Ψ Psychology Press
Taylor & Francis Group

RESPONSE INTERFERENCE AND WORKING MEMORY IN 12-YEAR-OLD CHILDREN

John F. Stins, J. C. (Tinca) Polderman, Dorret I. Boomsma, and Eco J. C. de Geus

Vrije Universiteit, Amsterdam, The Netherlands

A group of 69 12–year–old children performed three well-known response interference tasks: the Stroop task, the Eriksen flanker task, and the Simon task. Individual differences in accuracy and speed correlated across the tasks. However, there was no correlation between the interference effects on these three tasks. Stroop interference, but not the Simon or flanker effect, was correlated with working memory capacity, as obtained from the WISC-R. These results may help clarify the nature of ADHD, which is characterized by problems with response interference.

Keywords: *Stroop, Simon, Flanker, WISC, attention deficit*

In order to select and execute an appropriate response, humans often have to ignore certain information that is strongly linked to another, yet inappropriate, response. This ability to deal with sources of conflict is known as interference control, and underlies our ability to achieve long-term goals, despite the presence of sources of distraction in the environment. Interference control is thought to be mediated, in part, by working memory (e.g., de Fockert, Rees, Frith, & Lavie, 2001). For some people inadequate interference control and working memory problems may cause significant problems in their daily lives, and is generally thought to underlie neuropsychological deficits such as observed in attention-deficit/hyperactivity disorder (ADHD), a common childhood disorder (e.g., Barkley, 1997).

Neuropsychological studies and clinical investigations often make use of tasks that are designed to index response interference. Often used tasks are the Stroop task and the Eriksen flanker task. In both tasks, the subject is presented with a stimulus that simultaneously activates two conflicting response channels; one response is activated by the instructions, whereas the other response is activated by elements in the array that strongly invite an alternative—yet incorrect—response. For example, in the Color WORD Stroop task (Stroop, 1935), the instruction is to attend to the color of the ink in which a word is printed and name this color aloud. At the same time, the printed words may also read certain color names that are different from the color of the ink in which it is printed. As has been observed on numerous occasions, there is a strong tendency to respond to the

content of the word, and not to the ink color. This is evidenced by an increase in response time and a decrease in accuracy relative to a neutral control condition. Moreover, this pattern of response interference seems to be exaggerated in ADHD. For example, Carter et al. (1995; see also Salo, Henik, and Robertson, 2001) showed that children with ADHD have more difficulty suppressing the response associated with the word content than controls, as evidenced by the reaction time (RT) profiles of the two groups.

In a similar vein, in the Eriksen flanker task subjects have to respond to one element in a stimulus array and ignore other elements in the array that designate an other response. An often used paradigm is one in which subjects have to respond to the direction of a left or right pointing arrow, and ignore flanking arrows that point in the opposite direction as the target arrow (Eriksen and Schultz, 1976). Similar to the Stroop task, there is a tendency to respond to the distracting flanker elements, as evidenced by an increase in response time, relative to a neutral control condition. With respect to ADHD, Crone, Jennings, and van der Molen (2003) found that children with ADHD spent more time resolving the conflict between the competing responses than controls, although Jonkman et al. (1999) only found that children with ADHD made more errors than healthy controls, whereas the RTs were essentially the same for both groups.

A third task that is less well known in clinical circles, but that also manipulates response conflict, is the Simon task, also known as the spatial conflict task (e.g., Simon and Rudell, 1967; Hommel & Prinz, 1997). In the Simon task, subjects have to give a left-right response to a non-spatial stimulus attribute, e.g., its color or its shape. At the same time, the stimulus is presented (in a random fashion) left or right from fixation. It is often found that when the stimulus and the side of the response happen to coincide, RT is some-what faster than when they are opposite. There thus seems to be a strong tendency to respond to the side of stimulation, even though side of stimulation is uncorrelated with the task-relevant stimulus dimension. Several authors have already pointed to the logical equivalence of the Stroop task and the Simon task (e.g., Hasbroucq & Guiard, 1991).

At first blush, these three tasks seem to measure the same underlying construct, i.e., the ability to select the appropriate response from a set of response alternatives. Indeed, there is evidence that these tasks make use of similar cognitive processes. The tasks are supported by similar brain regions, most notably the anterior cingulate cortex (ACC). For example, a study by Petersen et al. (2002), using event-related fMRI, showed that the Simon task and the Stroop task activate similar brain regions, most notably the ACC. The importance of the ACC has also been repeatedly observed in the Stroop task (e.g., Ridderinkhof et al., 2002) and in the flanker task (e.g., Botvinick et al., 1999). Most recently, Fan et al. (2003), using fMRI, found that all three tasks activated the ACC and the left prefrontal cortex, although the interference scores were uncorrelated. Other researchers make use of sophisticated behavioral paradigms, based on information-processing models. For example, the papers by Lu and Proctor (1995) and by O'Leary and Barber (1993) reveal strong commonalities between the Simon paradigm and the Stroop paradigm, although it should be noted that those studies examined spatial Stroop tasks and not color word Stroop tasks.

However, an unresolved question is to what extent performance on these three tasks is influenced by the same underlying latent variable. More specifically, it could be the case that some subjects have in general more difficulty in dealing with response conflict than others. This, in turn, would imply that the amount of response interference is compa-rable across tasks. To this end, we decided to have subjects perform all three response interference tasks, and use correlation analyses to determine which characteristics of the

task are highly interrelated. The results are likely of interest both to researchers interested in fundamentals of information processing, and to clinicians who have to pick a suitable task to index a patient's response inhibition capabilities.

Our primary measure will be the interference effect, i.e., the RT difference between the congruent and inconcongruent conditions, but we will also examine overall RT, accuracy, and post-error slowing. This latter measure refers to the finding that subjects tend to slow down their response (i.e., increase their RT) directly after they have given an erroneous response. Post-error slowing is thought to represent adjustment of performance strategies in the face of an error. Post-error slowing is often found in the RT-literature (e.g, Ridderinkhof, 2002), and there is evidence that children with ADHD fail to adjust their strategy, relative to controls (e.g., Swaab-Barneveld et al.,2000).

In addition to investigating the interrelationship between response interference tasks, we also study whether distractor interference is related to working memory (WM) capacity. Barkley's (1997) theory implied that there should be a positive association between an individual's working memory capacity and the ability to inhibit inappropriate responses, since both are though to be mediated by limited attentional resources. Two recent papers indeed found evidence for this relation: First, Stevens, Quittner, Zuckerman, and Moore (2002) had children with ADHD and controls perform both the stop-signal task (a task of response inhibition) and a working memory task. In the latter task subjects had to store (and later reproduce) the colors and the numerical values of a sequence of digits. The authors found, for both groups, a significant correlation between working memory capacity and behavioral inhibition capacity (indexed by the inverse of stop-signal reaction time). A second piece of evidence is reported by Long and Prat (2002), who found that individuals with large WM-capacity were less affected by incongruent Stroop words than those with less WM-capacity. The authors found that high WM-span individuals adopted a strategy to overcome Stroop interference. In the present experiment, we tried to find additional evidence for the existence of an association between working memory and response interference. To this end, we employed a more standardized measure of WM-capacity, namely based on the WM subtests of the WISC. We examined whether the effect applies not only to Stroop interference (as suggested by the study of Long & Pratt, 2002), but also to Simon and flanker interference.

SUBJECTS

The subject group consisted of 12-year-old twins and, if present, their additional brother or sister. The twins, but not the sibs, participate in a longitudinal study in attention and attention problems. The twins are registered in the Netherlands Twin Registry (NTR), which is hosted by the Vrije Universiteit of Amsterdam. The twins were randomly selected from the NTR subject pool. None of the children suffered from severe mental or physical impairments. A study by van der Oord, Koot, Boomsma, Verhulst and Orlebeke (1995) and by Posthuma, de Geus, Bleichrodt and Boomsma, (2000) revealed that there are no significant differences between Dutch twins and singletons on measures of problem behavior and intelligence, so that we can safely generalize from our sample to the general population.

Twin pairs (and siblings, if present) were first asked in writing whether they were willing to participate in the study. Permission was also asked of the parents or caretakers. If permission was granted, the families received further information on the study, and were invited to come to the campus site to do the tests. On the day of testing, both the children and their parents / legal representatives signed an informed consent form.

A total of 164 subjects participated. Given the potential phenotypic interrelatedness between the twin pairs we decided to randomly select one individual from each twin pair (both MZ and DZ), so that our sample for this study consisted of 69 individuals (32 males and 37 females) The data from the other twins and their siblings will be analyzed in a future genetic study.

PROCEDURE

Twins (and siblings of the same family) were tested on the same day. The children performed a range of neuropsychological tests and six subtests of the WISC-R. Short breaks were given between tests. The entire session lasted approximately four hours per child. All children received the tests in the same order.

The Simon task and the Eriksen flanker task were performed on a computer. Subjects were seated in front of a computer monitor and a panel of two response buttons (left and right). The monitor and the response buttons were approximately aligned with the vertical meridian of each participant's body

In the Simon task subjects were first presented with a white fixation cross for 500 ms. Immediately after the cross had disappeared, a red or a green disk (1.9 cm in diameter) appeared for again 500ms, either left or right from fixation. The distance between the fixation cross and the inner edge of the disk was 2.5 cm. Stimulus color and stimulus location were uncorrelated. Subjects were instructed to press the left key in response to a green disk, and the right key in response to a red disk, regardless of stimulus location. Subjects received 60 red stimuli and 60 green stimuli, in random order. Half of each stimulus type was presented left, and other half was presented right. The trials on which the stimulus location happens to be on the same side as the required response are the congruent trials; the other trials are the incongruent ones. Prior to the experiment subjects received 12 practice trials that were not analyzed. The Simon task lasted approximately 10 minutes, including practice.

In the Eriksen flanker task subjects were first presented with a white fixation cross for 500 ms, which was immediately followed by a horizontal array of 5 equally sized and spaced white arrows for 800 ms. The array was 10.5 cm wide. Subjects were instructed to attend to the direction of the center arrow, and ignore the four flanking ones. Subjects were instructed to press the left key to a left facing center arrow, and the right key to a right facing center arrow. The flanking arrows could either all point in the same direction as the target arrow (e.g., < < < < <), or they all pointed in the opposite direction (e.g., < < > < <). The trials on which the flanking arrows point in the same direction as the target arrow are the congruent trials; the trials in which they point in the opposite direction are the incongruent trials. Subjects received 40 congruent and 40 incongruent trials in a random order, requiring an equal number of left / right responses. Prior to the experiment subjects received 12 practice trials that were not analyzed. The Eriksen flanker task lasted approximately 10 minutes, including practice.

The Stroop Colour and Word Test (Stroop, 1935) is a well known measure of selective attention. Subjects complete 3 pages, each with 10 columns of 10 items. Subjects have to name the items from the top-left corner to the bottom-right corner. Page 1 involves naming the words "red," "green," "yellow" and "blue" printed in black ink. Page 2 involves naming the colours of squares that are printed in different colors. Page 3 involves naming the ink color that the words "red," "green," "yellow" and "blue" are printed in. In page 3 word content and ink color never match, i.e., all color words are incongruent. Speed and accuracy are stressed in the instructions. Each page is scored as the time (in seconds) to complete the page. Time is recorded by the experimenter using a stop-watch.

The Stroop task lasted approximately five minutes. Even though this version of the Stroop test does not permit a trial-by-trial analysis, as in the Simon and flanker task, we decided to adopt this version because of its widespread use in clinical settings.

The six subtests of the WISC-R were similarities, arithmetic, vocabulary, digit span, block design, and object assembly.

DEPENDENT VARIABLES

For both the Simon task and the flanker task, we calculated: (a) processing speed, i.e., overall RT; (b) accuracy, i.e., percentage correct; (c) the interference effect, i.e., the RT-difference between congruent and incongruent trials. (d) the reaction time after an error had been committed ($RT_{after error}$); and (e) reaction time after a correct response ($RT_{after correct}$). A comparison between (d) and (e) allows us to assess whether subjects modify their processing strategy immediately following an error, i.e., whether they exhibit post-error slowing. Processing speed and the interference effect were calculated using only correct trials, and using trials that were not faster then 150 ms and not slower then 1500 ms. Post-error slowing was only calculated for subjects who made at least two errors during the session, regardless of the RT value.

For the Stroop task we calculated (a) overall processing speed, i.e., the total time to complete Page 3, and (b) the interference effect, i.e., the RT-difference between the neutral colors page (Page 2) and the color-incongruent words page (Page 3).

For the WISC-R subtests, we first standardized the scores obtained with each test. Next, we added the scores on the arithmetic and digit span subtest. This yielded our measure of an individual's working memory capacity, which we call WM-capacity.

RESULTS

Due to technical problems the data obtained with the Eriksen flanker task of one subject were lost. Also, the Stroop test was not administered with one subject due to practical problems. One subject had extremely high error scores on the Simon test but not on the other tests. This subject had an error rate of 39%, whereas the next-highest error rate in our sample was 26%. Finally, one subject failed to complete the subtests of the WISC-R.

We first present the descriptive statistics, i.e., overall performance on the tasks, and next we present correlation analyses, showing whether, and to what extent, the individual's performance covaries among the tasks.

DESCRIPTIVES

The variables of interest, derived from the flanker task and the Simon task, are presented in Table 1.

We performed a $2 \times 2 \times 2$ repeated-measures ANOVA on the RTs with task (Simon vs. flanker) and stimulus type (congruent vs. incongruent) as within-subject factors and sex (male vs. female) as between-subjects factor. All within-subject effects were significant. The main effect of task indicated that, on average, subjects were 125 ms faster on the Simon task than on the flanker task, $F(1,65) = 214.08$, $p < .001$. The main effect of stimulus type indicated that subjects were faster responding to congruent trials than incongruent ones, $F(1,65) = 425.54$, $p < .001$. Finally, the interaction indicated that the interference

Table 1 Performance on the flanker task, Simon task, and Stroop task

Variable	Eriksen Flanker	Simon	Stroop (per item)
RT (in ms):			
congruent / neutral	556	464	708
incongruent	661	502	1189
interference	105	38	481
Accuracy (percentage correct)	96.4	90.9	n.a.

effect was task dependent, $F(1,65) = 94.41$, $p < .001$. The effect was larger in the flanker task in the Simon task. No effects involving sex were significant.

We performed the same ANOVA on the percentages correct, which are shown in Figure 2 as a function of all three variables. The following effects were significant. First, the main effect of task $F(1,64) = 75.11$, $p < .001$, indicated that more errors were made in the Simon task than in the flanker task. Second, the main effect of stimulus type indicated that subjects were more accurate responding to congruent trials than incongruent ones $(F(1, 64) = 69.03, p < .001)$. Finally, task interacted with sex $(F(1,64) = 11.74, p < .01)$. As can be seen from Figure 1, males were less accurate than females in the Simon task, whereas there was no sex difference in the flanker task.

In order to test whether subjects modify their processing strategy after an error, we first calculated for each task adjustment scores by subtracting mean RT $_{after\ correct}$ from mean RT $_{after\ error}$. For the flanker task we found that RT was increased on average 44 ms directly after an error had been made relative to correct trials. A T-test revealed that this value was significantly greater than 0, $T(31) = 2.079$, $p < .05$. In a similar vein, we found for the Simon task a 57-ms effect of post-error slowing, which was also significant, $T(62) = 5.852$, $p < .01$. An ANOVA revealed no effects of task and of sex of the subject.
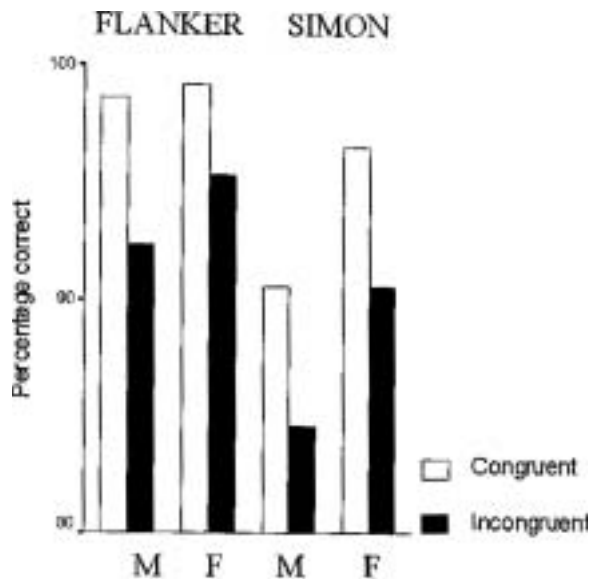
Figure 1 Accuracy (percentage correct) as a function of task (flanker or Simon), stimulus type (congruent or incongruent) and sex (M or F).

We analyzed Stroop performance by performing a 2 × 2 repeated-measures ANOVA on the completion times with Stroop page (Page 2 and Page 3) as within-subjects factor, and sex as between-subjects factor. Only the main effect of Stroop page was significant, $F(1,66) = 349.00$, $p < .001$. Performance on page 2 (neutral) was significantly faster than on Page 3 (incongruent): 70.84 vs. 118.96 s, respectively. Thus, we observed a robust 48-s Stroop effect, which is equivalent to an 480-ms effect per Stroop item.

WM-capacity ranged from 11 to 33 (mean 20.9). The correlation between the 2 subtests was .505, and was highly significant ($p < .001$).

RELIABILITY

From test theory we know that, in general, difference scores between two measures (such as interference scores in RT) will have high reliabilities when (a) the individual measures are highly reliable, and when (b) the correlation between the measures is low. Relatively high reliabilities of the Stroop test are often reported, but less is known about the reliabilities of flanker scores and of Simon scores. We estimated reliabilities of these tests by calculating the split-half correlations (based on the first and second 50% of the trials), and then applying the Spearman Brown correction. Both the flanker effect and the Simon effect had modest split-half reliabilities: .241 and .333, respectively. The split-half reliabilities for the congruent and incongruent RTs were much higher; for the flanker task they were .834 and .799, respectively, and for the Simon task they were .806 and .868, respectively.

CORRELATION ANALYSIS

Processing Speed

The correlation between overall response speed on the Simon task and the flanker task was .560, between the Simon task and the neutral Stroop page was .425, and between the flanker task and the neutral Stroop page was .404. All effects were significant at the .001-level.

Accuracy

The correlation between the number of correct responses for the Simon task and the flanker task was .559 and was highly significant ($p < .001$).

Distractor interference

The correlations between the Simon effect, the flanker effect, and the Stroop effect were all smaller than .2, and non-significant. The same results emerged when the Simon and flanker effects were expressed as percentage of overall RT. Figure 2 shows the scatterplot of the interference scores of the Simon task and the flanker task.

Working-Memory

The correlations between WM-capacity and the Simon and flanker effects were all smaller than .2 and non significant. However, the correlation with the Stroop effect was - .258 ($p < .05$). Thus, the larger an individual's WM-capacity, the less the Stroop interference.
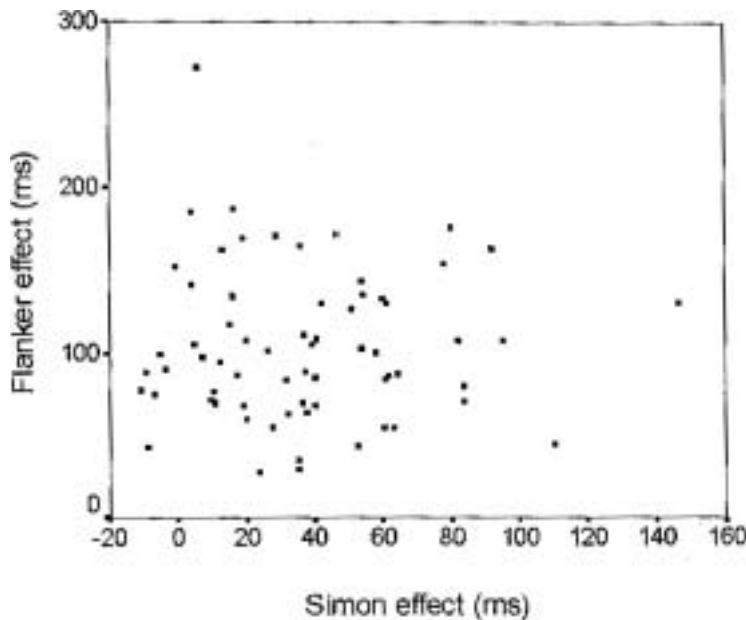
Figure 2 The size of the Simon effect in ms plotted as a function of the size of the flanker effect. Each point represents one child.

Correlation of WM-capacity with number of errors in the Simon task was not significant, but the correlation with number of errors in the flanker task was - .279. Thus, the larger the WM-capacity, the larger the accuracy in the flanker task. Finally, correlation with overall RT in the Simon task and the flanker task was also non-significant, although the correlation with the Stroop task was significant (-.49 for Page 2, and -.39 for Page 3). In other words, the larger the WM-capacity, the less time it takes to complete either version of the task. Finally, no significant interactions were found when we correlated WM-capacity with the amount of post-error slowing; neither for the flanker task nor for the Simon task.

Post-error Slowing

The correlation between the amount of post-error slowing for the Simon and flanker task was smaller than .3, and not significant.

DISCUSSION

In this study we compared performance across three response interference tasks, using a sample of 12-year–old children. First, we found evidence of substantial response interference; the RTs were higher in the incongruent conditions than in the congruent ones for all three tasks. Also, responding was less accurate in the incongruent conditions than in the congruent ones. Second, subjects responded more slowly directly after they had committed an error, regardless of task (Simon or flanker) and stimulus type. Third, as for the effect of sex, males appeared less accurate than females, but only in the Simon task. In addition, sex did not influence speed of responding. The literature on sex differences in

children on reaction time tasks is not very clear-cut, and sex differences appear to be strongly task-dependent, although there is some evidence for boys to be faster than girls (Lock & Berger, 1990). Fourth, we found that the larger an individual's working memory capacity, the less Stroop interference. We thus provided a qualitative replication of Long and Prat's (2002) results. However, we failed to observe substantial post-error slowing, or micro adjustments (Ridderinkhof, 2002) in both the Simon and flanker, but this variable did not correlate between the tasks, nor with WM. This could be due to the fact that subjects made only a handful of errors in each task, which of course limits the reliability of the measure. Future studies looking into this variable from an individual differences perspective will have to induce more errors, for example, by increasing the duration or the difficulty level of the task.

Finally, we studied whether the performance between our three tasks was correlated. If so, this would lend support to the claim that the tasks tap similar cognitive operations. The correlation analyses revealed low associations between performance across the three interference tasks. Although overall speed of responding between the three tasks was highly correlated, and although accuracy between the Simon and flanker task was also correlated, the amount of interference (i.e., the sizes of the Simon, Stroop, and flanker effects) was essentially uncorrelated. Thus, the extent to which a child suffers from incongruent Stroop words, say, has no predictive value of the distracting effects of incongruent flanker or Simon stimuli. An analysis of the reliabilities of the Simon and flanker task revealed that the low between-task correlations might have been due to the modest reliabilities of the interference scores of the tasks (.333 for the Simon task and .241 for the flanker task). This finding is at odds with that of Fan et al. (2002), who reported a .77 test-retest correlation for the flanker effect. This discrepancy, however, might have been due to the fact that Fan et al. (2002) tested adults, whose frontal executive network, though to be involved in interference control, is arguably more developed than that of our 12-year–olds. As evidence of this, paper by Rubia et al. (2000) reported that frontal activation, measured by fMRI, was stronger in healthy adults than in their adolescent counterparts. Although we have no data on the reliability of the Stroop effect, a study by Kindt, Bierman, and Brosschot (1996) showed low but significant reliabilities of interference scores of different versions of the Stroop test, and practically no significant correlations between the interference scores of the different Stroop formats.

A second reason for the low correlations between the interference scores may be that the tasks make use of similar, but not quite identical, cortical regions or cognitive resources. This was essentially the conclusion drawn by Fan et al. (2003), who studied the same tasks as ours, with some minor variations, using fMRI. Third, it may the case that the tasks make use of the same cortical regions or cognitive resources after all, but to varying extents. For example, the finding in our study that WM correlated only with Stroop interference and not with Simon or flanker interference could mean that the Stroop task (which demands verbal output and which involves a more diverse stimulus set) makes more heavily use of frontal executive resources than the two other tasks. Fourth, it may be the case that relatively uninteresting differences in task characteristics explain the lack of convergence. For example, in our version of the Stroop task congruent and incongruent items are blocked, whereas they are mixed in the Simon and flanker task. Also, in the Eriksen flanker task the incongruent items have greater visual complexity than congruent ones, whereas there is no such difference between congruent and incongruent Simon stimuli. Moreover, the timing in the tasks is different: the Stroop task is self paced, and subjects respond to each item at a rate of approximately 1 s. The Simon and flanker task, in contrast, is paced by the computer, and subjects receive a trail every 3 to 4 seconds.

To conclude, despite the fact that performance on these three tasks is hampered in ADHD, and that the tasks activate similar brain regions, our behavioral results revealed a low correspondence between the tasks, which suggest that the relationship between brain, behavior, and psychopathology is a truly complicated one.

REFERENCES

Barkley, R.A. (1997). Behavioral inhibition, sustained attention and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65–94.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., & Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181.

Carter, C.S., Krener, P., Chaderjian, M., Northcutt, C., & Wolfe, V. (1995). Abnormal processing of irrelevant information in attention deficit hyperactivity disorder. *Psychiatry Research, 56*, 59–70.

Crone, E.A., Jennings, J.R., & van der Molen, M.W. (2003). Sensitivity to interference and response contingencies in attention-deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry, 44*, 214–226.

De Fockert, J.W., Rees, G., Frith, C.D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291*, 1803–1806.

Eriksen, C.W., & Schultz, D.W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249–263.

Fan, J., Flombaum, J.I., McCandliss, B.D., Thomas, K.M., & Posner, M.I. (2003). Cognitive and brain consequences of conflict. *NeuroImage, 18*, 42–57.

Fan, J., McCandliss, B.D., Sommer, T., Raz, A., & Posner, M.I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340–347.

Hasbroucq, T., & Guiard, Y. (1991). Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 246–266.

Hommel, B., & Prinz, W. (Eds.). (1997). *Theoretical Issues in Stimulus-response Compatibility*. Amsterdam: North-Holland.

Jonkman, L.M., Kemner, C., Verbaten, M.N. van Engeland, H.,Kenemans, J.L., Camfferman, G., Buitelaar, J.K., & Koelega, H.S. (1999). Perceptual and response interference in children with attention-deficit hyperactivity disorder, and the effects of methylphenidate. *Psychofysiology, 36*, 419–429.

Kindt, M., Bierman, D., & Brosschot, J. F. (1996). Stroop versus Stroop: Comparison of a card format and a single-trial format of the standard color-word Stroop task and the emotional Stroop task. *Personality and Individual Differences, 21*, 653–661.

Lock, L.K. & Berger, R.A. (1990). Influence of sex, age and trial blocks on simple reaction times of elementary schoolchildren. *Perceptual and Motor Skills, 71*, 1379–1387.

Long, D.L. & Prat, C.S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition, 30*, 294–301.

Lu, C.-H., & Proctor, R.W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*, 174–207.

O'Leary, M.J., & Barber, P.J. (1993). Interference effects in the Stroop and Simon paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 830–844.

Peterson, B.S., Kane, M.J., Alexander, G.M., Lacadie, C., Skudlarski. P., Leung., H.-C., May, J., & Gore, J.C. (2002). An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cognitive Brain Research, 13*, 427–440.

Posthuma, D., de Geus, E.J.C., Bleichrodt, N., & Boomsma, D.I. (2000). Twin-singleton differences in intelligence? *Twin Research, 3*, 83–87.

Ridderinkhof, K.R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research*, *66*, 312–323.

Ridderinkhof, K.R., de Vlugt, Y., Bramlage, A., Spaan, M., Elton, M., Snel, J., & Band, G.P.H. (2002). Alcohol consumption impairs detection of performance errors in mediofrontal cortex. *Science, 298*, 2209–2211.

Rubia, K., Overmeyer, S. Taylor, E., Brammer, M., Williams, S.C.R., Simmons, A., Andrew, C., & Bullmore, E.T. (2000). Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. *Neuroscience and Biobehavioral Reviews. 24*, 13–19.

Salo, R., Henik, A., & Robertson, L.C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology, 15*, 462–471.

Simon, J.R., & Rudell, A.P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300–304.

Stevens, J., Quittner, A.L., Zuckerman, J.B., & Moore, S. (2002). Behavioral inhibition, self-regulation of motivation and working memory in children with Attention Deficit Hyperactivity Disorder. *Developmental Neuropsychology, 21*, 117–39.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28*, 643–662.

Swaab-Barneveld, H., de Sonneville, L., Cohen-Kettenis, P., Gielen, A., Buitelaar, J., & Van Engeland, H. (2000). Visual sustained attention in a child psychiatric population. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 651–659.

Van den Oord, E.J.C.G., Koot, H.M., Boomsma, D.I., Verhulst, F.C., & Orlebeke, J.F. (1995). A twin-singleton comparison of problem behaviour in2-3-year-olds. *Journal of Child Psyhology and Psychiatry, 36*, 449–458.