

The Stroop and Reverse Stroop Effects as Measured by an Interactive Tabletop

Holly C. Miller¹, Sébastien Kubicki², Delphine Caffier³, Christophe Kolski³, Janick Naveteur^{3,4}

¹Katholieke Universiteit Leuven, 3000 Belgium
hcmiller1661@gmail.com

²Lab-STICC UMR 6285 / UBO / ENIB, European Center for Virtual Reality
Technopôle Brest-Iroise F-29280 Plouzané, France
sebastien.kubicki@enib.fr

³Univ Lille Nord de France, F-59000 Lille
UVHC, LAMIH, F-59000 Valenciennes
CNRS, UMR 8201, F-59313 Valenciennes
{firstname.name}@univ-valenciennes.fr

⁴Univ Lille Nord de France, Lille 1, Neurosciences, F-59000 Lille

Correspondence concerning this article should be addressed to Janick Naveteur (Janick.Naveteur@univ-valenciennes.fr).

Abstract

An interactive tabletop equipped with RFID technology implemented the Stroop and Reverse-Stroop tasks. Participants moved a tangible object into one of four displayed virtual zones, which were identified either by a color border or the name of a color written in black. The correct target zone corresponded either to the color (Stroop) or meaning (Reverse Stroop) of a word displayed in congruent or incongruent colored LED lights (e.g., “yellow” displayed in yellow or red, respectively). Participants took more time to place the object in a zone, made more errors, and found the task more difficult when the word was lit with an incongruent than congruent color. This interference was influenced by both the types of task and response zone in a way that fits with a dual step processing account. Results also showed that the outcomes of the conflicting information streams might be judged as fun.

Keywords: Stroop effect, Reverse Stroop effect, Interactive tabletop, RFID, Tangible interaction, TangiSense

1. From Stroop’s cards to current technological supports

J.R. Stroop (1935) was the first to report that people take longer to verbalize the font color of a word when its meaning is incongruent (i.e., indicate “yellow” font when the written word is red) than they do to verbalize the color of a neutral stimulus such as a colored square. Since then, an overwhelming amount of research has explored this phenomenon (for a review see MacLeod, 1991), and has evidenced the validity and reliability of the Stroop effect. Stroop’s original study (1935) used a multiple-item card task, and performance was measured by timing the duration between the participants’ first and last response. Since then, computer versions have supplanted the card version, and the duration to process each item is now measured by the time required to press a corresponding (often arbitrary) key. Typewritten responses seem less often used despite providing a large Stroop effect (Logan & Zbrodoff, 1998). The size of the Stroop effect is smaller, with arbitrary-key presses, compared to recorded vocal responses with a voice-key. (MacLeod, 1991). To our knowledge, few computer versions have required a pointing response (Blais & Besner, 2007; Durgin, 2000, 2003).

2. Psychological Research on the Stroop Effect

The Stroop effect has been examined with varying methodologies, and evidence in favor of an account implicating response strength (mostly related to the degree of practice) and processing speed has accumulated (Blais & Besner, 2006; 2007; MacLeod & MacDonald, 2000). For instance, word meaning affects response time more than font color (Dunbar & MacLeod, 1984; MacLeod, 1991), a result that was first reported by Stroop (1935). When Stroop asked people to read the words printed in colors or printed in black, interference was not reliable. However, this interference, called reverse Stroop effect (RSE), increased after few days of practice (Stroop, 1935, Experiment 3). Later modifications of the standard Stroop task have produced stronger and more reliable RSE. As an illustration, Melara and Mounts (1993) obtained an RSE when the displayed words were small in size and difficult to identify. Dunbar and MacLeod (1984) observed an RSE when participants were told to identify words printed upside down and backwards and when many other non-response-set items were included in the task (Dunbar & MacLeod, 1984, Experiment 4). These RSE observations are of theoretical interest because they imply that cognitive interference occurs when two types of information are processed at different speeds.

In the standard Stroop paradigm, the word meaning is processed more quickly than color information, and thus interferes with responding. When modified such that color information is processed more quickly (by rendering the reading of word stimuli more difficult), an RSE is observable. Of relevance to the present study, an RSE has also been reported for participants who were asked to identify the meaning of a word by pressing a key on a response box which corresponded to the position of the color patch on a slide (Flowers, 1975) or by pointing to a location on a computer screen containing the corresponding color patch (Durgin, 2000; 2003). Supposedly, identifying the color by name requires a translation of lexical code into visual code, whereas responding to the physical color was accomplished without need of labeling the color. This is in line with the dual coding theory (Song & Hakoda, 2015), according to which the superficial structure of the stimuli, i.e. the imaging code (ink color) and the phonological code (referred to as “logogen code” by the authors) is first extracted, and then the coding corresponding to the response modality is produced (transformation of the imaging coding into phonological coding or vice versa according to the instructions). However, Blais and Besner (2007) also reported an RSE with a pointing task where the response zones were labeled with color names. There was no need here for a translation between the stimulus and the response codes (lexical code in both cases) and the authors proposed an explanation in terms of stimulus-response strength.

To date, the dependent variables of the Stroop task have focused on response time and/or error rates. Subjective difficulty was seldom taken into account (e.g., Boutcher & Boutcher, 2006). It is not known whether variations in the relative occurrence of Stroop and RSE performance are related to a coherent pattern of variations in the subjective evaluation of task difficulty. Yet this possibility, which would take into account the influence of response time and errors in the construction of the subjective experience of the participants, is relevant for understanding interference effects and also for the implementation of Stroop-like tasks as a game on new technological supports.

3. Hypothesis development

The current study was designed to address several issues related to interference between perceptual and conceptual information with different procedures implemented on the TangiSense (Kubicki, Lepreux, Lebrun, Dos Santos, Kolski, & Caelen, 2009) interactive tabletop. Tabletops have a large surface for interaction (Manches, O'Malley, & Benford, 2009, Fitzmaurice, Ishii, & Buxton, 1995) and mostly propose virtual objects manipulations. TangiSense is equipped with Radio Frequency Identification (RFID; Finkenzeller, 2003) which also permits tangible (Ishii, & Ullmer, 1997) objects detection (Lebrun, Adam, Mandiau, & Kolski, 2015; Baudisch, Becker, & Rudeck, 2010; Jordà, Geiger, Alonso, & Kaltenbrunner, 2007). We can also note that some new research aims to propose distributed interface using interactive tabletops (Jetter, Zöllner, Gerken, & Reiterer, 2012; Lepreux, Kubicki, Kolski, & Caelen, 2012; Kolski et al., 2014). Given the multidimensional aspects, concerns regarding interference are relevant, as perceptual and conceptual conflicts could impact user performance (Hapeshi & Jones, 1992). Previous unawareness of this possibility has already resulted in a Stroop-like impairment in preschoolers who were playing a color learning application (Kubicki, Lepreux, & Kolski, 2011).

The operational goal of the study was to record objective performance (response time and number of errors) and subjective performance (subjective difficulty) with a two-by-two between-subjects design. The standard Stroop and Reverse Stroop tasks were to be performed with either physical color or color word meaning as response guides. In order to respond, participants were required to move a tangible object into a target response zone identified by either a colored border or the name of a color written in black. There were four response zones that corresponded to the four color-related stimuli. These stimuli (that were presented one at a time) were words of colors written using incongruent LED lights (e.g., “red” was shown in yellow) and the participant was required to either guide their response by the color of the LED lights (for the above example, towards the yellow response zone or the zone in which the word yellow was written), or by the meaning of the word (for the above example, towards the red response zone or the zone in which the word red was written). Control trials were used to assess the effect of competing information on response behavior, the stimulus being written words of color using congruent LED lights (e.g., “red” was shown in red). Congruent control trials have been regularly adopted, though they tend to facilitate response speed (Sichel & Chandler, 1969; MacLeod, 1991) relative to control trials that consist of colored squares, XXXX’s, or arbitrary words. The control trials were intermixed with experimental trials, similar to the procedure of Fernandez-Duque and Knight (2008).

Given the robustness of color/name interference, our primary hypothesis was that users would experience interference on incongruent relative to control trials despite the fact that the transposition of this task to an interactive tabletop necessitated rather long tangible object displacements that could hypothetically increase response time variability and decrease number of errors due to possible trajectory correction.

The second hypothesis was that the type of task, i.e. standard Stroop or Reverse Stroop and the type of response zone (i.e. physical color or word target zone), would influence performance¹. In accordance with the translational account, more interference was expected for the normal Stroop task performed with the word-response zones and the reverse Stroop task performed with the color-response zones. As illustrated above (Blais & Besner, 2006, 2007; MacLeod & MacDonald, 2000) the possible relevance of other processing factors (processing speed and stimulus-response strength) did not lead us to expect the full disappearance of interference in the other basic conditions, at least with the normal Stroop task performed with the color-response zones if the word reading process is not totally inhibited. Nevertheless, the RSE was expected to be larger than the Stroop effect with the color-response zones and the reverse was expected with the word-response zone.

Third, the study investigated whether the interference in both the type of task and the type of response zone influenced the subjective difficulty experienced by participants. This search was motivated by the fact that subjective difficulty is an important facet of user experience but it is not a simple by-product of performance. It might itself influence performance (Maynard & Hakel, 1997; Mangos & Steele-Johnson, 2001), especially by modulating the vigor at which a task is undertaken (Kukla, 1972; Gendolla, Wright, & Richter, 2012). Emotional experience is also known to influence performance (Beal, Weiss, Barros, & MacDermid, 2005). According to Washburn (2003), the difference between a computer-based task, such as the Stroop task, and a computerized game is not clear-cut. The distinction might be measurable by the link between some kind of hedonic enjoyment and objective, subjective performance.

4. Research methodology

4.1. Participants

Eighty right-handed young adults (age range 18-34; M=22.7 years) whose native language was French and who had normal or corrected to normal vision were recruited at the University of Valenciennes and Hainaut-Cambrésis, France. There were 40 male and 40 female participants that were divided equally into four groups.

4.2. Materials

Description of the tabletop. The TangiSense interactive tabletop was designed by the RFIDées Company, and is made up of “tiles” of 2.5 cm², each containing 64 antennas (8 x 8). Each tile contains a DSP processor, which reads the RFID antennas, the antenna multiplexer, and the communication processor. The

reading strategies are prioritized, and the code is distributed between the processor reader antennas, the processor in charge of multiplexing, and the host computer.

The interactive tabletop used in the study measures 1 m square and contains 25 “tiles” slabs (5 x 5) of 1,600 antennas in total. Each antenna contains four RGB LEDs, which enable the display of virtual objects including colored zones, and word stimuli. The tiles are associated with a control interface connected to the host computer by an Ethernet link; for more details, the reader can see Lebrun, Adam, Kubicki, & Mandiau, (2010) or Kubicki, Lepreux, & Kolski (2012). For the description of the multi-agent software architecture of the tabletop, the reader can see Lebrun et al., 2010 or Kubicki, Lebrun, Lepreux, Adam, Kolski, & Mandiau (2013).

Description of tangible object and interface. The tangible object was a clear cylinder (4 cm diameter, 1 cm tall); see Figure 1. The start zone was a dark circle (8 cm diameter) outlined in white LED lights located 3.5 cm from the edge of the tabletop and centered horizontally. The stimuli consisted of words presented in colored LED lights (Red, Green, Yellow, Blue). The words projected (“RED”, “GREEN”, “YELLOW”, “BLUE”) were written in French with uppercase letters (“ROUGE”, “VERT”, “JAUNE”, “BLEU”), 11.5 cm tall, centered horizontally, and 70 cm from the start zone. The response zones were four dark circles (18 cm diameter) located an equidistance (32 cm) from the start zone, and arranged in a semi-circle. For half the participants these circles were outlined in one of four colors with LED lights and, for the other half, they were outlined in white LED lights and a white label (2 x 10 cm) was placed within the circle with the color words (“ROUGE”, “VERT”, “JAUNE”, “BLEU”) written in black (Arial police, uppercase letters, 2 cm tall)². The feedback stimulus presented for errors was as an “X” (14 cm tall) that appeared 11 cm above the start zone in white LED lights (see Figures 1 & 2).



FIGURE 1. The TangiSense Stroop task.

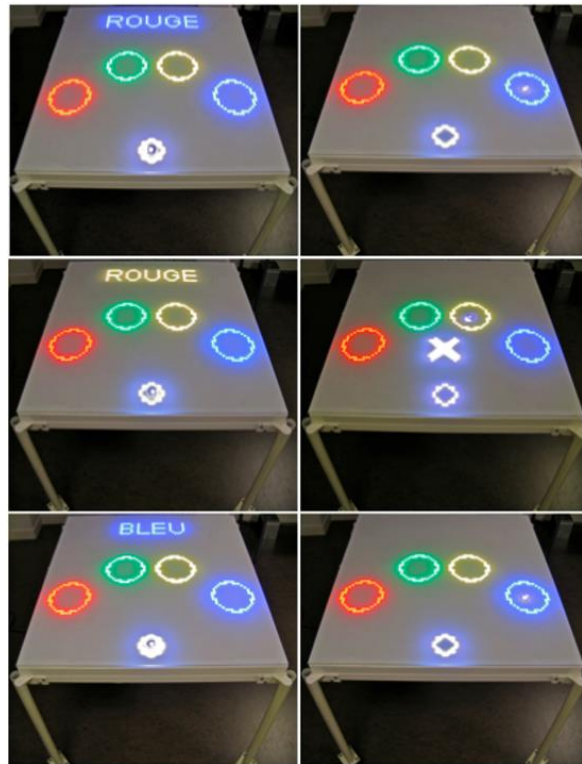


FIGURE 2. The TangiSense Stroop task. An incongruent trial where “RED” (“ROUGE” in French) is lit in blue (top left). A correct response for those in the STROOP condition (top right). An incongruent trial where “RED” is lit in yellow (middle left), and the incorrect response for this trial for those in the REVERSE Stroop condition (middle right). A congruent trial where “BLUE” (“BLEU”) is lit in blue (bottom left), and the correct response for this trial (bottom right).

4.3. Procedure

Participants were invited to play an application designed for the TangiSense interactive tabletop. The COLOR and WORD conditions were conducted at different times, and a random assignment was used for the standard Stroop or Reverse Stroop, which were run concurrently for each condition. Upon arrival participants signed a consent form that described the application. It was written that they would be asked to move a tangible object from a start zone to one of four equidistant colored zones. A word representing a color would be illuminated with colored lights on the tabletop. Participants in the two STROOP conditions were instructed to use the color of the lights to guide their responses and to ignore the meaning of the word, whereas those in the two REVERSE conditions were instructed to do the opposite.

Once the word was presented, the objective of the participants was to place the tangible object in the correct target zone as quickly and accurately as possible using their dominant right hand. As a reminder, the only difference between the COLOR and WORD conditions was the appearance of the response zones. The experimenter then verbalized the instructions and demonstrated how the tabletop worked. When the participants indicated that they were ready to begin, they were instructed to put the object in the start zone to initiate the first trial. The experimenter remained present in the room.

Each testing session consisted of 160 trials, with an equal number of congruent and incongruent trials presented in a randomized order. A testing trial was initiated once the tangible object (on which was placed an RFID tag) was placed in the start zone. One second after the RFID object tag was detected in the start zone a word (“GREEN”, “RED”, “BLUE”, “YELLOW”) was illuminated with either congruent or incongruent LED lights. Once displayed, a timer was activated to record the response time, i.e. the time to displace the object from the start zone to a response zone. Once detected in a response zone, a white halo appeared underneath the object regardless of whether it was placed correctly or incorrectly (see Figure 2). Further movement of the object (corrections) was not allowed nor was recorded. Accuracy was measured by recording the number of correct placements. An inaccurate placement was followed by the illumination of the white X. The feedback

was displayed until the participant moved the tangible object from the response zone to the start zone to begin the next trial. The participant was allowed to return the object at his/her own pace. The task lasted approximately 10 min.

Once the task was completed, participants were asked to answer questions using a 10 cm long visual analogue scale (VAS). They rated the perceived difficulty of congruent and incongruent trials (designed respectively by “when words were – or were not – written in their own color”) using a VAS anchored with “very easy” and “very difficult”. They also rated how boring or amusing they found the task overall (called “fun score”), using a VAS anchored with “very boring” and “very funny”. The total duration of participation was approximately 15 min.

4.4. Data analyses

The response time data for each participant was averaged across all trials and any response time that was 2 standard deviations above the individual’s average was eliminated for comparison regardless of trial type. This led to the exclusion of 1.7 % of the trials. For the response time, the overall average by trial type for each participant was then log transformed. This was done to correct for the positive skew inherent to response time data. These transformed averages were then subjected to analysis. The number of response errors was counted by trial type for each participant.

The response time and perceived difficulty were analyzed using a four-factor analysis of variance (ANOVA) with repeated measures on this factor (trial type: CONGRUENT vs. INCONGRUENT) and independent measures on the other factors (task: NORMAL STROOP vs. REVERSE STROOP; response zone: PHYSICAL COLOR vs. WORD; gender: FEMALE vs. MALE)³. The ANOVA performed on the fun score did not include the trial type factor. The LSD test was used for post-hoc analyses. Owing to an uneven distribution, non-parametric statistics were used to analyze the number of errors. Within-participants comparisons were performed with the Wilcoxon signed-rank test and between-participants comparisons were performed with the Kruskal-Wallis test complemented by the Mann-Whitney U-test. Attention was paid to effect size estimates (partial η^2 following the ANOVAs, and r following non-parametric comparisons). Two-by-two correlations were calculated between the three trial type differences (Incongruent – Congruent trials, I-C) obtained for (1) response time, (2) number of errors, and (3) subjective difficulty, as well as between these differences and the fun score. The Spearman’s rho test was used for correlations involving the number of errors and the Pearson product moment correlations was used for the other analyses; the decision was controlled for the false discovery rate (Benjamini & Hochberg, 1995). For concision, the four basic conditions for task and response zones are labeled below NStroop-C for normal Stroop task performed with the color response zones, NStroop-W for normal Stroop task performed with the word response zones, RStroop-C for reverse Stroop task performed with the color-response zones, and RStroop-W for the reverse Stroop task performed with the word-response zones.

5. Results

As expected, **response time** differed significantly by trial type $F(1, 72) = 114.11, p = .000, \eta_p^2 = .61$, evidencing that participants responded more quickly for the congruent than incongruent trials (mean raw difference : 0.112 s, corresponding to an increase of 8,25 % of the response time recorded in the congruent condition); a finding that was true for every basic condition ($ps < .0006$). There was a main effect of response zones ($F(1, 72) = 13.70, p = .001, \eta_p^2 = .16$), participants responded more quickly with the color-response zones than the word-response zones. There was no main effect of task ($p = .98$). There was a significant interaction between trial type, task and response zone $F(1, 72) = 8.41, p = .005, \eta_p^2 = .10$ (see Figure 3). ANOVAs conducted for each response zone separately confirmed that, with the color-response zones, a significant interaction between trial type and task arose ($F(1, 36) = 4.18, p = .048, \eta_p^2 = .10$) indicative of a larger RSE; with the word-response zones the interaction was in the opposite direction ($F(1, 36) = 4.28, p = .046, \eta_p^2 = .11$) indicative of a larger normal Stroop effect. ANOVAs performed on each task separately showed that, in the normal Stroop task, there was a significant interaction between trial type and response zone ($F(1, 36) = 6.20, p = .018, \eta_p^2 = .14$), indicative of greater interference with the word-response zones than with the color-response zones (reverse Stroop task: $F(1, 36) = 2.42, p = .128, \eta_p^2 = .06$). For comparative purposes with the results for the number of errors, the interaction between the three factors could also be presented by focusing on two-by-two comparisons performed on the interference (difference between incongruent and

congruent trials). There were no significant differences between NStroop-C and RStroop-W ($p = .639$), and between NStroop-W and RStroop-C ($p = .424$). The I-C differences were larger in the two latter than in the two former, but significance was reached only for the comparisons involving NStroop-W ($p = .008$ with NStroop-C and $p = .026$ with RStroop-W; for the comparisons involving RStroop-C: $p = .056$ with NStroop-C and $p = .147$ with RStroop-W).

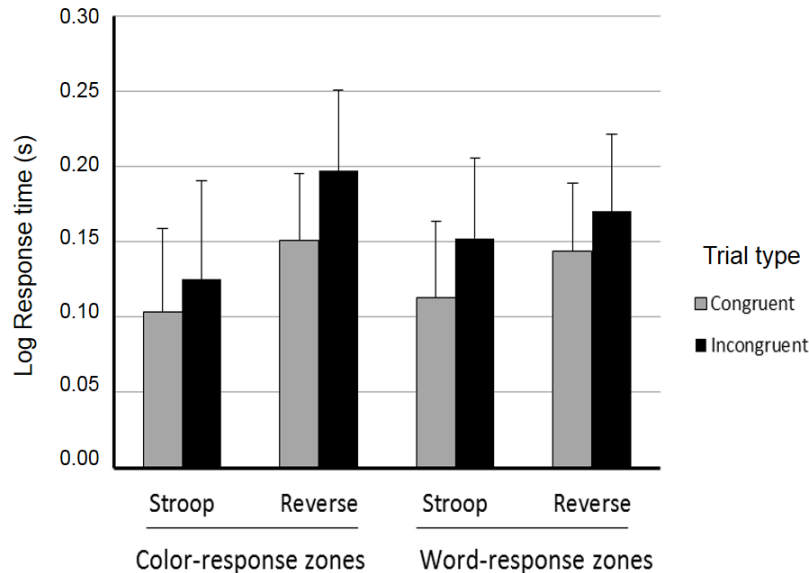


FIGURE 3. The mean response time for congruent and incongruent trials by Stroop task and type of response zone.

The **number of errors** varied between 0 and 19 (less than 12 % of the trials). It differed significantly by trial type ($Z = 5.58$, $p = .000$, $r = .62$, median difference between incongruent and congruent conditions: 1 error, range -1, +13) and the difference between trials was significant in every basic condition ($ps < .05$). The ANOVA performed upon the I-C difference revealed a significant difference between the four basic conditions ($H(3, 80) = 1.33$, $p = .002$, with the following sums of ranks: 528.0 for RStroop-W, 678.5 for NStroop-C, 925.5 for RStroop-C, and 1054.0 for NStroop-W). The two by two comparisons showed that there were no significant differences between RStroop-W and NStroop-C ($p = .432$), and between RStroop-C and NStroop-W ($p = .171$). The I-C differences were larger for the two latter than for the two former ($U = 99.5$, $p = .006$, $r = .25$, for RStroop-W – NStroop-W; $U = 101.5$, $p = .007$, $r = .30$ for RStroop-W – RStroop-C; $U = 107$, $p = .011$, $r = -.28$ for NStroop-C – NStroop W), but significance was not reached for the comparisons between NStroop-C and RStroop-C ($U = 132.5$, $p = .067$, $r = -.20$).

The ANOVA revealed that the **difficulty ratings** differed significantly by trial type $F(1, 72) = 137.89$, $p = .000$, $\eta_p^2 = .66$, such that participants perceived the incongruent trials as more difficult than the congruent trials (mean raw difference : 2,68, corresponding to an increase of 233,8 % of difficulty rating recorded in the congruent condition); the trial type difference reached significance in every basic condition (all $ps < .001$). The main effects of task and response zone were not significant ($ps < .10$). There was a significant interaction between trial type, task and response zone $F(1, 72) = 19.37$, $p = .000$, $\eta_p^2 = .21$ (see Figure 4). With the color-response zones, there was a significant interaction between trial type and task type ($F(1, 36) = 6.90$, $p = .013$, $\eta_p^2 = .16$) indicative of a larger RSE. With the word-response zones, the trial type x task interaction ($F(1, 36) = 12.76$, $p = .001$, $\eta_p^2 = 0.26$) was indicative of a larger normal Stroop effect. For the normal Stroop task, a significant interaction between trial type and response zone ($F(1, 36) = 10.67$, $p = .002$, $\eta_p^2 = 0.23$) was indicative of a larger interference for the word-response zones than for the color-response zones, whereas the opposite interaction arose for the reverse Stroop task ($F(1, 36) = 8.69$, $p = .006$, $\eta_p^2 = 0.19$). For comparative purposes with the previous set of results, the two by two comparisons performed on the interference showed that there were no significant differences between NStroop-C and RStroop-W ($p = .813$), and between

RStroop-C and NStroop-W ($p = .370$). The four comparisons between these two sets of data always reached significance (p ranging from .011 to .0004).

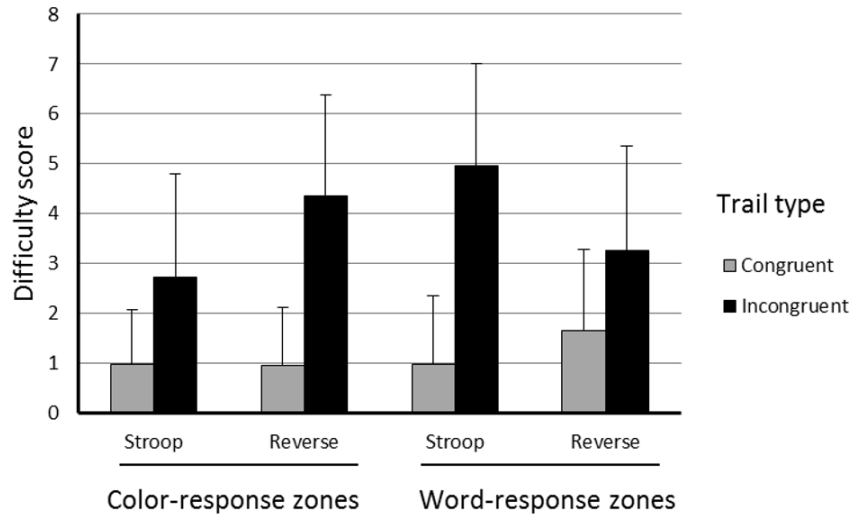


FIGURE 4. The mean difficulty rating (distance in centimeters on the visual analog scale for the questions assessing how difficult the trials were, where 0 = very easy, and 10 = very difficult) for the congruent and incongruent trials by Stroop task and type of response zone.

Table 1 summarizes the correlations between the trial type differences (Incongruent – Congruent trials) for (1) response time, (2) number of errors and (3) subjective difficulty. The correlations between these three differences and the fun score were also calculated. There was a positive significant correlation ($\alpha > \text{FDR}$ alpha) between, on the one hand, I-C difference for number of errors and, on the other hand, the I-C difference for response time and for subjective difficulty. The fun score was positively related to the I-C difference for both subjective difficulty and number of errors. The ANOVA performed on the fun score revealed a significant interaction between the type task and the response type ($F(1, 72) = 4.37, p = .040, \eta_p^2 = .06$), but the post hoc analysis pointed to only a trend to rate RStroop-W as less fun ($M = 5.07, SD = 2.07$) than NStroop-W ($M = 6.48, SD = 2.25; p = .053$) and RStroop-C ($M = 6.46, SD = 2.17; p = .057$; NStroop-C: $M = 5.76, SD = 2.39$).

Table 1. Correlational analysis of the interference (Incongruent-Congruent trials, I-C) for response time, number of errors and subjective difficulty and their link with the fun score (10-cm VAS anchored with boring and amusing).

		Coefficient*	Alpha	FDR alpha
I-C difficulty	Fun score	0.450	0.00003	0.008
I-C difficulty	I-C errors	0.332	0.003	0.017
I-C time	I-C errors	0.272	0.015	0.025
I-C errors	Fun score	0.265	0.017	0.033
<i>I-C difficulty</i>	<i>I-C time</i>	<i>0.150</i>	<i>0.184</i>	<i>0.042</i>
<i>I-C time</i>	<i>Fun score</i>	<i>-0.089</i>	<i>0.609</i>	<i>0.050</i>

* Pearson's r , except Spearman's rho for I-C errors. Decision controlled for the False Discovery Rate (Benjamini & Hochberg, 1995): significance is reached if $\alpha > \text{FDR}$ alpha. (an italic font was used for non-significant effects).

6. Discussion

Interference between perceptual and conceptual information was observed for every variation of the standard Stroop and Reverse Stroop tasks implemented with the TangiSense tabletop. Participants made more errors, and responded more slowly for incongruent than congruent trials. Incongruent trials were also perceived as more difficult than congruent trials. Our first hypotheses were thus upheld. Interference was observed despite the need for extended movements on the interactive tabletop, which made the response output-stage much greater than that required for comparable research requiring moving a cursor to a target on a computer screen (Blais & Besner, 2007; Durgin, 2000, 2003).

Results are indicative of an interaction between trial type and both, task and response zones for the three dependent variables. The relevance of the translational account is sustained by focusing on the influence of the response type in each type of task separately, i.e. when the required response was intrinsically identical but was provided by moving the tangible object towards either a perceptual or a conceptual target. A translational account is often used to explain the effects observed for the standard Stroop task requiring verbal responses (e.g., Glaser & Glaser, 1989; Sugg & McDonald, 1994): when a color has to be named, it must be translated from a color subsystem (visual) into a linguistic subsystem. Such an effect could explain the difference between the normal Stroop performed with the word and the color response zones that we observed for the reaction time, the number of errors and the subjective difficulty: a translation was required for the word response zones whereas no translation was required for the color response zones. According to Durgin (2000), there is also the possibility that translating perceptual color information (processed in a visual subsystem of the brain) into an action requires that verbal information (processed in a linguistic subsystem) be translated and thusly contributes to the RSE observed in the pointing tasks towards color patches (Durgin, 2000; 2003; Song & Hakoda, 2015)). In the present study, such a translation was required for the reverse Stroop task performed with a response given with the color response zones but was not required for the word response zones. This could therefore explain the greater number of errors and subjective difficulty for the former as compared to the latter. Despite non significance for distinguishing the two versions of the task with the word-response zones, our between-participant design did not evidence any clear cut quantitative differences between translations performed for the normal Stroop and for the reverse Stroop⁴, suggesting that the main source of interference at the response step is the switching from one code to the other, be the former perceptual or conceptual in nature. Besides, the overall longer response time for the word response zones as compared to the color response zones suggested that the localization of the chosen response zone was easiest with the color circles than with the color name written in black on white labels within white circles.

Interference still occurred, although less so, for the normal Stroop task performed with the color response zones and for the reverse Stroop task with the word response zones, despite that in both cases, no translation was required during the output stage. These findings argue that additional sources of interference may be found at the initial step of processing. A classical idea is that, for the Stroop task, the competition between perceptual and conceptual coding benefits the latter, probably because of response speed and/or response strength, and that interference arises when the instruction requires a perceptual code. For the reverse Stroop, the instruction requires a conceptual code, so the occurrence of interference is more puzzling. However, this result is in line with some previous occurrences of reliable RSE when the reading was rendered more difficult (Melara & Mounts, 1993; Dunbar & MacLeod, 1984). Such type of explanation may be related here to the material itself, specifically the LEDs. Their brightness may have increased the primacy of the font color, whereas the reading of the words may have been somehow hindered by the fact that the inner LED structure of the letters was clearly visible, which could have impeded reading. Since there were no significant differences in interference between these two conditions, it could be that the material used in the present experiment lead to a more equal difficulty to inhibit the physical coding and the conceptual coding according to the instructions at the first step of processing, and not to a strong dominance of reading as initially considered at the beginning of this paragraph.

There is no reason to believe that the interference at the initial step was not elicited in the four conditions differing by the task and the response zone. Therefore interference observed for both the normal Stroop task performed with the word response zones and the reverse Stroop task performed with the color response zones cannot be seen as the mere product of a translational effect. This conclusion supports a dual

coding theory (Song and Hakoda, 2015), with a first step extracting the imaging code (ink color) and the logogen code (phonological representation), and a second step that requires or not a transformation into a response code. However, the results are in line with the idea that interference can be produced at both steps by different underlying mechanisms in a cumulative way. The respective relevance of each step of interference is under the dependence of procedural details and material characteristics. This means that future search for neuroanatomical mechanisms must carefully identify these processes in the task at hand in the attempt to find neuroimaging evidence of dissociation between Stroop and reverse Stroop (Ruff et al., 2001; Song & Hakoda, 2015).

Despite that it has been seldom recorded in Stroop task experiments, the subjective difficulty parameter provided a pattern of results which was compatible with those reported with the response time and the error rate, though it was mostly related to the latter probably since errors were salient events. The subjective difficulty and the number of errors were also related to the fun score, an effect that could be explained a priori by the classical inverted U shape between complexity and arousal (Yerkes & Dodson, 1908), keeping in mind that the difficulty rating and the number of errors varied from low to moderate and assuming that the stimulating experience was mainly due to the occurrence of incongruent trials. This was corroborated by observations made during the execution of the task, such as laughing at errors, crying out at errors (“Damn!”) and providing self-praise for a succession of fast and quick responses. However, the idea that participants enjoyed performing these tasks should not be over-interpreted, since near the end of the task (that lasted about 10 min), many participants requested to know how much longer they would be asked to play the game. In a Stroop study, Boutcher and Boutcher (2006) reported on overall ratings regarding the difficulty and enjoyment of tasks that lasted 5 mn (participants sat 2.4 m from a computer screen where 5 cm words were printed). Self-reported difficulties were similar for Stroop versions requiring either verbal responses (color naming) or not (participants had only to think of the color). Both ratings were greater than that obtained for a monochrome version with word naming, but the difference for enjoyment was greater only for the Stroop task with vocalization (about 3.5/7). This suggests that complexity alone is not fully responsible for enjoyment. Further research might profitably explore other possible influential factors (social components, materials, response modes, positions of the participant...).

Many Psychologists enthusiastically embrace the integration of different technologies for studying the relationships between sensorial perception, information processing, and behavior (Fairclough & Houston, 2004). The findings reported here suggest that interactive tabletops such as TangiSense may provide the means to investigate such relationships. One possibility that such tabletops could offer in the future is the recording of trajectory shifts. The use of movement trajectories as a measure of mental processes has become increasingly popular for investigating the links between response decisions and execution. For example, several researchers have tracked the x/y pixel coordinates of computer mouse movements or have used a movement tracking system (e.g. Buetti & Kerzel, 2009; Finkbeiner, Song, Nakayama, & Caramazza, 2008; Miles & Proctor, 2011; Spivey, Grosjean, & Knoblich, 2005) and have reported that congruent/incongruent task features influence curvatures of trajectories. Cognitive-motor coordination was not finely measured in the current study, but the experimenters, and the participants themselves, often observed that when the trial was incongruent, the tangible object was often moved initially towards the wrong target zone and then towards the correct zone. This inhibitory phenomenon likely occurred in the past when the Stroop task was performed with color-word stimuli printed on cards which were sorted into bins by colors or by words (e.g., Flowers & Blair, 1976) and in versions that required pointing responses (Blais & Besner, 2007), but likely to a lesser degree than what we observed using the interactive tabletop which allowed a greater span of movement. Future technical improvements that would enable the quantification of these effects by interactive tabletops are under study. Knowing how cognitive interference affects object trajectory may ultimately benefit clinical researchers or neurological practitioners who, for example, study cognitive-motor disorders such as Parkinson’s disease or frontal brain damaged patients (Obeso et al., 2011). It may also be useful for studying developmental disorders in children (Adams, & Jarrod, 2011) since the interactive tabletop monopolizes a natural behavior (moving an object into a matching zone in the same plane). Moreover, the tangible objects can be highly diversified in terms of semantic, color and shape features.

7. Conclusion

The Stroop and Reverse Stroop effect emerged when conflicting information streams were displayed by an application designed for the TangiSense interactive tabletop. Irrelevant and discordant perceptual and conceptual information at the input or the output steps increased response time, decreased accuracy and increase the subjective difficulty. Thus, care should be taken when designing applications for tangible computing technology if the goal is to maximize user performance and experience. Overall, these results complement and extend previous demonstrations that stimulus-response compatibility can influence human-computer interactions. With regards to the interference itself, the present study was mostly guided by a replication strategy (e.g., Francis, 2012). As protection against false positive effects, the Stroop literature, as several other extensive literatures, often exemplified replication of effects with different methods. In line with this, our results contribute to the heuristic value of the phenomenon.

ACKNOWLEDGMENTS

This research was supported in part by the “Ministère de l'Éducation Nationale, de la Recherche et de la Technologie”, the “Région Nord Pas-de-Calais”, the CNRS, the FEDER, CISIT, and especially the “Agence Nationale de la Recherche” (TTT and IMAGIT ANR projects 2010 CORD 01701). The authors would like to thank Yoann Lebrun, and Maxence Courbez for their contributions.

FOOTNOTES

1. As for the response mode, similarities with other HCI tasks can be exemplified by at least three prototypical applications: (1) during a learning task the user could classify a set of instructions/data with a certain order in several response zones according to the importance or characteristics of the instructions/data into the learning procedure (see for instance Kubicki, Wolff, Lepreux, & Kolski, 2015), (2) while filling a form, a user could fill in (or not) a response zone that could be highlighted by colors (ISO 9241-143, 2012), (3) games are also tasks that would require responding to different types of response zones, for instance in order to make a choice during the game (Pagulayan, Keefer, Wixon, Romero, & Fuller, 2003).
2. Like the computer version, the response locations were fixed. Instead of typing D for Blue, and F for yellow etc., participants moved left for Red and right for Blue etc.
3. Gender effects are only sporadically reported in the Stroop literature. In the present study, the results related to gender were unsystematic and did not negate the conclusions that are given for the other factors. They will not be presented for concision.
4. To sustain this point further, it was possible to perform an ANOVA on the response time (log) and difficulty ratings with the response factor replaced by a translation factor (WITHOUT and WITH). As expected a significant interaction arose between trial type and translation ($F(1, 72) = 8,04$; $p = .005$, $\eta_p^2 = .10$, and $F(1, 72) = 19,37$; $p = .00003$, $\eta_p^2 = .21$) showing that the translation increased interference, and the higher order interaction with task did not reach significance ($p = .397$, and $p = .636$).

References

- Adams, N. C., & Jarrold, C. (2002). Inhibition and the validity of the Stroop task for children with autism. *Journal of Autism Development Disorder*, 39(8), 1112–1121.
- Baudisch, P., Becker, T., & Rudeck, F. (2010). Lumino: tangible blocks for tabletop computers based on glass fiber bundles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)* (pp. 1165–1174). New York: ACM.
- Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology*, 90(6), 1054–1068.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Blais, C., & Besner, D. (2006). Reverse Stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1345–1353.
- Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review*, 14(3), 466–469.
- Boutcher, Y. N., & Boutcher, S. H. (2006). Cardiovascular response to Stroop: effect of verbal response and task difficulty. *Biological psychology*, 73(3), 235–241.
- Buetti, S., & Kerzel, D. (2009). Conflicts during response selection affect response programming: Reactions towards the source of stimulation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 816–834.
- Dunbar, K., & MacLeod, C. M. (1984). A horse race of a difference color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 622–639.
- Durgin, F. H. (2000). The Reverse Stroop Effect. *Psychonomic Bulletin & Review*, 7, 121–125.
- Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics*, 65, 367–378.
- Fairclough, S. H., & Houston, K. (2004). A metabolic measure of mental effort. *Biological Psychology*, 66(2), 177–190.
- Fernandez-Duque, D., & Knight, M. (2008). Cognitive control: Dynamic, sustained, and voluntary influences. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 340–355.
- Finkbeiner, M., Song, J., Nakayama, K., & Caramazza, A. (2008). Engaging the motor system with masked orthographic primes: A kinematic analysis. *Visual Cognition*, 16(1), 11–22.
- Finkenzeller, K. (2003). *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. New York: John Wiley & Sons.
- Fitzmaurice, G.W., Ishii, H., & Buxton, W. (1995). Bricks: Laying the Foundations for Graspable User Interfaces. In *Human Factors in Computing Systems, CHI '95 Conference Proceedings* (pp. 442–449). New York: ACM.
- Flowers, J. H. (1975) "Sensory" interference in a word-color matching task. *Perception & Psychophysics*, 18(11), 37–43.
- Flowers, J. H., & Blair, B. (1976). Verbal interference with visual classification: Optimal processing and experimental design. *Bulletin of the Psychonomic Society*, 7, 260–262.
- Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, 7(6), 585–594.
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42.
- Gendolla, G. H., Wright, R. A., & Richter, M. (2012). Effort Intensity: Some Insights From the Cardiovascular System. In R. M. Ryan (Ed.), *The Oxford handbook of Human motivation* (pp.420–438). Oxford: Oxford University Press.

- Hapeshi, K., & Jones, D. (1992). Interactive multimedia for instruction: A cognitive analysis of the role of audition and vision. *International Journal of Human-Computer Interaction*, 4(1), 79–100.
- Ishii, I., & Ullmer, B. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Human Factors in Computing Systems, CHI '97 Conference Proceedings* (pp. 234–241). New York: ACM.
- ISO 9241-143 (2012). ISO 9241-143:2012, Ergonomics of human-system interaction -- Part 143: Forms.
- Jetter, H.C, Zöllner, M., Gerken, J., & Reiterer, H. (2012). Design and Implementation of Post-WIMP Distributed User Interfaces with ZOIL. *International Journal of Human-Computer Interaction*, 28 (11), 737–747.
- Jordà, S., Geiger, G., Alonso, M., & Kaltenbrunner, M. (2007). The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proc. of the 1st international conference on Tangible and embedded interaction (TEI '07)* (pp. 139–146). New York: ACM.
- Kolski, C., Garbay, C., Lebrun, Y., Badeig, F., Lepreux, S., Mandiau, R., & Adam, E. (2014). Interactive Surfaces, Tangible Interaction: Perspectives for Risk Management. In P. Millot (Ed.), *Risk Management in Life critical Systems* (pp. 351-373). London: ISTE-Wiley.
- Kubicki S., Lepreux S., Lebrun Y., Dos Santos P., Kolski C., & Caelen J. (2009). New Human-Computer Interactions Using Tangible Objects: Application on a Digital Tabletop with RFID Technology. In J.A. Jacko (Ed.), *Human-Computer Interaction, 13th International Conference, HCI International 2009* (San Diego, CA, USA, July 19-24, 2009), *Proceedings, Part III, LNCS 5612* (pp. 446–455). Berlin/Heidelberg: Springer-Verlag.
- Kubicki, S., Lepreux, S., & Kolski, C. (2011). Evaluation of an interactive table with tangible objects: Application with children in a classroom. In *Proc. 2nd Workshop on Child Computer Interaction "UI Technologies and Educational Pedagogy"*, at CHI'2011, Vancouver, May.
- Kubicki, S., Lepreux, S., & Kolski, C. (2012). RFID-driven situation awareness on TangiSense, a table interacting with tangible objects. *Personal and Ubiquitous Computing*, 16 (8), 1079–1094.
- Kubicki, S., Lebrun, Y., Lepreux, S., Adam, E., Kolski, C., & Mandiau, R. (2013). Simulation in Contexts Involving an Interactive Table and Tangible Objects. *Simulation Modelling Practice and Theory*, 31, 116–131.
- Kubicki, S., Wolff, M., Lepreux, S., & Kolski C. (2015). RFID interactive tabletop application with tangible objects: exploratory study to observe young children' behaviors. *Personal and Ubiquitous Computing*, 19 (8), 1259-1274.
- Kukla, A. (1972). Foundations of an attributional theory of performance. *Psychological Review*, 79(6), 454–470.
- Lebrun, Y., Adam, E., Kubicki, S., & Mandiau, R. (2010). A Multi-agent System Approach for Interactive Table Using RFID. In Y. Demazeau, F. Dignum, J.M. Corchado & J.B. Perez (Eds.), *8th International Conference on Practical Applications of Agents and Multi-Agent Systems, PAAMS 2010* (pp. 125–134). Berlin/Heidelberg: Springer.
- Lebrun, Y., Adam, E., Mandiau, R., & Kolski, C. (2015). A model for managing interactions between tangible and virtual agents on an RFID interactive tabletop: case study in traffic simulation. *Journal of Computer and System Sciences*, 81, 585–598.
- Lepreux, S., Kubicki, S., Kolski, C., & Caelen, J. (2012). From Centralized interactive tabletops to Distributed surfaces: the Tangiget concept. *International Journal of Human-Computer Interaction*, 28 (11), 709–721.
- Linnman, C., Carlbring, P., Ahman, A., Andersson, H., & Andersson, G. (2006). The Stroop effect on the internet. *Computers in Human Behavior*, 22, 448–455.
- Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 978.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4(10), 383–391.

- Manches, A., O'Malley, C., & Benford, S. (2009). Physical manipulation: evaluating the potential for tangible designs. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, TEI'2009* (pp. 77–84). New York: ACM.
- Mangos, P. M., & Steele-Johnson, D. (2001). The role of subjective task complexity in goal orientation, self-efficacy, and performance relations. *Human performance*, 14(2), 169–185.
- Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, 10(4), 303–330.
- Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, 21, 627–645.
- Miles, J. D., & Proctor, R. W. (2011). Colour correspondence effects between controlled objects and targets. *The Quarterly Journal of Experimental Psychology*, 64(10), 2044–2064.
- Miller H., Kubicki S., Kolski C., & Naveteur J. (2013). Implementation of the Stroop task using an Interactive Table: An experimental study. In *ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '13* ACM, Paris, (pp. 253–258). New York: ACM.
- Obeso, I., Wilkinson, L., Casabona, E., Bringas, M. L., Álvarez, M., Álvarez, L., Pavón, N., Rodríguez-Oroz, M. C., Macías, R., Obeso, J. A., & Jahanshani, M. (2011). Deficits in inhibitory control and conflict resolution on cognitive and motor tasks in Parkinson's disease. *Experimental Brain Research*, 212(3), 371–384.
- Pagulayan, R. J., Keefer, K., Wixon, D., Romero, R. L., & Fuller, T. (2003). User-Centered Design In Games. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook. Fundamentals, Evolving, Technologies and Emerging Applications* (pp. 883-906). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Ruff C. C., Woodward T. S., Laurens K. R., & Liddle P. F. (2001). The role of the anterior cingulate cortex in conflict processing: evidence from reverse stroop interference. *Neuroimage*, 14(5), 1150–1158.
- Sharma D., & McKenna F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory and Cognition*, 26(2), 1033–1040.
- Sichel, J. L., & Chandler, K. A. (1969). The color-word interference test: The effects of varied color-word combinations upon verbal response latency. *Journal of Psychology*, 72, 219–231.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398.
- Song, Y., & Hakoda, Y. (2015). An fMRI study of the functional mechanisms of Stroop/reverse-Stroop effects. *Behavioural Brain Research*, 290, 187-196.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Sugg, M. J., & McDonald, J. E. (1994). Time course inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception & Performance*, 20, 647–675.
- 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 (6), 987–999.
- Washburn, D. A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers*, 35(2), 185-193.
- Wylie, G., & Allport, A. (2000). Task switching and the measurement of “switch costs.” *Psychological Research*, 63, 212–233.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.

Author Biographies

Holly C. Miller obtained her doctorate in Experimental Psychology at the University of Kentucky. She is now a post-doctoral researcher who examines the biological underpinnings of cognitive behavior, and how unconscious processes influence human responding.

Sébastien Kubicki obtained his Ph.D. in 2011. He is an assistant professor of Computer Science at the Brest National School of Engineering (France) and a member of the IHSEV team of the Lab-STICC laboratory. He is specialized in human-computer interaction, context awareness, systems and new user interfaces like interactive tabletops.

Delphine Caffier is a doctoral student in Engineering and Ergonomics of Physical Activity at the University of Valenciennes (France). Her main research interest concerns the links between emotions and movement.

Christophe Kolski obtained his Ph.D. in 1989. He is a professor of Computer Science at the University of Valenciennes (France). He specializes in human-computer interaction, software engineering for interactive systems, adaptive User Interface, tangible and distributed interaction.

Janick Naveteur is an associate professor in the Biology department of the University of Lille 1; with a research position at the University of Valenciennes. Her focus is on emotions, attention and pain, with a special interest in time pressure and transportation. Methodologies include psychophysiological, and self-reports.