The Stroop Effect: Why Proportion Congruent has Nothing to do with Congruency and Everything to do with Contingency

by

James R. Schmidt

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2007

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

Participants are slower to identify the print colour of incongruent colour words (e.g., the word ORANGE printed in green) than of congruent colour words (e.g., ORANGE printed in orange). The difference in time between these two conditions is the Stroop effect. The item-specific proportion congruent (ISPC) effect is the observation that the Stroop effect is larger for words that are presented mostly in congruent colours (e.g., BLUE presented 75% of the time in blue), and smaller for words that are presented mostly in a given incongruent colour (e.g., YELLOW presented 75% of the time in orange). One account of the ISPC effect, the modulation hypothesis, is that participants use the distracting word to modulate attention to the word (i.e., participants allow the word to influence responding when it is presented mostly in its congruent colour). Another account, the contingency hypothesis, is that participants use the word to predict the response that they will need to make (e.g., if the word is YELLOW, then the response is probably orange). Reanalyses of data from Jacoby, Lindsay, and Hessels (2003) along with results from new experiments are inconsistent with the modulation hypothesis, but entirely consistent with the contingency hypothesis. A mechanistic account of how responses are predicted is generated from the contingency hypothesis.

Acknowledgements

This research was supported by an NSERC Canada Graduate Scholarship. I thank Derek Besner for his thoughtful comments on earlier versions of this thesis, and Larry L. Jacoby and Carole Jacoby for providing the raw data from Jacoby et al. (2003). Correspondence concerning this thesis should be addressed to James R. Schmidt, Department of Psychology, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1. E-mail: j4schmid@watarts.uwaterloo.ca.

Table of Contents

List of Figures

The Stroop Effect: Why Proportion Congruent has Nothing

to do with Congruency and Everything to do with Contingency

Contingency learning shows signs of becoming a lively area of research (e.g., Jacoby, Lindsay, & Hessels, 2003; Musen & Squire, 1993; Schmidt, Crump, Cheesman, & Besner, 2007). One popular domain for this research is the Stroop paradigm. The standard Stroop effect is the finding that participants take longer to identify the print colour of an incongruent colour word (e.g., the word GREEN printed in red; $\text{GREEN}_{\text{red}}$) relative to a congruent colour word (RED_{red}; Stroop, 1935; see MacLeod, 1991, for a review). The magnitude of the Stroop effect changes when the proportion of congruent items is manipulated (Lowe & Mitterer, 1982). In particular, the Stroop effect increases as the proportion on congruent trials increases. The standard account of the influence of proportion congruent (Cheesman & Merikle, 1986; Lindsay & Jacoby, 1994; Lowe & Mitterer, 1982) is that the detection of these proportions allows participants to modulate attention to the word, thereby changing the size of the Stroop effect. However, Schmidt et al. (2007) suggest that simple *contingency learning* provides a sufficient account of the proportion congruency effect.

Here, I provide a reanalysis of the Jacoby et al. (2003) data and report new results from my own laboratory that are consistent with the Schmidt et al. (2007) suggestion and at the same time are problematic for the modulation hypothesis. Possible mechanisms by which participants use contingency information to control responding will also be considered.

Proportion Congruency and the Modulation Hypothesis

A number of experiments (Cheesman & Merikle, 1986; Lindsay & Jacoby, 1994; Lowe & Mitterer, 1982) have shown that the magnitude of the Stroop effect can be modulated by varying the proportion of congruent trials. Specifically, the Stroop effect is larger when most of the items in the experiment are congruent (high proportion congruent) than when most of the items are incongruent (low proportion congruent). The standard explanation of this effect (Cheesman & Merikle, 1986; Lindsay & Jacoby, 1994; Lowe & Mitterer, 1982), here termed the *modulation hypothesis*, is that participants strategically modulate attention to the word depending on the proportion of congruent items. For instance, when the word and colour match most of the time (high proportion congruent), participants attend to the word more than usual. This will speed up responses on congruent trials (increased facilitation) and slow down responses on incongruent trials (increased interference), making for a larger Stroop effect. In contrast, when the word and the colour mismatch most of the time (low proportion congruent), participants make a greater effort to ignore the word. This decreases both facilitation from congruent words and interference from incongruent words, making for a smaller Stroop effect. Thus, the modulation account holds that participants use information about proportion congruency to decide the degree to which they will attend to the word and thus to allow the word to impact performance in colour identification.

The modulation hypothesis seems to be the most widely accepted explanation for the proportion congruent effect. For instance, Lowe and Mitterer (1982) conclude that their findings "demonstrate the strategic modulation of selective attention" (p. 698) and Lindsay and Jacoby (1994) state that the effect "suggests that when most items are incongruent, subjects somehow inhibit the influence of word-reading processes, relative to when most items are congruent" (p. 225).

Although intuitively appealing, the modulation hypothesis has difficulty explaining some findings. Given the assumption that participants modulate attention to the word, it would be expected that this modulation would be task-wide. That is, words should be ignored throughout a low proportion congruent block and attended to throughout a high proportion congruent block. However, Jacoby et al. (2003) manipulated proportion congruency for each item (i.e., each colour word) such that some words were presented most often in their congruent colour (e.g., $\text{BLUE}_{\text{blue}}$) and other words were presented most often in a particular incongruent colour (e.g., $\text{ORANGE}_{\text{yellow}}$). A proportion congruent effect was still observed, even though high and low proportion congruent stimuli were intermixed. Jacoby et al. called this finding the item-specific proportion congruent (ISPC) effect. As Jacoby et al. point out, this finding is difficult to accommodate within the modulation hypothesis framework, because it would have to be assumed that participants are modulating attention to the word on a trial-by-trial basis depending on the *identity* of the word (e.g., if the word is BLUE, then the word is attended, but if the word is ORANGE, then the word is ignored). In essence, to defend the modulation account it would have to be maintained that participants decide whether to attend to the word after they have already read it.

Response Prediction and the Contingency Hypothesis

A different account of the ISPC effect, here termed the *contingency hypothesis*, is that participants implicitly learn contingencies (i.e., correlations) between words and responses and then use said contingencies to predict the *specific response* associated with each distracting word (Schmidt et al., 2007). For instance, if the word ORANGE is presented most often in yellow, then upon processing the word ORANGE participants will (unconsciously) predict that the correct response will be yellow. This response prediction allows participants to shortcut some processing (thus speeding responding) if the predicted response is indeed the correct one. When a word accurately predicts the correct response (e.g., $\text{BLUE}_{\text{blue}}$, where BLUE is presented most often in blue), this is called a *high contingency trial.* When the word predicts the wrong response (e.g., $BLUE_{green}$), this is called a *low contingency trial.* When the word does not predict a specific response (e.g., $PINK_{brown}$, where PINK is presented equally often in all colours), this is called a *medium contingency trial.*

Although the modulation hypothesis has difficulty explaining the ISPC effect, the contingency hypothesis does not. According to the contingency hypothesis, in the high proportion congruent condition responses to congruent trials will be faster than usual, because participants can use the word to successfully predict the response (e.g., $BLUE_{blue}$; high contingency). The same advantage does not occur for incongruent trials in this condition because, for instance, BLUE does not accurately predict an incongruent green response (low contingency). Thus, by speeding congruent but not incongruent trials, the difference between incongruent and congruent trials (the Stroop effect) will be larger (i.e., relative to a condition where words are not predictive of responses). Similarly, in the low proportion congruent condition responses will be faster for *incongruent* items, because participants can use the word to successfully predict the response (e.g., $ORANGE_{yellow}$; high contingency). The same advantage does not occur for congruent trials in this condition because, for instance, ORANGE predicts a yellow, not an orange response (low contingency). Thus, by speeding incongruent but not congruent trials, the overall Stroop effect will be smaller.

According to the contingency hypothesis, then, proportion congruency manipulations are confounded with contingency. Specifically, in the high proportion congruent condition the magnitude of the Stroop effect is inflated due to confounding higher word-response contingencies for the congruent (.75) relative to incongruent (.25) items. Similarly, in the low proportion congruent condition the Stroop effect is reduced due to confounding higher word-response contingencies for the incongruent (.75) relative to congruent (.25) items. Fixing this confound is as simple as rearranging the cells in the design. Rather than comparing *high contingency* congruent trials with *low contingency* incongruent trials (the high proportion congruent condition), we can compare *high contingency* congruent trials (from the high proportion congruent condition) with *high contingency* incongruent trials (from the low proportion congruent condition). Similarly, rather than comparing *low contingency* congruent trials with *high contingency* incongruent trials (the low proportion congruent condition), we can compare *low contingency* congruent trials (from the low proportion congruent condition) with *low contingency* incongruent trials (from the high proportion congruent condition). After making this adjustment, the contingency hypothesis predicts a main effect of Stroop trial type (congruent, incongruent), a main effect of contingency (high, medium, low), and, more critically, *no interaction* between the two. In contrast, the modulation hypothesis predicts an interaction, because incongruent trials should be more affected by attention given that the majority of the Stroop effect is interference, with little or no facilitation from congruent trials (see MacLeod, 1991, for a review).

Facilitation and Interference

The contingency hypothesis claims that participants use response prediction to speed responses on high contingency trials. The specific mechanism that can be proposed for explaining how this occurs is that participants prepare for a response by lowering the threshold for the expected response. As shown in Figure 1, for instance, if the word ORANGE is presented most often in yellow, then presentation of the word ORANGE will lead the participant to reduce the threshold for a yellow key response. Consequently, it will take less activation of this potential response for it to be made. This mechanism therefore predicts response time *facilitation* on high contingency trials (where the predicted response is correct) relative to medium contingency trials (where no prediction is made). However, no interference for response time is expected on low contingency trials (where the predicted response is incorrect) relative to medium contingency trials, because on both low and medium contingency trials the response threshold for the *correct response* is at the normal level. Thus, the contingency hypothesis predicts response facilitation, but not response interference for response times.

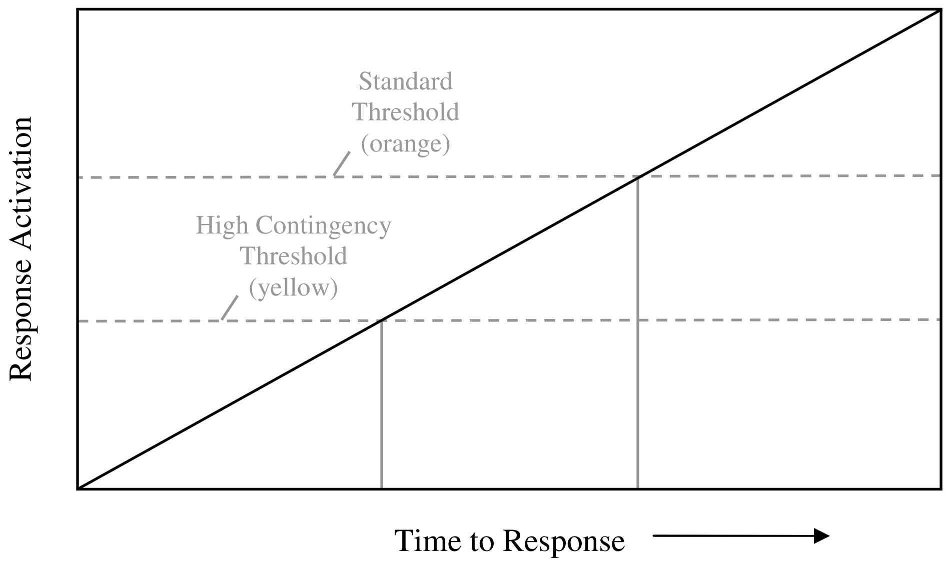

Figure 1. Response threshold model of contingency effects. If ORANGE is presented most often in yellow, then when ORANGE is presented the threshold for the yellow response will be lower than the threshold for all other potential responses (e.g., orange).

It is important to articulate why this account does not predict interference on low contingency trials. On a low contingency trial (e.g., $\text{BLUE}_{\text{green}}$), the distracting word is predictive of a specific response (i.e., BLUE predicts a blue key response). Thus, the threshold for this predicted response will be lowered. However, the response threshold for the remaining colours (green, yellow, and orange) will be unaltered. Because the *correct* response (green) is not the predicted response (blue), *correct* response latencies will not be speeded.

Summary

The first goal of the present investigation is to reanalyze the Jacoby et al. (2003) ISPC

Stroop experiment to test predictions derived from the contingency hypothesis. Specifically, it is expected that contingency and Stroop effects will act separately and that there should therefore be no interaction between Stroop trial type and contingency. That is, the size of the Stroop effect should be the same for words with .75 contingencies (i.e., *high-proportion-congruent* congruent words vs. *low-proportion-congruent* incongruent words) as it is for words with .25 contingencies (i.e., *low-proportion-congruent* congruent words vs. *high-proportion-congruent* incongruent words) or any other contingency. Congruent and incongruent trials should be faster in the high contingency condition, but the *difference* between congruent and incongruent trials should not vary with contingency. In a final analysis, high contingency and low contingency trials are compared to medium (chance) contingency trials to assess, respectively, facilitation and interference. Based on the response prediction model, the expectation is that the contingency effect will be solely facilitative for response latencies. In other words, although there is a benefit in responding to items where the words have greater than chance contingencies, there is no cost in responding to items where the words have lower than chance contingencies.

Reanalysis 1–Jacoby et al. (2003) Response Latencies

Method

Sixteen participants named the colour (green, white, blue, yellow) of colour words (GREEN, WHITE, BLUE, YELLOW). Individual words were presented 75%, 50%, or 25% of the time in their congruent colour, and in a selected incongruent colour the remaining trials. Further details on the method of this experiment can be found in the original paper by Jacoby et al. (2003). For the purpose of these analyses, the means for congruent and incongruent trials in Experiment 2a of Jacoby et al. were rearranged as a function of word-response contingency: high (.75), medium (.50), and low (.25). Thus, the high contingency condition consists of congruent trials with high proportion congruent words and incongruent trials with low proportion congruent words, the medium contingency condition consists of congruent and incongruent trials with medium proportion congruent words, and the low contingency condition consists of congruent trials with low proportion congruent words and incongruent trials with high proportion congruent words.

Results

The response latency data (along with the original organization of the data) are presented in Figure 2. A 3 (contingency; high, medium, low) x 2 (Stroop trial type; congruent, incongruent) ANOVA for response latencies revealed a main effect for contingency, $F(2,30) = 3.848$, $MSE = 1513$, $p = .033$, $\eta_p^2 = .204$, and a main effect for Stroop trial type, $F(1,15) = 35.235$, $MSE = 4902$, $p < .001$, $\eta_p^2 = .701$. Critically, and as predicted, no interaction was observed between contingency and Stroop trial type, $F(2,30) = .284$, $MSE = 668$, $p = .755$, $\eta_p^2 = .019$. Although Type II error is always a concern when interpreting a null finding, the measure of effect size (η_p^2) indicates that the interaction term explains less than 2% of the variance in the results,[1] thus reinforcing the claim that this is a true null (or at least very small) interaction.

Congruent and incongruent trials for each contingency level were then averaged and planned comparisons were conducted to test the hypothesis that the entire contingency effect is driven by response facilitation on high contingency trials. As expected, there was a significant 24 ms advantage for high contingency trials (609 ms) relative to medium (chance) contingency trials (636), $t(15) = 2.293$, $SE_{diff} = 11$, $p = .036$, $\eta_p^2 = .260$. The 1 ms numerical difference between medium contingency (636) and low contingency trials (635) was not significant, $t(15) = .180$, $SE_{diff} = 10$, $p = .859$, $\eta_p^2 = .002$.

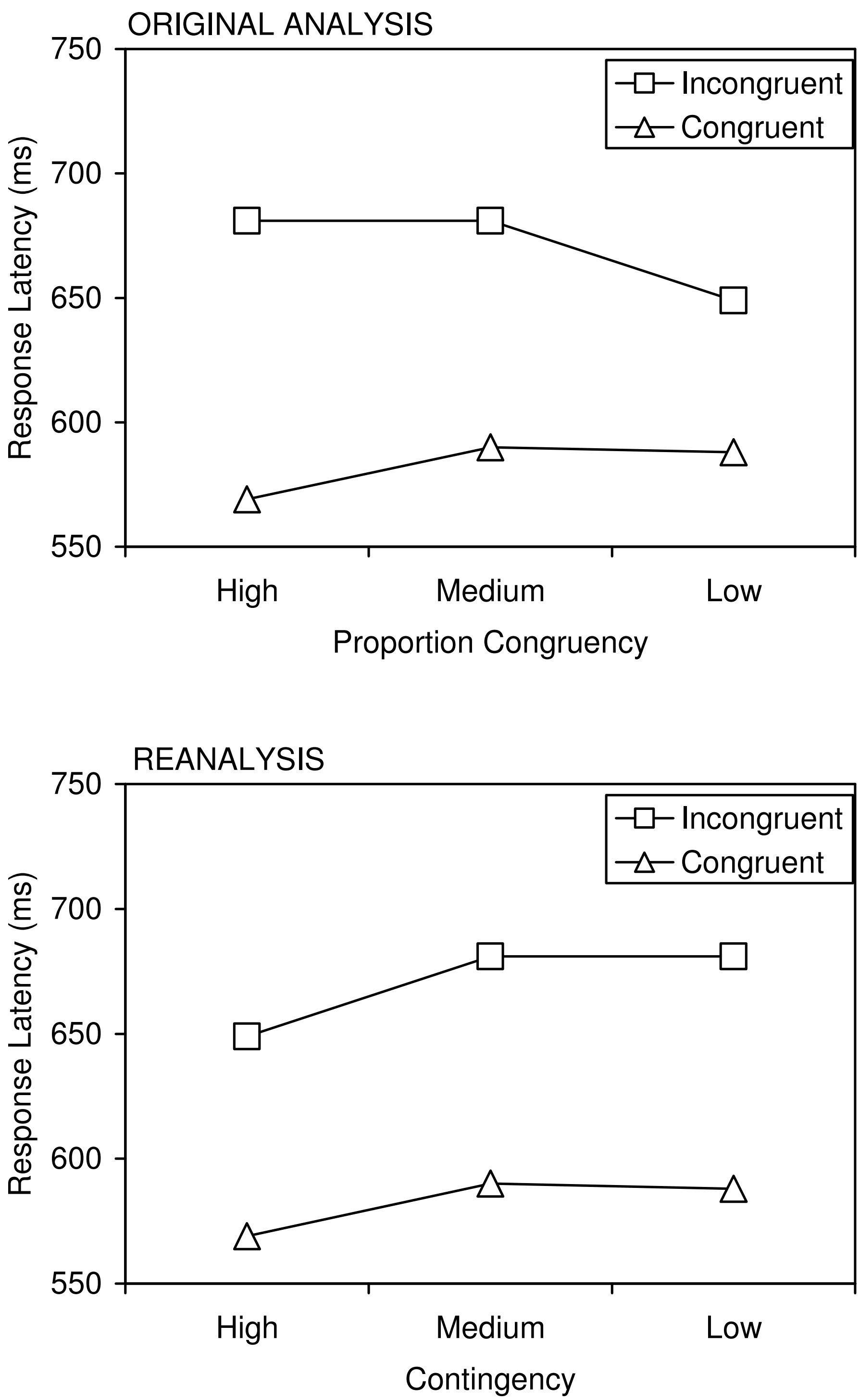

Figure 2. Bottom: mean response latencies in milliseconds from Jacoby, Lindsay, and Hessels (2003) for congruent and incongruent trials with high, medium, and low contingencies. Top: original organization of the data.

Discussion

The results of this reanalysis confirm the two hypotheses stemming from the contingency hypothesis. First, contingency and Stroop effects act separately, in that these factors produced no interaction. Second, the contingency effect is entirely facilitative. As discussed in the introduction, these findings provide further empirical support for the contingency hypothesis, which holds that participants use distracter words to predict responses thereby shortcutting some processing when this prediction is successful (high contingency trials) relative to when response prediction is unsuccessful (low contingency trials) or not attempted (medium contingency trials).

These findings undermine the modulation hypothesis, which asserts that participants modulate attention to the word based on the proportion congruency of the distracter words. Of particular importance, if it is assumed that participants increase attention to the word on high proportion congruent trials, then not only should congruent trials (high contingency) be speeded relative to the chance condition due to increased facilitation (observed), but incongruent trials (low contingency) should be slowed relative to the chance condition due to increased interference (not observed). Similarly, if participants suppress the word on low proportion congruent trials, then not only should incongruent trials (high contingency) be speeded relative to the chance condition due to decreased interference (observed), but congruent trials (low contingency) should be slowed relative to the chance condition due to decreased facilitation (not observed). Further, because interference is substantially larger than facilitation in the Stroop task, incongruent trials should have been more influenced by attention than congruent trials, resulting in an interaction between contingency and Stroop trial type (not observed). Thus, these data entirely support the contingency hypothesis, but are in important ways inconsistent with the modulation hypothesis.

Consistent with the present findings, Logan, Zbrodoff, and Williamson (1984, Experiment 3) reported a null interaction between contingency and Stroop trial type for a four-choice Stroop task with a standard task-wide proportion congruent manipulation. In that experiment, the difference in response latencies between high and low contingency congruent trials was the same as the difference between high and low contingency incongruent trials. Logan et al. also reported a pair of two-choice experiments (Experiments 1 and 2) where an interaction between contingency and Stroop trial type *was* observed. This could be regarded as a potential problem for the contingency hypothesis described here. However, it is important to note that task-wide proportion congruency manipulations (two-choice especially) are subject to confounding sequential effects (i.e., trial-to-trial modulations of word reading). For instance, if the word on the previous trial was congruent, then the Stroop effect will be larger on the current trial (probably due to increased attention to the word; Gratton, Coles, & Donchin, 1992; see Schmidt et al., 2007, for a similar finding of sequential contingency effects). Given that more trials are preceded by a congruent trial in a high proportion congruent block than in a low proportion congruent block, it seems that simple proportion congruency experiments are ill suited to test the present hypotheses.[2] The ISPC paradigm used here, on the other hand, is able to rule out the sequential trial confounds, because each trial type is preceded by the same proportion of every type of trial (i.e., there are no systematic sequential confounds).

Reanalysis 2–Jacoby et al. (2003) Percentage Error

The first reanalysis demonstrated that when the proportion of congruent items is manipulated the resultant variations in the size of the Stroop effect can be explained by confounded contingencies. In the current section, I expand on the contingency hypothesis used to generate the predictions for the first analysis and test some novel predictions that fall out of this account. In particular, whereas the response threshold mechanism described earlier predicts only facilitation in response latencies, it predicts both facilitation *and* interference in errors. The reason that facilitation is predicted on high contingency trials for errors is the same as that for response times: the threshold for the correct (predicted) response is lower. Consequently, it takes less time to make a correct response and it is highly unlikely that participants will make another (incorrect) response due to the relatively lower threshold for the correct response compared to the incorrect responses.

Unlike for latencies, where no response interference was expected, response interference *is* predicted for errors because the threshold for one of the competing (incorrect) responses is lowered relative to the correct response and to the other incorrect responses. For instance, for the stimulus $BLUE_{green}$, the threshold for the blue response is lowered. However, the blue response is not the correct response. Thus, there is a heightened probability that participants will make an error and select this predicted incorrect response (blue) over the unpredicted correct response (green).

Method

There were 16 participants in Experiment 2b. Experiment 2b was identical in all respects to Experiment 2a, except that participants had only 550 ms (rather than 2000 ms in Experiment 1) to respond. This procedure increases the number of errors, which increases the power to detect effects in the error data. For the purpose of these analyses, the means for congruent and incongruent trials in Experiment 2b of Jacoby et al. were rearranged as a function of word-response contingency: high (.75), medium (.50), and low (.25).

Results

The percentage error data (along with the original organization of the data) are presented in Figure 3. A 3 (contingency; high, medium, low) x 2 (Stroop trial type; congruent, incongruent) ANOVA for percentage error revealed a main effect for contingency, $F(2,30) = 8.827$, $MSE = 0.8$, $p < .001$, $\eta_p^2 = .233$, and a main effect for Stroop trial type, $F(1,15) = 114.481$, $MSE = 2.9$, $p < .001$, $\eta_p^2 = .798$. As predicted, no interaction was observed between contingency and Stroop trial type, $F(2,30) = .126$, $MSE = 0.8$, $p = .882$, $\eta_p^2 = .004$. Again, the effect size measure (η_p^2) suggests that this is a true null (or at least very small) interaction, explaining less than 1% of the variance in the results.

Congruent and incongruent trials for each contingency level were then averaged and planned comparisons were conducted to test the hypothesis that the contingency effect has both a facilitative effect on high contingency trials and an interfering effect on low contingency trials. As expected, there was a significant 4.4% advantage for high contingency trials (32.3%) relative to medium (chance) contingency trials (36.6%), $t(15) = 1.854$, $SE_{diff} = 2.4$, $p = .042$ one tailed, $\eta_p^2 = .106$, and a significant 5.2% disadvantage for low contingency trials (41.9%) relative to medium contingency trials, $t(15) = 2.315$, $SE_{diff} = 2.3$, $p = .035$, $\eta_p^2 = .156$.

Discussion

The results of the second reanalysis provide further support for the response threshold mechanism of the contingency hypothesis. Specifically, whereas only facilitation was observed in response latencies, both facilitation on high contingency trials and interference on low contingency trials were observed in the error data. It is atypical for an account to make differential predictions for response latencies and errors, so the present findings constitute strong and thoroughly consistent support for the contingency hypothesis. The experiments that follow test some even more fine-grained hypotheses generated from the contingency hypothesis.

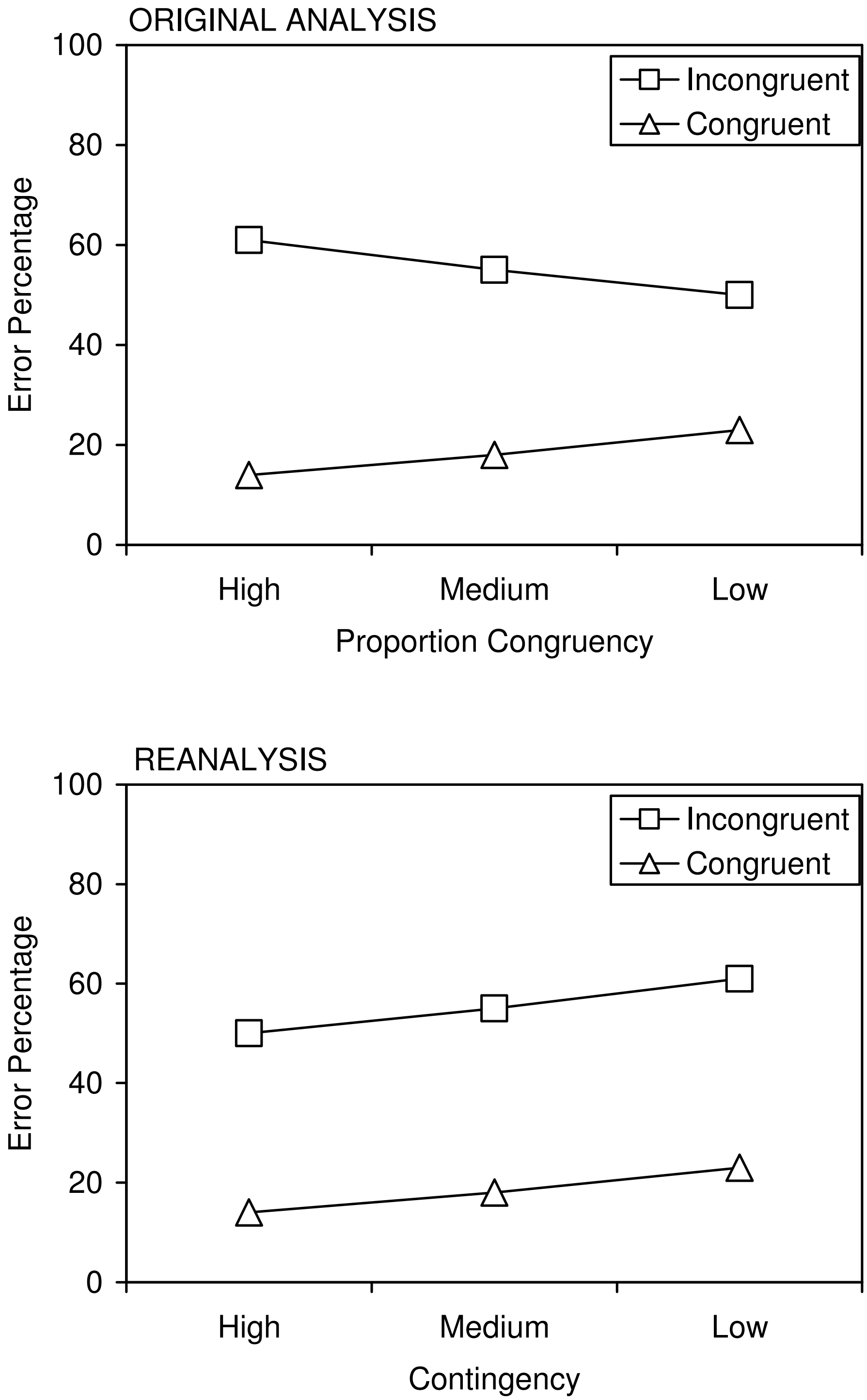

Figure 3. Bottom: error percentages from Jacoby, Lindsay, and Hessels (2003) for congruent and incongruent trials with high, medium, and low contingencies. Bottom: original organization of the data.

Experiment 1

The contingency hypothesis presented here holds that participants implicitly learn contingencies and implicitly use them to reduce the response threshold for the predicted response. This account successfully predicted facilitation with no corresponding interference when response latency was the dependent variable, but both facilitation and interference when error percentage was the dependent variable. Further, contingency effects operate separately from the Stroop effect, a finding supported by the null interactions between contingency and Stroop trial type. To further test the contingency hypothesis, Experiments 1 and 2 investigate contingency effects in the contingency learning paradigm developed by Schmidt et al. (2007).

Schmidt et al. (2007; see also Musen & Squire, 1993) report several experiments that provide further support for the contingency hypothesis interpretation of the ISPC effect by showing that contingency effects can be observed in the absence of a Stroop effect. In their experiments, colour-unrelated words (e.g., MOVE) were presented most often in a given colour (e.g., MOVE 75% of the time in blue, SENT 75% of the time in green, etc.). Key press responses were faster to high contingency trials (e.g., $\text{MOVE}_{\text{blue}}$) than low contingency trials (e.g., $\text{MOVE}_{\text{green}}$). Thus, the results of these experiments confirm that participants are able to use words (even when they are colour-unrelated) to predict what response to make. The modulation hypothesis cannot accommodate these findings, because the words and colours have no congruency relation that can be used to decide whether or not to attend to the word.

Experiment 1 was a two-choice, rather than a four-choice, task (in part to ensure that the interactions observed by Logan et al., 1984, were not attributable to something more than sequential confounds). In Experiment 1, colour-unrelated words were presented most often in a randomly-assigned colour. For instance, SEVEN may have been presented 75% percent of the time in blue and 25% of the time in green, whereas CHAIR may have been presented in blue and green equally often. This type of manipulation creates high contingency (.75; e.g., $SEVEN_{blue}$), medium contingency (.50; $CHAIR_{blue}$), and low contingency (.25; e.g., $SEVEN_{green}$) trials. Experiment 1 focuses primarily on response latencies, with the goal of replicating the facilitation-only pattern of results observed in Reanalysis 1.

Experiment 2 uses the response deadline procedure used in Jacoby et al. (2003), where participants are given a very short amount of time to respond, to maximize errors and focuses on error percentages. Experiment 2 attempts to replicate the findings observed in Reanalysis 2 and tests new predictions from the response threshold model. Both experiments are key press, rather than naming, tasks.

Method

Participants. Thirty-four University of Waterloo undergraduates participated in Experiment 1 in exchange for course credit.

Apparatus. All stimuli were presented on a standard computer monitor and responses were made on a QWERTY keyboard. Participants pressed the 'f' key for blue and the 'j' key for green. Stimulus presentation and response timing were controlled by E-Prime (Psychology Software Tools, 2002).

Materials and Design. There were three display words (SEVEN, GLIDE, CHAIR) and two display colours (blue, green) in the experiment. The words were selected to be of the same length. For each participant, one of the words (e.g., SEVEN) was presented most often (three out of four times per block) in blue, another word (e.g., GLIDE) was presented most often (three out of four times per block) in green, and the final word (e.g., CHAIR) was presented equally often (twice each per block) in blue and green. Assignment of words to colours was counterbalanced across participants. Words were presented in bold 18 point Courier New font. The RGB values for the stimulus colours were 0,0,255 (blue) and 0,255,0 (green). There were three contingency levels in the experiment: high (.75; e.g., $\text{SEVEN}_{\text{blue}}$), medium (.50; e.g., $\text{CHAIR}_{\text{blue}}$), and low (.25; $\text{GLIDE}_{\text{blue}}$).

Procedure. There were 420 trials in this experiment, consisting of 35 blocks of 12 trials each. On each trial, a white (255,255,255) fixation cross (i.e., "+") was presented in the middle of a black screen for 250 ms. This was followed by another 250 ms of blank screen, followed by the stimulus display. The stimulus display was presented until a response was made or until the trial timed out at 2000 ms. Correct responses were followed by a blank screen for 250 ms before the next fixation cross. Incorrect and null responses were followed by the messages "Incorrect" and "Too Slow," respectively, for 1000 ms in red (255,0,0).

Results

The dependent measures for Experiment 1 were mean correct response latencies and error percentages. All responses shorter than 250 ms or longer than the response deadline were considered spoiled trials and were excluded from analysis (less than 0.2% of the trials).

The mean correct response latencies for Experiment 1 are presented in Figure 4. A one-way ANOVA on contingency (high, medium, low) for response latencies was significant, $F(2,66) = 3.902$, $MSE = 162$, $p = .025$, $\eta_p^2 = .106$. As expected, planned comparisons revealed that high contingency trials (437 ms) were responded to faster than medium contingency trials (444 ms), $t(33) = 3.223$, $SE_{diff} = 2.1$, $p = .003$, $\eta_p^2 = .239$. Also as predicted, no difference was found between medium and low contingency (445 ms) trials, $t(33) = .395$, $SE_{diff} = 3.4$, $p = .696$, $\eta_p^2 = .005$. For this latter comparison, there was high power (.8) to detect an effect as small as 6 ms.

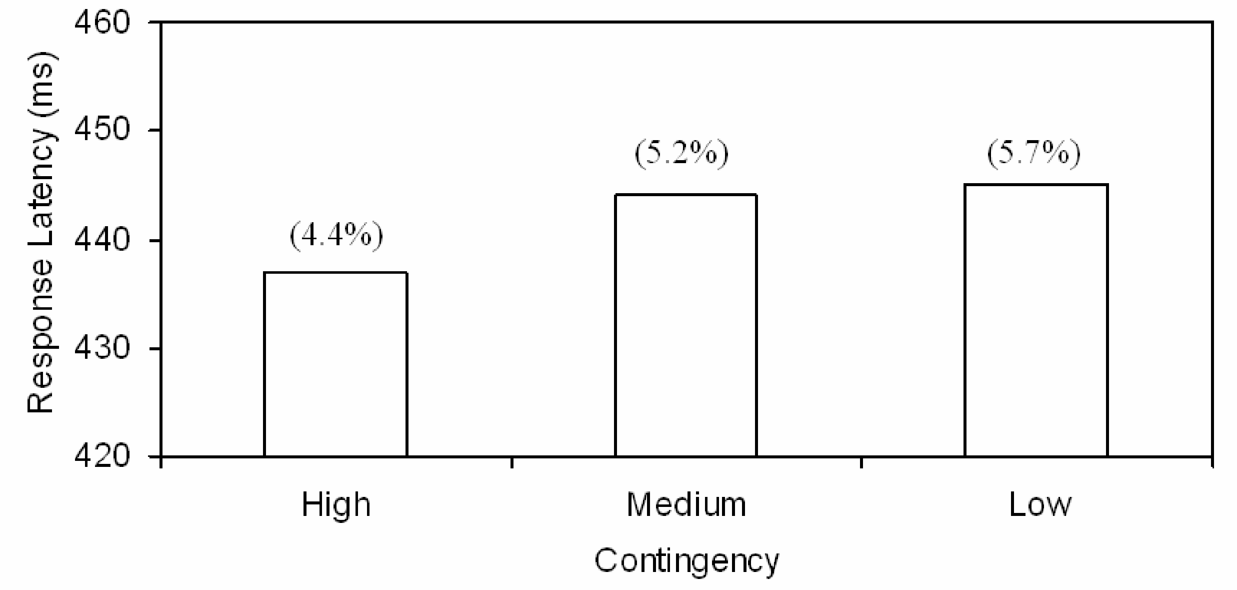

Figure 4. Experiment 1 mean response latencies in milliseconds for high, medium, and low contingency trials. Error percentages are in parentheses.

Errors were infrequent in Experiment 1, as shown in Figure 4. An ANOVE on contingency (high, medium, low) for error percentages was significant, $F(2,66) = 3.987$, $MSE < .1$, $p = .023$, $\eta_p^2 = .108$. Although the means were numerically in the expected direction, there was only a marginal difference between high (4.4%) and medium contingency trials (5.2%), $t(33) = 1.763$, $SE_{diff} = .5$, $p = .087$, $\eta_p^2 = .086$, and a non-significant difference between medium and low contingency trials (5.7%), $t(33) = .941$, $SE_{diff} = .5$, $p = .353$, $\eta_p^2 = .026$.

Discussion

Experiment 1 successfully replicated the finding of response facilitation for high contingency trials with no response interference for low contingency trials relative to medium (chance) contingency trials in a task with colour-unrelated words. This provides further support for the claim that the "proportion congruent" effect is in fact not due to modulations in the Stroop effect, but is instead due to a confounding of Stroop and contingency effects.

Experiment 2

Experiment 2 investigated whether the same pattern of facilitation and interference in errors observed in the ISPC task would be seen with a simple contingency task. In addition, Experiment 2 tested a novel prediction from the contingency hypothesis. The argument for the presence of response interference on low contingency trials is that, because the threshold for the (incorrect) predicted response is lowered, participants will be more likely to erroneously select this predicted response rather than the correct response for the colour. For instance, for the stimulus $\text{MEET}_{\text{green}}$, where MEET is presented most often in blue, participants will predict a blue key response based on the word. This will result in the threshold for the blue response being lowered and will therefore increase participants' tendency to incorrectly make this response (i.e., because lowering a response threshold can be likened to loosening the trigger for this response to fire). An observation of response interference in errors supports the notion of a response threshold mechanism, but this account further predicts that the inflation of errors in the low contingency condition is due solely to an increase in the *specific predicted response* rather than just an increase in errors in general. For instance, in a four-choice task, we should expect increased blue key errors for the stimulus $\text{MEET}_{\text{green}}$, but *no* increased yellow or orange key errors. Although yellow and orange would also be incorrect responses, they are not the *predicted* incorrect response. Thus, it should be expected that more than a chance number of errors in the low contingency condition (33% of incorrect responses) should be the predicted response. Moreover, after correcting for this inflation of errors for the predicted response, response interference should no longer be observed.

Method

Participants. Ninety-five University of Waterloo undergraduates participated in

Experiment 2 in exchange for course credit. None had participated in Experiment 1.

Apparatus. The apparatus for Experiment 2 was identical in all respects to Experiment 1, with one exception. Experiment 2 was a four-choice task (which was necessary to test the hypotheses regarding predicted versus unpredicted errors); participants pressed the "a" key for blue, the "z" key for green, the "m" key for yellow, and the "k" key for orange.

Materials and Design. The materials and design for Experiment 2 were identical to Experiment 1, with the following exceptions. There were six new display words (LOOP, FINS, MEET, SLID, CALL, TUBE) and four display colours (blue, green, yellow, orange). For each participant, one of the words was presented most often (six out of twelve times per block) in blue, another most often in green, another most often in yellow, and another most often in orange. The remaining two words were presented equally often (three times each per block) in all colours. Assignment of words to colours was counterbalanced across participants. The RGB values for the new stimulus colours were 255,255,0 (yellow) and 255,125,0 (orange). There were three contingency levels in the experiment: high (.50), medium (.25), and low (.17).

Procedure. The procedure was identical in all respects to Experiment 1, with the following exceptions. There were 432 trials in this experiment, consisting of 6 blocks of 72 trials each. To increase error frequency, participants had to respond before a 550 ms deadline when the stimulus display terminated. To avoid discouraging errors, only the "Too Slow" message for null responses had the 1000 ms "penalty duration" used in Experiment 1, whereas the "Incorrect" message for errors was set to the 250 ms duration used on correct trials.

Results

The dependent measure for Experiment 2 was error percentages. Responses not made within the response deadline were considered spoiled trials and were excluded from analysis (approximately 20% of the data, reflecting the difficulty of this task).

The error percentages for Experiment 2 are presented in Figure 5. To broadly characterize the data, although the manipulation was able to induce adequately high errors, contingency effects were small in this experiment. This may have been due to the relatively weaker contingency manipulation (.17-.50 rather than .25-.75) or due to difficulty learning contingencies with such a brief response window. A one-way ANOVA on contingency (high, medium, low) for errors was significant, $F(2,188) = 12.068$, $MSE = 0.3$, $p < .001$, $\eta_p^2 = .114$. As expected, planned comparisons revealed that high contingency trials (27.7%) generated fewer errors than medium contingency trials (30.1%), $t(94) = 2.890$, $SE_{diff} = .8$, $p = .005$, $\eta_p^2 = .082$, and low contingency trials (31.4%) generated more errors than medium contingency trials, $t(94) = 1.929$, $SE_{diff} = .7$, $p = .028$ one tailed, $\eta_p^2 = .038$.

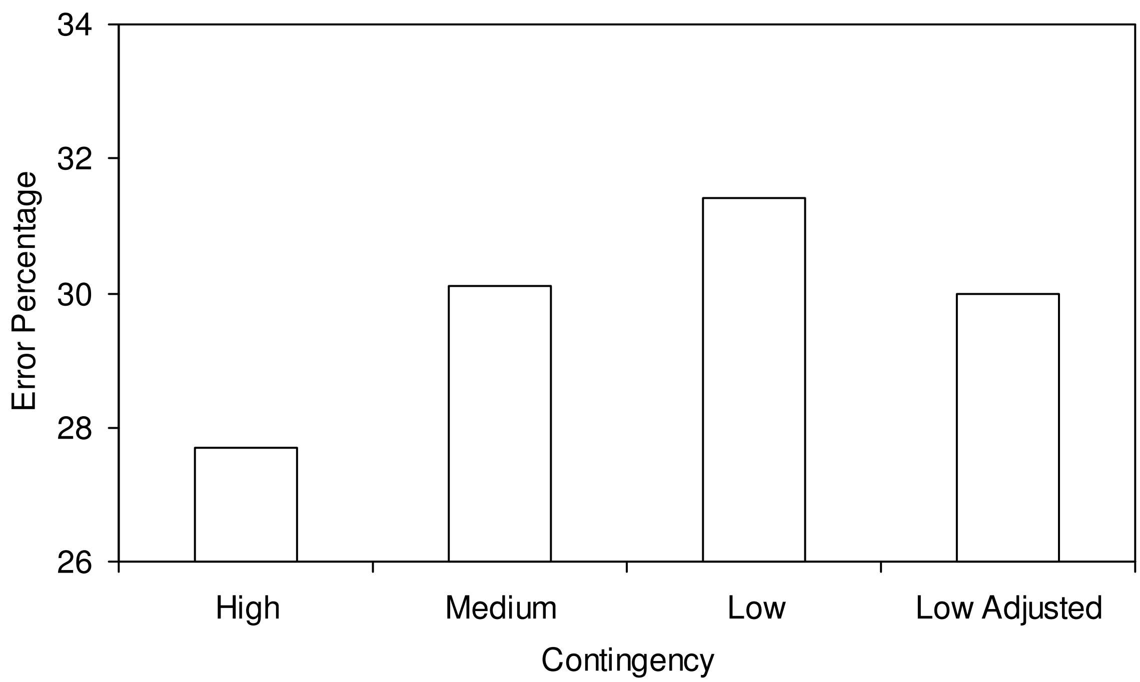

Figure 5. Experiment 2 error percentages for high, medium, and low contingency trials.

Critically, I also assessed the hypothesis that this increase in errors for low contingency trials was due to an increase in the (incorrect) predicted response and found that the 36.4% errors for the predicted response was significantly greater than chance (33.3%), $t(94) = 2.800$, $SE = 1.1$, $p = .006$, $\eta_p^2 = .077$. I then tested the assumption that the interference observed for low contingency trials would be eliminated by removing the extra predicted errors. To do this, I calculated a fourth condition (shown in Figure 5), low contingency adjusted, that consists of the error frequency of only the non-predicted responses adjusted to the same scale as the other trial types. This was done with the formula: (low contingency) x (percentage unpredicted errors) x (3 / 2). As predicted, the resulting low contingency adjusted condition showed significantly fewer errors (30.0%) than the unadjusted low contingency condition (31.4%), $t(94) = 3.036$, $SE_{diff} = .4$, $p = .003$, $\eta_p^2 = .089$, and no difference in errors compared to medium contingency trials (30.1%), $t(94) = .113$, $SE_{diff} = .7$, $p = .910$, $\eta_p^2 < .001$. For this last comparison, there was high power (.8) to detect an effect as small as 2%.

Discussion

Experiment 2 replicated the findings of facilitation and interference in errors observed in Reanalysis 2. Additionally, these analyses support the prediction of the contingency hypothesis that the interference effect in errors is due solely to the increase in errors for the predicted response: participants made more of the predicted errors than expected by chance, and when this increase was controlled for there was no difference in errors for medium and low contingency trials. This is inconsistent with any version of the modulation hypothesis that I can think of and provides further support for the notion that participants use the contingencies between words and responses to reduce the response threshold for the predicted response.

General Discussion

Two new analyses of the data from Jacoby et al.'s (2003) along with the results of two new experiments converge with the results of other recent investigations (e.g., Schmidt et al., 2007) to confirm a simple contingency account of the ISPC effect (and probably the majority of the proportion congruent effect). When Stroop effects are analyzed as a function of the predictability of the words (i.e., contingency), Stroop trial type (congruent, incongruent) does not interact with contingency in either the response latency or error data. This finding confirms the view that the entire ISPC effect (i.e., the highly replicated interaction between proportion congruency and Stroop trial type) can be accounted for by a main effect of Stroop trial type and a main effect of contingency.

Reanalysis 1 and Experiment 1 also demonstrated that contingency effects are solely facilitative in response times. That is, high contingency trials are speeded relative to medium (chance) contingency trials, but there is no disadvantage for low contingency trials relative to medium contingency. These results are inconsistent with the modulation hypothesis, but are entirely as predicted by the contingency hypothesis. According to the contingency hypothesis, response time facilitation occurs because participants are able to successfully predict the correct response based on the word on these trials and shortcut processing by reducing the response threshold for the predicted response. This same mechanism further predicts facilitation for high contingency trials and interference for low contingency trials in errors. These predicted results for errors were observed in Reanalysis 2 and Experiment 2. Error facilitation is predicted according to this account because the threshold for the correct response is lowered when it is predictable (i.e., on high contingency trials) and it is therefore less likely that participants will make any other response. Error interference is also predicted, because the response threshold for one of the *incorrect* responses is lowered when it is predictable (i.e., on low contingency trials) and it is therefore more likely that participants will make this particular error. Further analyses in Experiment 2 confirmed that the increase in errors in the low contingency condition is due solely to the increase in the *specific* predicted incorrect response.

Conflict Monitoring

These findings have important implications for the conflict monitoring framework of Botvinick, Braver, Barch, Carter, and Cohen (2001). Botvinick et al. propose that many of the effects observed in Stroop and Stroop-like paradigms can be explained by a conflict monitoring mechanism thought to lie in the anterior cingulate cortex (ACC). According to their account, the ACC detects conflict when it occurs (e.g., when a word and display colour indicate two different potential responses). After detecting conflict, attention is more stringently focused on the target. Through simulation work, Botvinick et al. have demonstrated that conflict monitoring can explain sequential Stroop effects (because attention to the target is increased following an incongruent trial) and the standard proportion congruent effect (because the greater number of incongruent trials in the low proportion congruent condition leads to more attention to the target dimension of colour block-wide).

Clearly, the conflict monitoring account is a variant of the modulation hypothesis and suffers from the same shortcomings. In particular, the modulation hypothesis is unable to explain the ISPC effect, because the amount of conflict experienced prior to high and low proportion congruent trials is the same (i.e., because they are mixed within the same block). However, Blais, Robidoux, Risko, and Besner (in press) present a variant of the conflict monitoring account in which conflict is monitored separately for each item. For instance, each time YELLOW is presented as the distracter word, the level of conflict will be monitored. If YELLOW is presented most often in an incongruent colour, then overall conflict will be high. As such, when YELLOW is again encountered, attention to the colour will be highly focused, thus minimizing the impact of the word on performance. However, within the same block, BLUE may be presented most often in its *congruent* colour, resulting in low overall conflict for this word. As such, when the word BLUE is presented, attention to the colour will not be as focused, thus allowing the word to impact performance relatively more. Blais et al.'s variant of the conflict monitoring account can therefore explain the variations in the size of the Stroop effect when the proportion of congruent items is manipulated across items. But Blais et al.'s account simulates only the interaction between proportion congruency and Stroop trial type, and not the various other characteristics of the data (e.g., the fact that the effect is driven by the speeding of congruent trials in the high proportion congruent condition and incongruent trials in the low proportion congruent condition). More critically, their account cannot explain why contingency effects are still observed with colour-unrelated words: there is simply no differential conflict to monitor between the various conditions.

In summary, although the conflict monitoring model of Botvinick et al. (2001) provides a compelling explanation for sequential Stroop effects (which undoubtedly contribute to the simple proportion congruent effect), the weight of the evidence suggests that the ISPC effect is better explained by contingency learning. Further, although sequential (trial n-1) effects contribute to the simple proportion congruent effect, it is my suggestion that the remainder (and bulk) of the effect is explained by contingency.

Implications for Future Stroop Research

These results, successfully predicted by contingency hypothesis, illuminate an error in the popular conceptualization of proportion congruency manipulations. For instance, in the high proportion congruent condition of a two choice task, *congruent* words with *high contingencies* (e.g., $BLUE_{blue}$) are being compared to *incongruent* words with *low contingencies* (e.g., $BLUE_{green}$). That is, words are 75% predictive of the correct response in the congruent condition, but only 25% predictive of the correct response in the incongruent condition. The reverse is true in the low proportion congruent condition, where *congruent* words with *low contingencies* (e.g., $ORANGE_{orange}$) are being compared to *incongruent* words with *high contingencies* (e.g., $ORANGE_{yellow}$). In this sense, Stroop effects (i.e., facilitation and interference resulting from conflict in meaning between related concepts) are confounded with contingency effects (i.e., prediction of responses based on word-response correlations).

Given that the contingency hypothesis explanation of the proportion congruent effect appears sufficient, it would be prudent to question the "purity" of all Stroop effects emerging from methodologies with non-chance contingencies. Algom and colleagues have reached a similar conclusion (Dishon-Berkovits & Algom, 2000; Malara & Algom, 2003; Sabri, Melara, & Algom, 2001). The main methodological point is that, although non-chance contingencies may "maximize" the size of the Stroop effect (and thus its detectability), the said Stroop effect measure is confounded with contingency effects. Thus, any data collected with such a design may not be applicable to the standard effect with chance contingencies. For instance, in a four-colour Stroop task, each word should be presented in each colour 25% of the time (which is typically not the case). Any deviation from such contingencies will necessarily confound Stroop effects with contingency learning effects. This methodological point applies to all paradigms in which contingency information may aid responding.

Conclusions

The results of two new analyses of previously published data (Jacoby et al., 2003) along with the results of two new experiments disconfirm the generally accepted and intuitively appealing modulation hypothesis, which explains the proportion congruent effect in terms of control over attention allocated to processing the word. The modulation hypothesis consistently failed in its predictions. On the other hand, these analyses provide consistent support for the contingency hypothesis, which explains the proportion congruent effect in terms of the predictability of words (i.e., contingency). Specifically, the results are consistent with the proposed response mechanism where participants decrease the threshold for the expected (high contingency) response. Of particular note, the contingency hypothesis successfully predicted differential results for response latencies and error rates.

Confirmation of the contingency hypothesis will likely be received with both frustration and excitement by researchers in cognitive psychology. The results may seem frustrating because they disconfirm the highly appealing modulation hypothesis and consequently bring into question the proper interpretation of many of the experiments that have made use of the paradigm over the last 25 years. On the other hand, the results should be exciting because they demonstrate the processing power and control of implicit processes and offer new insights for the growing field of implicit learning.

Footnotes

[1] Note that even a true random factor will explain some proportion of variance with a partial eta squared (η_p^2) due to maximization on random error. This is true of all measures of effect size, such as R squared (R^2).

[2] The issue of sequential trial confounds can become even more complicated given that all congruent trials have high contingencies and all incongruent trials have low contingencies in the high proportion congruent block, whereas the reverse is true in the low proportion congruent block.

References

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

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.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40*, 343-367.

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.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480-506.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review, 10,* 634-644.

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,* 219-234.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color-word Stroop task. *Bulletin of the Psychonomic Society, 22,* 135-138.

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

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

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110,* 422-471.

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

Psychology Software Tools. (2002). E-Prime. http://www.pst-net.com

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.

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

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