The Effect of Response Set Size on the Stroop Interference

Adam Chuderski (adam.chuderski@gmail.com)
Institute of Philosophy, Jagiellonian University
Grodzka 52, 31-044 Krakow, Poland

Tomasz Smolen (tomasz.smolen@mailplus.pl)
Institute of Psychology, Jagiellonian University
Mickiewicza 3, 31-120 Krakow, Poland

Abstract

Previous research on the effect of response set size on the interference resulting from processing in the Stroop task – a paradigmatic test of executive control – brought equivocal conclusions. In the present paper, we analyze what predictions regarding this effect should be drawn from the most influential computational models of the Stroop as well as from our own new model. Then we test these predictions in an experiment, by manipulating response set size as well as the stimuli/response set sizes proportion, finding both evidence in favor of our model and data which is not explained by any Stroop model.

Introduction

An intensively studied human mental faculty is executive control, being the ability of the human mind to influence and organize its own cognitive processing, including control over perceptual stimulation and motor programs. Work on various executive control functions, such as preponent response inhibition, task switching, and multitasking, has inspired models of integrated control of human cognitive architecture (Gray, 2007).

The Stroop task (MacLeod, 1991) is probably the most popular test measuring an interesting aspect of executive control, namely interference resolution, which in general requires focusing on a new, weakly-learned process when dealing with a stimulus, while overriding another process, which is well-learned, strongly associated with that stimulus and automatically activated by it. The Stroop task has been used as an important tool for the verification of several computational models of executive control (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Cohen, Dunbar, & McClelland, 1990; Herd, Banich, & O'Reilly, 2006; Roelofs, 2003; Smoleń & Chuderski, 2010; van Maanen, van Rijn, & Borst, 2009; Verguts & Notebaert, 2008)

In its standard version, the Stroop test requires the naming of the ink color of a word, which itself refers to another color. The so-called interference effect is being observed, which consists of increased response latency in such incongruent trials, when compared to RT in neutral trials (i.e., when it is required to name the color of a color-unrelated string, like XXXXX). If a color and a word match, the facilitation effect is also being observed, which consists of a decreased response latency in such congruent trials in comparison to neutral trials. Other interesting experimental effects have been also found (cf. MacLeod, 1991).

One of the most intensively examined effects related to Stroop interference is the *response set size effect* – in some studies, increasing the number of color-response pairs increased the observed interference effect (this effect should not be mistaken for the response-set effect: a word meaning a color, which is not associated with any response, usually yields little interference. This latter effect is not the focus of this paper). The response set size effect has been used in the testing of some Stroop models (e.g., Cohen et al., 1990; Kanne, Balota, Spieler, & Faust, 1998).

However, numerous studies (for a review see MacLeod, 1991) did lead to ambigous findings regarding whether increasing the number of possible responses in the Stroop task really increases the interference or whether it decreases it or does not have any effect. This mutually contradictory data might have resulted from differences in experimental designs, as both the standard Stroop (so-called the color-word) task and its analogs (e.g., the picture-word or word-word tasks) were used in various settings. Moreover, most of the studies confounded response set size with stimulus set size, as most commonly one-to-one SR mappings were applied (especially, in an oral version).

In his paper, MacLeod (1991) cited several studies (published between the sixties and the eighties) which reported no effect of stimulus/response set size on the amount of interference. He also cited three studies which showed an increase in interference resulting from an increase in set size (e.g., Williams, 1977), and three other studies, which demonstrated the opposite effect (e.g., La Heij & Vermeij, 1987). In two more recent studies (Kanne et al., 1998; La Heij & van den Hof, 1995) the increasing interference was found as a result of an increase in the number of SR mappings. However, in the former study, set size was manipulated only in range from two to four stimuli/responses. The two-stimuli condition is an unusual one, because if the exact stimulus is not repeated, the presentation of two consecutive incongruent trials always involves the priming of the second target (i.e., if in a preceding incongruent trial a distractor word is presented, then it *has* to be a target in a following trial). In three- or four-stimuli versions, consecutive incongruent trials need not involve priming.

The difference in the amount of priming may cause differences in interference effects (Chuderski & Smoleń, unpublished data). In fact, the increase of interference in Kanne et al.'s study was only significant between two- and three-stimuli conditions (Δ = 61 ms) but not between three-and four-stimuli conditions (Δ = 9 ms). On the contrary, in La Heij & van den Hof's study, the difference in set sizes was so huge (4 vs. 16) that the set size factor might have been substantially confounded with other variables (e.g., working memory load). In general, the pattern of the cited results seems to be mixed and unconvincing.

Moreover, all cited experiments were applied with the use of oral versions of the Stroop task (or its analogs), which require naming the color. It would be interesting to see what results would be observed if manual responses had been required, as such responses are in fact the most common procedure in psychological experiments.

In the present paper we tested the effect of increasing the number of manual reactions required in the Stroop task on the amount of Stroop interference. Moreover, we attempted to investigate the influence of the ratio of stimuli to responses, namely what will happen if not only one but two or three stimuli are associated with one response. The main goal of the study was theoretical: as we believe that the effects of set size on Stroop performance are important indicators of processes responsible for coping with the interference, we wanted to test some existing models of Stroop against data from our experiment on set size effects.

Predictions of Stroop models regarding stimulus/response set size

We analyzed predictions of three (groups of) Stroop models: (a) Cohen et al.'s (1990) connectionist model and its extensions (Cohen, Usher, & McClelland, 1998), which explain the Stroop interference in terms of differences in practice (strength) between color and word naming as well as of the attentional modulation of color/word processing, (b) Roelofs' (2003) theory and similar models (Altmann & Davidson, 2001; van Maanen et al., 2009), which identify the interference as resulting from access to declarative memory, and (c) our own new model (Smoleń & Chuderski, 2010), which localizes the causes of the interference in the resolution of response conflicts. We focus only on how each model would handle increasing set sizes. For details of particular models see the original papers.

The model by Cohen at al. (1990) was a feed-forward network, which represented processing pathways for color naming and word reading as two separate interconnections of input, hidden, and output nodes, which shared only the output layer. Nodes which processed reading were associated more strongly than those for color naming. However, as an additional task-unit activated the color naming pathway, this pathway was able to determine a response, but at the cost of coping with the interference yielded by the other path. Though Cohen et al. simulated the Stroop task versions including two and ten stimuli, they did not directly compare interference effects generated by these two model versions. However, Kanne et al. (1998) attempted to test Cohen et al.'s model against the results of their experiment cited above. They extended the two-response architecture to account for three responses, and then for four responses. Surprisingly, the model showed the opposite behavior than the subjects, as the simulated interference effect decreased with larger set sizes. Cohen et al. (1998) responded to this test with a modified model. Three model versions had the same architecture, which accounted for multiple responses, but they differed in the number of stimuli/responses that attention had been allocated to. The simulations yielded virtually no size set effect (equaling 8.8, 8.9, and 8.9 model's cycles, for set sizes 2, 3, and 4, respectively).

Another influential Stroop model was developed by Roelofs (2003), who proposed distinct mechanisms for color and word naming, based on differences in assumed language production architecture. His model included three levels of word representations: concepts, lemmas (syntactic representations), and word forms. Color perception, via related concepts, activated a corresponding lemma, which had to be retrieved in order to select a proper word form. However, a perceived word was able to directly initiate the relevant lemma and form representations. In the Stroop task, color naming could be achieved by an additional selection process, modulated by a color concept representation, which acted as a goal. However, due to a shorter route from perception to response in the case of words, the interference emerged. Two other models, which are similar to Roelofs' in their assumptions about access to memory and which identify the loci of interference in memory retrievals, have been implemented in ACT-R cognitive architecture (Altmann & Davidson, 2001; van Maanen et al., 2009).

In the context of predictions of response set size effects, the most important assumption shared by all those models is that the latency of retrieval of a representation, which is crucial for color naming RT, depends on the activity of other memory representations related to the very task. For example, this assumption is precisely expressed by van Maanen et al., who implemented the RACE/A (retrieval by accumulating evidence in an architecture; van Maanen & van Rijn, 2007) theory of the time course of memory retrievals on short time scales. This model predicts that the latency of a retrieval (the time needed by the activation of a retrieved memory representation to reach a retrieval threshold) will be inversely proportional to the ratio of activation of the yet to be retrieved chunk to the sum of the activations of other relevant chunks, which also compete for retrieval. The ratio is expressed as a respective Luce's (1986) formula. Analogous formulations of retrieval rate can also be found in Roelofs (2003; Appendix A) and Altmann and Davidson (2001; Equation 1). Neither of these three models was used in order to replicate set size effects. However, on the assumption that all color names and color concepts relevant for a particular Stroop administration will form a set of potentially competing memory representations, such a pre-diction directly follows: with an increasing set size, the number of competing representations will increase, so the denominator of Luce ratio will increase. Thus, this very ratio will decrease, which will result in larger latency of the retrieval of color representations and, consequently, of color naming. As word reading, which is believed to be involved in congruent trials, is not sensitive to set size manipulation (MacLeod, 1991), some increase in latency difference bet-ween incongruent and congruent trails should be expected within each of the afore mentioned models.

Finally, we (Smoleń & Chuderski, 2010) have recently presented the architecture implementing the Stroop task. The architecture was primarily focused on describing executive control, so “ordinary” processes (like perception, visual attention, declarative memory etc.) were simplified. Elementary cognitive processing was represented as a choice among the set of potential cognitive and behavioral actions expressed as production rules. Rules were added to the choice set only if they matched the current context, defined as the contents of visual and central attention (the latter constituted the most available part of working memory). A utility value was assigned to each rule, which reflected the history of reinforcement-based learning related to the use of this rule. When no conflict was present, the model simply used the rule which would probably be the most successful one according to model’s learning history (i.e., the rule of the highest utility). However, if more than one rule matched the context, a conflict arose. The model estimated the amount of conflict as a ratio of the sum of all utilities of rules, which would lead to different actions than a rule of the highest utility, to the sum of all utilities. In other words, it estimated the conflict value as the function of how strong the competitors to the most dominant rule(s) are. Then, the model increased the goal-related control as a function of conflict. The less a given rule was related to the goal, the more this top-down control lowered its utility. Due to the control, a non-dominant but highly appropriate rule could be selected for further processing. However, the model assumed that the control takes time, so the more the utilities of goal-unrelated rules had to be decreased (i.e., the higher the conflict was), the longer the selection of the rule to be fired required, which finally resulted in larger RTs.

In the context of the set size effect, the most important property of our Stroop model is that only the rules which exactly match the visual input will be included in the choice set. So, assuming that each perceptual aspect of the stimulation (i.e., words and colors) related to the Stroop task is processed by one associated rule, and assuming that utilities for all rules processing words are equal and that the same is true for color processing rules (but the former have higher utility than the latter, as reading is trained more than naming colors), one can predict that the number of stimuli/responses will not affect the interference effect. No matter how large the set size is, both in incongruent and congruent trials only two rules will be considered, one which processes a color and one which processes a word. In congruent trials, both rules will lead to the same response, so the conflict will be low and latency small. On the contrary, in incongruent trials, a color naming rule will become a competitor to a word reading rule and thus the conflict and a resulting interference will be large.

In the administered experiment we aimed to test whether increasing the number of S-R mappings would increase the interference or would have no effect on it. Moreover, if there was an increase, it would be interesting to know whether it is the number of responses (i.e. potential response keys) that matters or whether it is the number of stimuli that counts. In order to answer this latter question, we designed experimental conditions which assigned more than one stimulus to one and the same response key.

Experiment 1

Participants

The recruitment was conducted via publicly available social networking websites in Krakow, Poland. Seventy nine women and forty three men participated (122 people in total). Mean age was 22.9 years (SD = 4.4, range 18 – 45). For a two-hour session each participant received the equivalent of seven euro in Polish zloty. All participants had normal or corrected-to-normal vision.

Materials and design

The figure-word analog of the Stroop task was used. The participants were randomly assigned to one out of four task conditions. Each condition involved either four or six stimuli and either two responses or the same number of responses as stimuli. This resulted in four conditions: four stimuli – two responses, four stimuli – four responses, six stimuli – two responses, and six stimuli – six responses.

Six geometric figures (approx. 5 cm × 5 cm in size) were used: square, rhombus, rectangle, circle, oval, and ring, presented in blue, with black outlines, on a gray background. Each stimulus was presented at the center of a computer screen. A word naming a figure, in Polish, printed in black (approx. 3 cm × 3 cm in size), was placed in the center of each figure. Congruent stimuli had the same meaning of the word as the shape of the figure. Incongruent figures were different than words. The same distractor word was always associated with a particular figure (e.g., “ring” was always put into a rectangle and vice versa). In each two-response condition, the distractor primed a response with the opposite hand than a hand associated with a target. Direct stimuli repetitions were not allowed. In four-stimuli conditions, only the square, rhombus, circle, and oval were used.

In each condition, the task started with a training sequence including 10 congruent and 30 incongruent trials in four-response condition or 15 congruent and 45 incongruent trials in six-response condition. Then a test sequence was presented in random order, which included 72 congruent and 48 incongruent trials in four-response conditions or 108 congruent and 72 incongruent trials in six-response conditions. The six-stimuli sequences were longer in order to give an equal number of presentations of each stimulus. Stimuli were presented for 2.5 s and then were followed by a mask which was shown for 1 s.

In two-response conditions, the square, rhombus, and rectangle (the latter only in the six-stimuli condition) were assigned to the ‘Z’ key, while the circle, oval, and ring (the latter only in the six-stimuli condition) were assigned to the ‘M’ key. In conditions with more than two responses, the square was assigned to the ‘Z’ key, rhombus – to ‘X’, rectangle – to ‘C’ (but only in the six-response condition), circle – to ‘B’, oval – to ‘N’, and ring – to ‘M’ (again, only in the six-response condition). So, one, two, or three fingers of each hand were dedicated to responding, depending on the condition. An instruction told participants to avoid reading and to press quickly and accurately the button assigned to the shape of the presented figure. In order to place equal demands on participants' working memory in both four- and six-responses conditions, the hints reminding stimulus-response mappings were placed at the bottom of the screen. Incorrect responses were signaled with a beep sound. The independent variables were: the number of stimuli (four or six), the number of responses (two or as mant as the no. of stimuli), and the trial (congruent/ incongruent) The dependent variable (DV) was the latency of correct responses directly following a correct response. RT less than 250 ms or more than 2200 ms were excluded.

Procedure

First, participants solved an analogy-making test which was not related to the present study. Then, the Stroop task was applied. Testing took place in a large, dimly lit room, in groups of no more than four people. Each participant was equipped with headphones and was sitting at a visually isolated desk.

Results

The mean proportion of errors was .047 and did not differ significantly between stimuli and response conditions. This indicated that correct respondse was not difficult in either four- or six-response conditions.

The mean latencies for all conditions are presented in Figure 1. All main effects were highly significant: participants responded more slowly in incongruent trials than in congruent ones, $F[1, 114] = 99.0$, $p < .001$, when six stimuli were involved in comparison to the case of four stimuli, $F[1, 114] = 115.5$, $p < .001$, and in one-to-one SR mapping conditions in comparison to two-response conditions, $F[1, 114] = 5655.4$, $p < .001$. A two-way interaction between the number of stimuli and the number of responses was also significant, $F[1, 114] = 36.2$, $p < .001$, and indicated that mean latency was larger when six stimuli were applied than when four stimuli were used, but it increased more in one-to-one SR mapping conditions than in many-to-one mapping conditions.

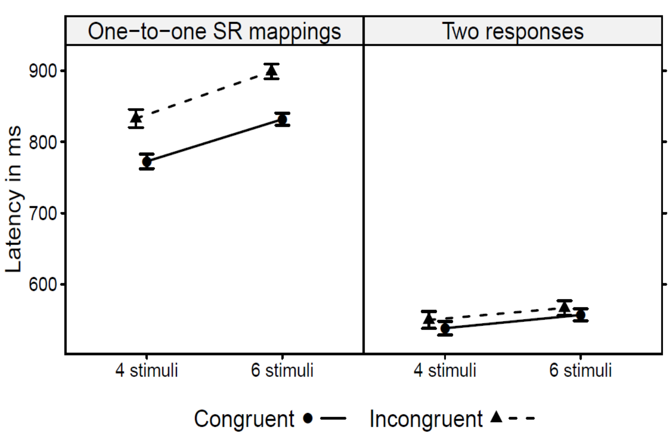

Figure 1: Mean response latency in all conditions of Experiment 1, for both congruent (solid lines) and incongruent (dashed lines) trials. Bars represent 95% confidence intervals.

For the purpose of the present study, the most important are the interactions regarding the type of trial and two other factors, namely whether both numbers (of stimuli and of responses) have any influence on the interference effect. The two-way interaction between the type of trial and the number of responses was highly significant, $F[1, 114] = 50.2$, $p < .001$, demonstrating that association of more than one stimulus with each response dramatically reduces the interference effect. However, the number of stimuli had virtually no effect on interference ($F = .1$) in any response condition (i.e., a three-way interaction was not significant, $F = .3$). The interference effect equaled 60 ms and 67 ms in four- and six-responses conditions, respectively, and fell to 11 ms on average in two-response conditions. However, the latter effect was significant, $F[1, 114] = 4.3$, $p = .038$.

No meaningful change in the above presented analyses occurred when only the first 120 trials from six-stimuli conditions were analyzed (i.e., when the equal numbers of trials for all conditions was taken into account).

Discussion

The main result yielded by this study is virtually no difference in the interference effects ($\Delta = 7$ ms) between four- and six-response conditions. Though the mean RT increased in the latter condition compared to the former one, it did so in equally the same amount for both congruent and incongruent stimuli.

A surprising result regarded two-response conditions. An extremely low interference effect was observed in both these conditions, in comparison to versions of the Stroop with one-to-one SR mappings, even if within the former conditions an incongruent stimulus always primed both competing responses. A question naturally arises: was this residual but significant interference observed in two-response conditions anyway related to response conflict or was no response conflict elicited, but the interference was rather related to the non-matching aspects of stimulus?

In order to answer this question a two-response condition, which includes the incongruent stimuli priming a single response, should be examined. If a small but significant interference effect still shows up, then it will suggests that factors other than response conflict are responsible for interference in manual versions of Stroop. However, if the effect disappears, then the conflict which is present in stimuli but is not related to responding, should not be taken into account as a factor causing the Stroop inter-ference. Thus, another experiment was administered in order to test these hypotheses.

Experiment 2

Participants

The recruitment procedure was the same as in Experiment 1. Twenty four women and twenty men participated (44 people in total). Mean age was 22.2 years ($SD = 2.8$, range 18 – 32). Again, for a two-hour session each participant received the equivalent of seven euro. All participants had normal or corrected-to-normal vision.

Materials, design, and procedure

The materials, DV, and procedure were analogous as in Experiment 1, with two exceptions. Firstly, only the four-stimuli, two-response condition was applied. Secondly, a distractor always primed the same response as one assigned to a target figure, namely the word "square" (in Polish) was always a distractor assigned to the shape of rhombus (and vice versa) and the word "circle" (in Polish) was always a distractor placed in the shape of oval (and vice versa). The only factor was the type of the trial.

Results

The mean proportion of errors was .053. The only relevant effect surpassed the adopted level of statistical significance, $F[1, 42] = 4.2$, $p = .042$. However it was a reversed effect: the mean latency observed in congruent trials was larger than the mean latency observed in incongruent trials (525 vs. 516 ms, respectively).

Discussion

No Stroop effect was found. When both a target figure and a non-matching word did prime the same response, the response latency was even smaller than when a word matched a figure. This unexpected result surely needs replication, perhaps with a material different than geometric figures. However, an unequivocal conclusion can be formulated that no amount of Stroop interference, which was observed in the manual task version, can be related to the very conflict in stimulus appearance, when a response conflict was eliminated in the incongruent trials.

General discussion

A methodological issue regarding the present research concerns whether increasing set size from four to six elements was indeed an effective experimental manipulation. We believe it was. On one hand, four is the lowest reasonable set size in Stroop tasks because, as we mentioned above, set sizes of two and of three impose methodological problems concerning stimuli repetitions and negative priming. On the other hand, the set size of eight elements is the upper limit that can be tested with manual versions of the Stroop. So, the choice of four and six SR mappings seemed reasonable. The substantial increase in mean response time for set size six, in comparison with set size four, indicates that indeed the task became more difficult in the former condition. However, in future studies it would be interesting to also test also set sizes of eight.

The fact that an increase in the number of stimulus-response mappings in the manual version of the Stroop has virtually no effect on the amount of the Stroop interference is in concord with the predictions of two models, which explain Stroop phenomenon as the resolution of conflict either between processing paths (Cohen et al., 1990) or between response tendencies (Smoleń & Chuderski, 2010). Although the former model was shown to wrongly predict RT distributions yielded by the Stroop task (Mewhort, Braun, & Heathcote, 1992), as well as to miss the effect of temporal asynchrony between the presentation of colors and words on the interference effect (Roelofs, 2003), at least in the case of the response set size effect, this model correctly predicts the observed data. Thus, the critique of the model made by Kanne et al. (1998) may have missed the point, because it relied on artefactual effects from the experimental design, which used only two stimuli. However, the lack of set size effect was an accidental rather than an intended property of Cohen et al.'s model.

On the contrary, our model's prediction on the lack of set size effects is a direct consequence of model's theoretical assumptions. The response conflict resolution, which is the main cause of the emergence of interference effects in the model, always relates to only those responses, which are primed by actual stimuli presented to the model. All other potential stimuli-response mappings, which are not related to the actual stimuli, have no effect on the value of the conflict being resolved. This assumption naturally also explains the lack of the interference effect observed in the many-to-one Stroop task version, in which non-matching aspects of a stimulus prime the same response. Our model predicts that in such a situation simply no conflict is present (i.e., there are no competing responses), so there is no need for conflict resolution and thus no interference is involved.

However, for both our and Cohen et al.'s models, the outstanding decrease in interference from one-to-one to many-to-one SR mappings would be a problematic phenomenon to explain. In the case of many-to-one SR mappings, in both models, two competing response tendencies/ processing paths (depending on the model) would still be activated and remain in conflict, leading to similar interference effects as in the one-to-one mapping task. In order to account for the effect of the ratio of stimuli set size to response set size, probably some additional assumptions would have to be adopted. For example, one might seek an explanation in a categorization processes, preceding the processing/response conflict. Maybe, before the activation of a certain path or response, the cognitive system dealing with many-to-one SR mappings firstly needs to categorize a stimulus as assigned to the proper response, and this process somehow stops the conflict and, consequently, lowers the interference. However, the present study does not provide any explanation as to why when there is a decrease in the number of potential responses, while a number of stimuli is constant, it results in such a huge reduction of Stroop interference. Surely, some further studies are needed.

Evidently, overtly expressed predictions of the models, which explain the Stroop interference in terms of additional memory retrievals needed for color or picture naming, were not supported by the present study. According to these models, an increase in the number of stimuli within a task should probably have made the color/picture name retrievals more difficult because there were more candidate memory chunks to be selected from. This should have resulted in a significant increase in Stroop interference with increasing set sizes. Such a prediction was not supported by results of our experiment, which applied the manual Stroop task. In fact, there is little evidence for this prediction even in the case of oral versions (but see La Heij & van den Hof, 1995; Williams, 1977). We suppose that if memory retrievals account for any part of Stroop interference at all, then they do it only when oral responding is involved, which may probably require recalling the elements of the motor program for oral response. On the contrary, manual response is much simpler, so it may require just activating the proper effector not mediated by any memory retrieval.

However, though these differences in experimental paradigms are superficial and not related to the main purpose of the Stroop task, namely measurement of the cognitive costs of executive control, they seem to result in discrepant experimental effects. Thus, it is very interesting whether one general model of the Stroop task, which would explain processing in both oral and manual versions of the Stroop, can be constructed. If it can, then it will constitute some evidence for the general control mechanism (see van Maanen et al., 2009) processing Stroop interference, which is responsible for a variety of Stroop-related phenomena. If such a model cannot to be found, then the view of executive control as a bunch of low-level, local regulatory processes (Egner, 2008), each focused on one specific type of conflict (e.g., regarding attentional focusing, memory retrievals, response selection etc.) would seem more probable, at least in the domain of coping with interference.

Acknowledgments

This work was sponsored by Polish Ministry of Science and Higher Education (grant N106 417140).

References

Altmann, E. M., & Davidson, D. J. (2001). An integrative approach to Stroop: Combining a language model and a unified cognitive theory. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 21-26): Hillsdale, NJ: Laurence Erlbaum.

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.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review, 97,* 332-361.

Cohen, J. D., Usher, M., & McClelland, J. L. (1990). A PDP approach to set size effects within the Stroop task: Reply to Kanne, Balota, Spieler, and Faust (1998). *Psychological Review, 105,* 188-194.

Egner, T. (2008). Multiple conflict-driven mechanisms in the human brain. *Trends in Cognitive Sciences, 12,* 374-380.

Gray, W. D. (2007). *Integrated models of cognitive systems*. New York: Oxford University Press.

Herd, S. A., Banich, M. T., & O'Reilly, R. C. (2006). Neural mechanisms of cognitive control: An integrative model of Stroop performance and fMRI data. *Journal of Cognitive Neuroscience, 18,* 22-32.

Kanne, S. M., Balota, D.A., Spieler, D. H., & Faust, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's (1990) connectionist model of Stroop performance. *Psychological Review, 105,* 174-187.

La Heij, W., & van den Hof, E. (1995). Picture-word interference increases with target-set size. *Psychological Research, 58,* 199-133.

La Heij, W., & Vermeij, M. (1987). Reading versus naming: The effect of target set size on contextual interference and facilitation. *Perception and Psychophysics, 41,* 355-366.

Luce, R.D. (1986). *Response times*. New York: Oxford University Press.

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

Mehwort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 872-882.

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

Smoleń, T., & Chuderski, A. (2010). Modeling strategies in Stroop with a general architecture of executive control. In S. Ohlsson, R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 931-936). Austin, TX: Cognitive Science Society.

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16,* 987-999.

Williams, T. (1977). The effects of amount of information in the Stroop color word test. *Perception and Psychophysics, 22,* 463-470.

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control. *Psychological Review, 115,* 518-525.