second version. For the non-response set trials, the irrelevant words used in the 2NR condition were ‘pink’ and ‘blue’; and ‘green’ and ‘yellow’ in the respective versions, while the 6NR contained the additional words ‘red’, ‘brown’, ‘white’, ‘orange’ for version one and ‘pink’, ’red’, ‘brown’, ‘purple’ in version two. Neutral trials were included but only to keep to the original design as closely as possible.

Procedure

Each participant completed three sets of blocks: one set of pure blocks and two separate sets of mixed blocks. The pure blocks set contained blocks of each of the four trial types (neutral, 6NR, 2NR and response set) while each set of mixed blocks consisted of three blocks of neutral, response set, and non-response set trials with either 6 or 2 non-response set colours, with each block containing an equal number of trials of each trial type randomly presented. In other words, participants went through 10 experimental blocks (4 pure blocks of neutral, 6NR, 2NR and response set trials, 3 mixed blocks of neutral, 2NR and response set trials, and 3 mixed blocks of neutral, 6NR and response set trials) with 72 trials in each block. A practice block made up of 48 trials preceded the experimental blocks, which resulted in a total of 768 trials performed by each participant. The order of the sets was counterbalanced across participants, as was the order of the trial types in the pure block format[4].

Results

Using the same criteria as Experiment 1, the total number of valid responses in the pure, mixed 2NR and mixed 6NR sets were .967 (*SD* = .022), .964 (.014) and .965 (.018) respectively. The mean RTs of each trial type are detailed in Table 3.

The magnitudes of the response set effects were calculated in the following ways: For the two mixed blocks, the effects were calculated by

[4] As before, presentation order was included in a separate omnibus test and unlike the previous experiment, the interaction was found to be significant ($F(5,34) = 2.57$, $p = .045$, $r = .265$). Although It should be noted that the number of orders to control for (6) would require a lot more participants in to have enough power to accurately determine order effect.

taking the difference between the RTs to response set trials and the corresponding non-response set trials of the block, while in the pure block set, two response set effects were obtained by taking the difference between the response set trials block and each of the two non-response set blocks. This led to four measures of the response set effect, one in each of the mixed block conditions and two in the pure block presentation condition. The response set effect in the two mixed block conditions were non-significant (6NR: $t(39) = -1.35$, $p = .185$; $r = .211$, 2NR: $t(39) = 1.28$, $p = .208$, $r = .201$), while the response set effects in the two pure block conditions were significant 6NR: $t(39) = 3.06$, $p = .004$; $r = .440$, 2NR: $t(39) = 3.50$, $p = .001$, $r = .489$).

Table 3: Mean RTs and accuracy (and SEs) of all trial types and mean and response set effects of Experiment 2. NR refers to non-response set colours. So e.g. 2NR means there were 2 non-response set colours.

	Mixed (2NR)		Mixed (6NR)		Pure	
	RTs(ms)	Accuracy(%)	RTs(ms)	Accuracy(%)	RTs(ms)	Accuracy(%)
Neutral	637.77 (13.85)	96.63 (0.355)	628.23(14.56)	96.28 (0.446)	619.01 (13.00)	96.08 (0.456)
2NR	652.19 (15.29)	96.35 (0.354)	-		629.25 (15.08)	96.74 (0.412)
6NR	-		651.59(18.74)	97.26 (0.418)	631.02 (14.29)	97.15 (0.358)
Response set	660.40 (15.64)	96.15 (0.396)	641.91(15.70)	95.87 (0.486)	667.61 (18.68)	95.83 (0.556)
Response set effects 2NR	8.21ms (6.42)				38.36*ms(10.96)	
Response set effects 6NR			-9.68ms (7.18)		36.59* ms(11.97)	

*p < .05

To determine the effect of presentation format, a one-way ANOVA on the four response set effects yielded a significant effect ($F(3,117) = 7.95$, $p < .001$, $r = .252$). Planned comparisons revealed a non-significant difference between the two response set effects in the pure blocks ($t(39) = 0.18$, $p = .855$, $r = .029$), but larger response set effects in the pure blocks compared to the corresponding response set effect in the mixed blocks (6NR: $t(39) = 3.34$, $p = .002$; $r = .472$, 2NR: $t(39) = 2.58$, $p = .014$, $r = .382$). Consistent with the findings of Experiment 1, these analyses revealed an effect of trial type mixing on the response set effect where response set effects were larger in pure blocks compared to mixed blocks.

To determine the effect of trial type mixing on non-response set effects, a one-way ANOVA on the four measures of non-response set effects (two in pure block and one each in the mixed blocks) was conducted. The analysis indicated the differences between the effects were statistically non-significant ($F(3,117) = 0.66$, $p = .581$, $r = .075$).

The effect of having different number of activated colour concepts in the irrelevant dimension was investigated by comparing the magnitude of the response set effect in the two mixed blocks. A pairwise comparison between them showed that the response set effect was larger when there were fewer non-response set colours ($t(39) = 2.62$, $p = .013$, $r = .387$). Thus, the new finding here is that as predicted, the response set effect was larger when there were fewer non-response set colours.

Separate one-way ANOVAs were conducted on the RTs of each trial type (i.e. response set, non-response set, neutral) to compare them across sets of blocks (e.g. pure, mixed 6NR). Analysis of the effects of presentation format revealed a non-significant trend for both Neutral (F(2,78) = 2.52, p = .087, r = .177) and non-response set trials (both 2NR and 6NR) trials (F(3,117) = 2.54, p = .060, r = .146), but a statistically significant effect for response-set trials (F(2,78) = 3.28, p = .043, r = .201). Pairwise comparisons within the response set trials showed only the difference between the trials in the 6NR colours pure and mixed blocks to be statistically significant (t(39) = 2.82, p = .008, r = .412). The difference between the response set trials in the two mixed blocks (2NR vs. 6NR) showed a non-significant trend (t(39) = 1.95, p = .058, r = .298); and for the mixed (2 NR colours) and pure blocks (t(39) = 0.59, p = .558, r = .094) the comparison was non-significant.

Accuracy analysis

The one-way ANOVA on the accuracy rates for the four response set effects was statistically non-significant (F(3,117) = 1.58, p = .198, r = .115)

Discussion

The results from this experiment replicated the effect of trial type mixing on the magnitude of the response set effect observed in Experiment 1. Furthermore, the results showed that the size of the response set effect is smaller when a larger number of non-response set colours is present in the irrelevant dimension, which is consistent with the notion that the magnitude of response set effect is influenced by the number of colour concepts present in the relevant and irrelevant dimension in any experimental block or at any one time. In other words the response set effect is diluted when more colour concepts are active in a block of trials. This finding is contrary to predictions made by models assuming a more strategic biasing of the response set colours. Finally, another result to note is that there was no difference in the magnitude of the non-response set effect across conditions, further supporting the notion that the effect is mainly affecting response set trials.

The comparisons of each trial type across the 2NR mixed, 6NR mixed, and pure blocks revealed that the only statistically significant effect was faster RTs to response-set trials in the 6NR mixed blocks compared to its counterpart pure block. This finding is consistent with Experiment 1 showing that the mixing effect is driven by facilitation of responses to response set trials when presented in mixed blocks. This we have argued is due to the reduced ability to establish the salience of the response set colours in mixed blocks which would reduce the resting activity level of those colours, rendering them easier to ignore when in the irrelevant dimension. However, despite there being a reduced response set effect when there were only 2 non-response set colours, this was not driven by a reduction in RTs to response set trials. A numerical decrease was observed to the response set trials in the 2NR set (660ms vs. 668ms), but this decrease was not significant. This finding presents a challenge to the stimulus-driven account presented here since it was predicted that the reduction in the response set effect would be driven by a reduction in RTs to response set trials. Notably however, when taking neutral trial RTs into account the results do not seem so inconsistent with our argument. The neutral trial RTs in the mixed 2NR condition are almost 18ms slower than in the neutral trials in the pure condition. The reasons for this increase are unclear but could be due to general mixing effects (there is a similar, but smaller increase in neutral trial RTs in the mixed 6NR condition). Whatever the reason for this increase, its implications are important since it suggests that all trial types in the mixed 2NR condition could be inflated. If this were the case and one wanted to remove the influence of the errant effect one would subtract this increase from the RT to the response set trials in the same condition. Doing this reduces the RT to those response set trials to ~642ms, a figure that is ~26ms shorter than its counterpart in the pure condition, and a figure thus is consistent with our stimulus-driven account.

As in Experiment 1 there is a numerical increase in the RTs to the non-response set trials in both the 2NR and 6NR conditions. However, also consistent with Experiment 1 is that the increases were not significant (for either the 2NR or 6NR condition), and thus cannot be used to draw conclusions. This issue will be further discussed in the General Discussion below in the context of mixing effects in other literatures.

General Discussion

The experiments in this study set out to investigate the effect of presentation format on response set effects in the Stroop task. Data from both experiments showed response set effect to be smaller and even statistically eliminated when the trials were presented in mixed blocks. Although only response set trials were significantly affected by the mixing effect, the overall pattern of results in both experiments are consistent with the notion that there was some effect of trial type mixing on all trial types (see ‘Alternative accounts of trial type mixing effects’ section below for a deeper discussion of this issue),.

Experiment 2 was conducted to test a new account of the response set effect based on a stimulus-driven mechanism that processes the number of colour concepts activated in the task. This account contrasted with the putative account of the response set effect based on strategic, selective biasing or flagging of the colours in the relevant dimension only. The new account was tested by varying the number non-response set colours under the assumption that the more colour concepts presented the harder it would be for any particular colour to rise to prominence or gain special status. The results showed a smaller response set effect in the mixed condition that had more non-response set colours. It should be noted that since only the number of non-response colour words were manipulated, the results do not allow us to identify whether the effect is limited to variation in the number of non-response set colour concepts or whether manipulating the number of activated colour concepts in either the response set or non-response would have the same effect. However, increasing the number of items in the response set would burden working memory capacity, likely adding to task difficulty and therefore not representing as pure a test of the effect under study as the manipulation of the irrelevant dimension.

The present research offers important insights into the processes involved in the mechanisms of selective attention. Our results suggest that response set effects are not just the result of the ability to better ignore colour concepts that have not been identified as task relevant via a fixed, pre-set top-down bias or flagging. Although being part of the response set makes a distractor more difficult to inhibit, as shown by response set trials having slower RTs compared to non-response set trials in pure blocks, the amount of interference is modulated by the number of non-response set colours in the same block. If task relevant colours were somehow identified and fixed according to task instructions or even after a few trials, there would be no effect of presentation format. Our results provide some evidence of the influence of bottom-up processes in performing the Stroop task that popular top-down models ignore.

The results observed in the current study do not conform to those of Lamers et al. (2010) who showed the benefits of previewing the distractor dimension of a Stroop stimulus. In their study’s second experiment, they showed that previewing the irrelevant colour word facilitates RTs to incongruent (response set) trials, indicating a benefit to pre-processing the irrelevant dimension. Increasing the number of non-response set colours in mixed blocks as we have done would decrease the predictability of the irrelevant colour word, and yet we showed that this also results in decreased RTs to response set trials. However, Lamers et al. (2010) cued the irrelevant colour word 2000 ms prior to target presentation, which would give participants the chance to inhibit the irrelevant word by the time the Stroop stimulus appeared. Hence this is an entirely different manipulation from the one used in the present work.

Another implication of the present work for current models of Stroop interference is that our experimental manipulation did not significantly affect the magnitude of the non-response set effect (non-response set trial RTs – Neutral trial RTs). This finding is inconsistent with predictions from models suggesting that response set and non-response set trials should be affected in tandem (Roelofs, 2003). However, this finding is a null result and thus should be interpreted with caution. See below for a fuller discussion of this issue.

The results from the present experiments also run counter to predictions from the proactive control literature that investigates proportion congruency effects. In such studies (e.g. Kane & Engle, 2003; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979; Lowe & Mitterer, 1982; West & Baylis, 1998) the proportion of incongruent and congruent trials within blocks of trials are manipulated (i.e. mostly congruent trials vs. mostly incongruent trials) and typically report larger Stroop interference when there are fewer incongruent trials. This effect was also observed in studies where proportion of neutral and incongruent trials were manipulated (e.g. Goldfarb & Henik, 2013;Tzelgov, Henik, & Berger, 1992). It is thought that when there are more incongruent trials, participants tend to strategically focus more on the colour dimension of the Stroop stimulus to provide information about the correct response. A proactive control account would predict that it would be easier to strategically allocate attention in pure blocks to either the relevant dimension (incongruent block) or irrelevant dimension (congruent block) to deal with conflict. One prediction from this account of the proportion congruency effect is that RTs to incongruent trials should be shorter when presented in pure blocks than when presented in mixed blocks; a prediction not supported by the findings from the present study.

Another account of the proportion congruency effect holds that when there is a greater proportion of incongruent trials reaction times are likely to benefit from conflict adaptation (Botvinick et al., 2001). Conflict adaptation, a form of reactive control, refers to when conflict on trial n is reduced because conflict on trial n-1 has primed the system to better deal with conflict on the next trial. Hence, the conflict adaptation account would also predict smaller

RTs to incongruent trials presented in pure blocks; a prediction that contrasts with the present results (see also Egner, 2014; Mordkoff, 2012; Schmidt & Besner, 2008).

Our data do not permit us to draw strong conclusions as to why the data contradict predictions based on congruency sequence effects. However, even if incongruent trial reaction times were benefitting from proactive or reactive control mechanisms in the pure blocks, our data would show that the mixed block benefits from a large reduction in conflict that is greater than benefits from other control mechanisms, resulting in reduced RTs in the mixed blocks. Furthermore, one might question how different the two conditions are in terms of likely benefits from congruency proportion effects; the mixed blocks still mainly comprise conflict-inducing trials with only a small proportion being non-conflict neutral trials. Finally, it should also be noted that proportion congruency effect has been argued to be due response contingency (see Schmidt & Besner, 2008; Schmidt, Crump, Cheesman, & Besner, 2007; for an in depth discussion). Response contingency refers to the incidental learning of associations between irrelevant words and response colours when these associations occur more often for certain combinations. When trials are mostly congruent, participants implicitly learn that, for example, the word *red* is most often associated with the colour red. When exceptions to that association are presented (e.g. the word *red* in blue), it is harder to respond correctly to the stimulus colour, pushing up RTs to incongruent trials and thus the Stroop effect compared to when incongruent trials are more frequent. The contingency effect in the present study has been controlled for (see Appendix

1) and thus it might not be surprising that the results are not consistent with the proportion congruency effect.

A reverse response set effect?

The argument presented in the present work is that colour concepts in the irrelevant dimension contribute to colour concept salience computation such that more non-response set colours that are present the less likely it is that a response set will be established. A question that remains is whether the non-response set could in effect steal the salience from the response set colours such that they are responded to more slowly that the response set colours, producing a reverse response set effect. This possibility is suggested by the 6NR condition where non-response set trials are responded to more slowly than response set trials, albeit non-significantly. There are questions that have to be answered when arguing that non-response set trials can become a more activated set (steal salience) when they are greater in number than response set trials. For example, how it is they become perceived as a set? It can only be by contrast with the response set which means that the response set needs to be established in some way. Importantly, we are not arguing that there is no top-down involvement in establishing the response set, just that the irrelevant dimension influences the process of establishing response set salience to the point that it can result in eliminated response set effects (as in Experiment 2). Furthermore, it is difficult to argue that the non-response set colour concepts could compete with the response set colour concepts for salience given the latter are encountered on every trial. The only way that this could happen is if salience is computed from the irrelevant dimension only. Our data do not allow us to make such a strong conclusion at this stage.

If it were posited that only words in the irrelevant dimension contribute to salience computation (i.e. a completely stimulus-driven account of response set effects), then since there are more non-response set colours (6) than response set colours (3) in the 6NR condition, the non-response set colours would be more salient (if perceived as a set) and hence RTs would increase and should produce a reversed response set effect which we did not observe statistically. This would not happen in NR2 because there are too few of the NR colours (2) to compete; the RS colours would still be greater in number, if only by a small amount (1). This would predict that RTs to non-response set trials in the NR2 and NR6 conditions would differ, but they do not appear to. However, the baseline neutral trial differs between the two NR conditions and the pure condition. This could account for why we see no difference between the non-response set trials in 2NR and 6NR – any difference might be hidden by changes in the neutral trials. These differences in neutral trial RTs could be due to a more general mixing effect that was not predicted or just random differences. Nevertheless, the change in the neutral trial RTs across the conditions means that we must draw conclusions based on difference values and not absolute values when comparing conditions. Thus, our data suggest the possibility of a reverse response set effect. Future research should address this interesting possibility, as it would extend the stimulus-driven account of response set effects presented here. An alternative possibility is that non-response set colours do not steal salience. Without the extra salience the only source of conflict is semantic which would be nearing equality between the two trial types. So under this account RTs to response set trials should decrease but there would be no effect on non-response set trials.

Alternative accounts of the observed effect of trial type mixing

Before drawing conclusions we must also consider other potential accounts of the effect. Mixing effects have been observed in other literatures that might be able to account for the effect of trial type mixing on the response set effect. Two general patterns of results have been reported, both describing the size of an effect being smaller in mixed blocks compared to pure blocks (referred to as *mixing effect);* thus similar to what has been observed here. One way to determine if a mixing cost can explain the difference in performance between our presentation formats is to observe an asymmetry in mixing cost. This is when a relatively slow trial is not slowed down as much as a faster trial in mixed blocks, which would result in effects being smaller in mixed blocks compared to pure blocks and is the reason smaller effects are observed in the former (Los, 1996). In the present context this would be represented by the response set trials being slowed in the mixed block relative to the pure block condition but not as much as non-response set or neutral baseline trials. Hence the response set effect (response set – non-response set trials) would be smaller in the mixed block as a result of an increase in RTs to non-response set trials. This account can be immediately abandoned since response set trials were speeded up, not slowed down and the trial types were not affected in the same direction.

Another observation from studies presenting trials in mixed and pure blocks is the effect of *homogenisation* described by Lupker, Brown and Colombo (1997) and Lupker, Kinoshita, Coltheart and Taylor (2003) in research on word naming. Unlike the mixing cost described above, compared to pure blocks, the RTs of trials in mixed blocks tend to move towards the overall mean RT of the different trial types in the block (i.e. the slower trials become faster while faster trials become slower). This effect is driven by the averaging of the threshold for the decision making process towards the mean of the all trial types in each block (see Lupker et al., 2003 for a more comprehensive explanation), which results in the RT of the faster trials increasing while the RTs of slower trials decrease in the mixed blocks. In the present context such an account would find support if RTs to response set trials decrease whilst RTs to non-response set trials increase. As noted above, we have observed a pattern similar to this in the present data. However, whilst there is a good reason to consider this a potential modulating factor (i.e. the overall RT patterns), there are equally good reasons as to why this cannot be the only process responsible for the effects observed. First, the only trial type to be significantly affected by mixing in both experiments was response set trials. Whilst there were trends for affects for the other trial types, these did not reach significance. Second, this pattern is more apparent in Experiment 2, than Experiment 1. If the effect were based solely on this there is no good reason why it should be more apparent in Experiment 2. In Experiment 1 RTs to neutral trials actually decrease (non-significantly) in the mixed block, which is contrary to homogenisation predictions. In Experiment 2, neutral trial RTs increase (non-significantly) which means that the best summary of the effect of mixing on neutral trials is that on average they are not affected by trial type mixing. There is no reason according the homogenisation account why one stimulus type would not be affected by the mixing. However, for the sake of argument one might consider the homogenisation account to predict that the mixing effect to be limited to category-similar stimuli (e.g those with colour words in the irrelevant dimension), which leads on to the third point: Homogenisation predicts a response threshold change based on average response times for each trial type. The 2NR and 6NR non-response set trial RTs are almost identical, but despite this, only in the 6NR condition was the RT to the response set trials significantly different from that in the pure block. This larger effect of trial type mixing on response set trials in the 6NR condition is only predicted by the stimulus-driven account presented here.

Of course, the above reasoning could also be taken as a challenge to the stimulus-driven account presented here. For example, the stimulus-driven account of response set effects as presented above makes no claims about neutral trials being unaffected by the mixing. The neutral words (e.g. marvel, story) are concepts that could potentially dilute the response set effect. Future research will need to test this aspect of the theory. It might well be that dilution or shared salience can only occur within the goal-relevant category (e.g. colour or colour related words). It is clear however that the homogenisation account, based on modifying task-wide response threshold, is more tied to the notion that all trial types would be affected. Arguably, the present evidence is best interpreted as contrasting with the homogenisation account. Finally, it is important to reiterate that the only significant effect on a trial type observed across both experiments was on the response set trials; an effect consistent only with the stimulus-driven account of response set effects. The fact that this is not observed in the 2NR condition of Experiment 2 is potentially problematic for the stimulus-driven account as stated earlier, and is an issue that needs to be addressed by future research, but the effect of mixing on the neutral trial baseline must be considered before drawing any conclusions. Doing so suggests a relative decrease in RTs to response set trials in the 2NR condition.

Implications for the debate on the automaticity of reading

The inability to prevent the irrelevant colour word from interfering with colour naming has been taken as evidence for word reading being an automatic (happening without intent and not requiring attentional resources) and ballistic (cannot be stopped once started) (Brown, Gore & Carr, 2002; Neely & Kahan, 2001; and Posner & Snyder, 1975). However, the demonstration that Stroop interference can be reduced using manipulations such as the narrowing of spatial attention (e.g. Besner, 2001; Besner, Risko, & Sklair, 2005; Besner, Stolz, & Boutilier, 1997; Labuschagne & Besner, 2015, Stolz & McCann, 2000) social priming (Goldfarb, Aisenberg, & Henik, 2011) and a post-hypnotic suggestion (e.g. MacLeod & Sheehan, 2003; Parris, Dienes & Hodgson, 2012; Raz & Campbell, 2011; Raz, Moreno- Iñiguez, Martin, & Zhu, 2007; Raz, Kirsch, Pollard, & Nitkin-Kaner, 2006; Raz et al., 2003; Raz, Sharipo, Fan & Posner, 2002) has been taken as evidence against the notion of that word reading is automatic.

In their reviews of the studies above, Augustinova and Ferrand (2014a) and Flaudias and Llorca (2014) pointed out that Stroop interference is made up of both semantic and response based processes. Augustinova and Ferrand (2014a) argued that only the former is assumed to be automatic, and as such studies need to show that their manipulations affect semantic processes before a claim for control over ‘automatic’ processes can be made. They also argued that the use of manual responses, which are the norm for such studies, is not appropriate for measuring semantic processes since they have been shown to mainly index response conflict in the Stroop task (Sharma & McKenna, 1998; see also Kinoshita, De Wit, & Norris, 2016, for evidence showing that only vocal responses result in interference from the lexical properties of the irrelevant stimulus). Therefore, they argued, that even when these studies showed an elimination of Stroop interference, they were unlikely to have demonstrated a reduction in semantic processing and instead were affecting response conflict processes only. Instead in a series of studies Augustinova and Ferrand have shown that semantic interference appears to be unaffected by these experimental manipulations (Augustinova & Ferrand, 2012a; 2012b; 2014b; Ferrand & Augustinova, 2014). To show this they employed semantic-associative Stroop trials to index semantic conflict. Semantic-associative Stroop trials are trials on which the irrelevant word is semantically or associatively related to a colour e.g. *sky*, which is related to the colour blue, presented in green. They have shown that for example neither narrowing spatial attention, social priming nor the word blindness post-hypnotic suggestion affect the magnitude of the semantic-associative Stroop effect (semantic-associative Stroop trial RT – neutral trial RT). However, all of their studies involved the use of the mixed trial type presentation format.

The findings from the present research suggest that a significant amount semantic conflict is involved in manual response Stroop tasks, especially in mixed blocks, findings that are consistent with those from Brown and Besner (2001) who presented a reanalysis of the Sharma and McKenna (1998) paper on which Augustinova and Ferrand's argument is based. Moreover, if one accepts the stimulus-driven account presented here, and perhaps even if not, given that the response set effect is the key index of response competition, the present results suggest that the contribution of response competition is substantially reduced (Experiment 1) and even eliminated (Experiment 2) when trial types are mixed. Thus, by measuring response conflict as the difference between incongruent (response set) trials and semantic-associative trials when in a mixed trial type context, it is likely that at least half of the interference identified as response conflict is likely to actually be semantic conflict. Evidence of a large reduction in Stroop interference would therefore represent a concomitant reduction in semantic conflict. In short we are suggesting that it is conceivable that semantic-associative Stroop trials do not fully capture all of the semantic conflict that is present in the Stroop task. Furthermore, we are suggesting that semantic conflict not indexed by semantic-associative trials, as measured by the difference between non-response set trial and semantic-associative trials, might reveal that the previously mentioned manipulations might in fact affect semantic conflict (see Sharma & McKenna, 1998, explaining how non-response set trials includes additional semantic competition). Any claims that semantic conflict is not affected by experimental manipulations could therefore potentially be undermined. Future research in this area could employ pure block presentation format to confirm their findings. Notably they have often employed vocal responses instead of the manual response employed here, and the effect of trial type mixing on the response set effect has yet to be shown using vocal responses.

Conclusion

By demonstrating the modulation of the response set effect under conditions of trial type mixing we have presented evidence that is difficult for extant models that proffer an account of response set effects based on strategic, top-down biasing of attention to explain. Response set effects have heretofore been considered a well-established component of Stroop interference, but we have presented evidence for the complete elimination of the response set effect under certain conditions. We have presented and tested an alternative stimulus-driven account of the response set effect that is the best fit to the data compared to alternative accounts discussed. We have argued that response sets are established by computing relevant and irrelevant perceptual components, and that irrelevant components can, somewhat ironically, dilute those selective attention mechanisms responsible for facilitating goal-oriented behaviour. The mere computation of this irrelevant content represents a failure of selective attention indicating it is not the result of optimal selective mechanisms. This finding it is not however necessarily consistent with automatic lexical access. Rather, the finding can be explained as a consequence of a mechanism computing goal-related, if not goal-relevant information; a suggestion that is consistent with theoretical positions arguing that the goal of the task is the ultimate arbiter of what is processed and what is not (Roelofs, 2003; Kinoshita et al., 2016). Finally, we have proposed further ways to test and extend the theory and have highlighted the implications of these findings for automaticity of reading debates.

References

Augustinova, M., & Ferrand, L. (2012a). Suggestion does not de-automatize word reading: Evidence from the semantically based Stroop task*. Psychonomic Bulletin & Review,19*(3), 521-527.

Augustinova, M., & Ferrand, L. (2012b). The influence of mere social presence on Stroop interference: New evidence from the semantically-based Stroop task. *Journal of Experimental Social Psychology*, *48*(5), 1213-1216.

Augustinova, M., & Ferrand, L. (2014a). Automaticity of Word Reading Evidence From the Semantic Stroop Paradigm. *Current Directions in Psychological Science*, *23*(5), 343-348.

Augustinova, M., & Ferrand, L. (2014b). Social priming of dyslexia and reduction of the Stroop effect: What component of the Stroop effect is actually reduced?. *Cognition*, *130*(3), 442-454.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459.

Besner, D. (2001). The myth of ballistic processing: Evidence from Stroop's paradigm*. Psychonomic Bulletin & Review, 8(*2), 324-330.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4*(2), 221-225.

Besner, D., Risko, E. F., & Sklair, N. (2005). Spatial attention as a necessary preliminary to early processes in reading. *Canadian Journal of Experimental Psychology, 59*(2), 99-108.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624 - 652.

Brown, M., & Besner, D. (2001). On a variant of Stroop’s paradigm: Which cognitions press your buttons? *Memory & Cognition, 29*(6), 903-904.

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*(2), 220-240.

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: Does the response set matter? *Cognition*, *75*(2), B51-B64.

Caramazza, A., & Costa, A. (2001). Set size and repetition in the picture–word interference paradigm: Implications for models of naming. *Cognition*, *80*(3), 291-298.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332.

Cohen, J.D. & Huston, T.A. (1994) Progress in the use of interactive models for understanding attention and performance. In C. Umilta & M. Moscovitch (Eds.) *Attention and Performance XV* pp. 453–456, Cambridge, MA: MIT Press.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

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

Ferrand, L., & Augustinova, M. (2014). Differential effects of viewing positions on standard versus semantic Stroop interference. *Psychonomic bulletin & review*, *21*(2), 425-431.

Flaudias, V., & Llorca, P. M. (2014). A brief review of three manipulations of the Stroop task focusing on the automaticity of semantic access. *Psychologica Belgica, 54*(2), 199-221.

Goldfarb, L., Aisenberg, D., & Henik, A. (2011). Think the thought, walk the walk–Social priming reduces the Stroop effect. *Cognition, 118*(2), 193-200.

Goldfarb, L., & Henik, A. (2006). New data analysis of the Stroop matching task calls for a reevaluation of theory. *Psychological Science, 17*(2), 96-100.

Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology, 5*:1157.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83*(2p1), 299-303.

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*(1), 47-70.

Kinoshita, S., De Wit, B., & Norris, D. (2016, September 22). The Magic of Words Reconsidered: Investigating the Automaticity of Reading Color-Neutral Words in the Stroop Task. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication. http://dx.doi.org/10.1037/xlm0000311

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology, 77*(4), 576-588.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400-410.

Labuschagne, E. M., & Besner, D. (2015). Automaticity revisited: when print doesn't activate semantics. *Frontiers in Psychology, 6*:117.

Lamers, M. J., Roelofs, A., & Rabeling-Keus, I. M. (2010). Selective attention and response set in the Stroop task. *Memory & cognition*, *38*(7), 893-904.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 219 - 234.

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*(3), 166-174.

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, *94*(2), 145-188.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology, 36*(4), 684 - 700.

Luo, C. R. (1999). Semantic competition as the basis of Stroop interference: Evidence from color-word matching tasks. *Psychological Science, 10*(1), 35-40.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 570.

Lupker, S. J., Kinoshita, S., Coltheart, M., & Taylor, T. E. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language*, *49*(4), 556-575.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163-203.

MacLeod, C. M. (2005). The Stroop task in cognitive research. *Cognitive Methods and Their Application to Clinical Research*, 17-40.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126-135.

MacLeod, C. M., & Sheehan, P. W. (2003). Hypnotic control of attention in the Stroop task: A historical footnote. *Consciousness and cognition, 12*(3), 347-353.

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, *12*(3), 467-473.

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*(4), 750-757.

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H.L. Roediger, J.S. Nairne, I. Neath, & A.M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69–93). Washington, DC: American Psychological Association.

Parris, B. A. (2014). Task conflict in the Stroop task: When Stroop interference decreases as Stroop facilitation increases in a low task conflict context. *Frontiers in Psychology*, *5*: 1182.

Parris, B. A., Dienes, Z., & Hodgson, T. L. (2012). Temporal constraints of the word blindness posthypnotic suggestion on Stroop task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 833-837.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. *Attention and Performance V*, 669-682.

Proctor, R. W. (1978). Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, *23*(5), 413-419.

Raz, A., & Campbell, N. K. (2011). Can suggestion obviate reading? Supplementing primary Stroop evidence with exploratory negative priming analyses. *Consciousness and Cognition, 20*(2), 312-320.

Raz, A., Moreno-Íniguez, M., Martin, L., & Zhu, H. (2007). Suggestion overrides the Stroop effect in highly hypnotizable individuals. *Consciousness and Cognition, 16*(2), 331-338.

Raz, A., Kirsch, I., Pollard, J., & Nitkin-Kaner, Y. (2006). Suggestion reduces the Stroop effect. *Psychological Science, 17*(2), 91-95.

Raz, A., Landzberg, K. S., Schweizer, H. R., Zephrani, Z. R., Shapiro, T., Fan, J., & Posner, M. I. (2003). Posthypnotic suggestion and the modulation of Stroop interference under cycloplegia. *Consciousness and Cognition, 12*(3), 332-346.

Raz, A., Shapiro, T., Fan, J., & Posner, M. I. (2002). Hypnotic suggestion and the modulation of Stroop interference. *Archives of General Psychiatry, 59*(12), 1155-1161.

Risko, E. F., Schmidt, J. R., & Besner, D. (2006). Filling a gap in the semantic gradient: Color associates and response set effects in the Stroop task. *Psychonomic Bulletin & Review*, *13*(2), 310-315.

Roelofs, A. (2001). Set size and repetition matter: Comment on Caramazza and Costa (2000). *Cognition*, *80*(3), 283-290.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88-125.

Scheibe, K. E., Shaver, P. R., & Carrier, S. C. (1967). Color association values and response interference on variants of the Stroop test. *Acta Psychologica*, *26*, 286-295.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition, 26*(5), 1033-1040.

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*(3), 514-523.

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *The Quarterly Journal of Experimental Psychology*, *31*(1), 121-132.

Stolz, J. A., & McCann, R. S. (2000). Visual word recognition: Reattending to the role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*(4), 1320-1331.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. *Neuroimage, 27*(3), 497-504.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging, 13*(2), 206-217.

West, R., Bowry, R., & McConville, C. (2004). Sensitivity of medial frontal cortex to response and nonresponse conflict. *Psychophysiology*, *41*(5), 739-748

Appendix 1: Breakdown of the number of word stimuli (leftmost column) presented in each colour and the resultant proportion contingency.
(note that counterbalanced versions used different colour and word stimuli but proportions remain the same)

Experiment 1	Number of trials			contingency (%)		
	YELLOW	PINK	GREEN	button 1	button 2	button 3
yellow	0	16	16	0	50	50
pink	16	0	16	50	0	50
green	16	16	0	50	50	0
blue	0	16	16	0	50	50
purple	16	0	16	50	0	50
orange	16	16	0	50	50	0
wall	0	16	16	0	50	50
marvel	16	0	16	50	0	50
story	16	16	0	50	50	0

Experiment 2 (2NR)						
	YELLOW	PURPLE	GREEN	button 1	button 2	button 3
yellow	0	12	12	0	50	50
purple	12	0	12	50	0	50
green	12	12	0	50	50	0
pink	12	12	12	33	33	33
blue	12	12	12	33	33	33

Experiment 2 (6NR)						
	YELLOW	PURPLE	GREEN	button 1	button 2	button 3
yellow	0	12	12	0	50	50
purple	12	0	12	50	0	50
green	12	12	0	50	50	0
pink	3	3	3	33	33	33
blue	3	3	3	33	33	33
red	3	3	3	33	33	33
brown	3	3	3	33	33	33
white	3	3	3	33	33	33
orange	3	3	3	33	33	33