Subjective aspects of cognitive control at different stages of processing

EZEQUIEL MORSELLA
San Francisco State University, San Francisco, California
and University of California, San Francisco, California

LILIAN E. WILSON
Yale University, New Haven, Connecticut

CHRISTOPHER C. BERGER
San Francisco State University, San Francisco, California

MIKAELA HONHONGVA
Yale University, New Haven, Connecticut

ADAM GAZZALEY
University of California, San Francisco, California

AND

JOHN A. BARGH
Yale University, New Haven, Connecticut

Although research on cognitive control has addressed the effects that different forms of cognitive interference have on behavior and the activities of certain brain regions, until recently, the effects of interference on subjective experience have not been addressed. We demonstrate that, at the level of the individual trial, participants can reliably introspect the subjective aspects (e.g., perceptions of difficulty, competition, and control) of responding in interference paradigms. Similar subjective effects were obtained for both expressed and unexpressed (subvocalized) actions. Few participants discerned the source of these effects. These basic findings illuminate aspects of cognitive control and cognitive effort. In addition, these data have implications for the study of response interference in affect and self-control, and they begin to address theories regarding the function of consciousness.

More than 30 years ago, the classic flanker task first appeared in this journal (then called *Perception & Psychophysics*; B. A. Eriksen & Eriksen, 1974). Since that time, through empirical and theoretical developments (Brown & Braver, 2005; van Veen & Carter, 2006), the impact that cognitive interference of various kinds has on behavior and the activity of certain brain regions has been examined (Botvinick, 2007; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Brown & Braver, 2005; Curtis & D'Esposito, 2009; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; van Veen & Carter, 2006), but for the most part, the effects of interference on subjective experience have not been addressed[1] (cf. Mayr, 2004; Mayr, Awh, & Laurey, 2003; Mulert, Menzinger, Leicht, Pogarell, & Hegerl, 2005; Rosen, McGuire, & Botvinick, 2007). Today, the relationship between the objective and subjective indices of cognitive control/effort remains mysterious (Baumeister & Vohs, 2004; Grahek, 2007; Preston

& Wegner, 2009). Thus, the fleeting urges, inclinations, and tendencies that thousands of participants have experienced when performing interference tasks such as the flanker task (C. W. Eriksen & Schultz, 1979) or the Stroop task[2] (Stroop, 1935) continue to be a scientific terra incognita.[3] Are these subjective effects ephemeral and capricious or systematic and reliable?

Consider the classic flanker task (C. W. Eriksen & Schultz, 1979). In this task, participants are first trained to press one button with one finger when presented with the letter S or M and to press another button with another finger when presented with the letter P. After training, the participants are instructed to respond to targets that are flanked by distractors. For example, they are instructed to respond to the stimulus presented in the center of an array (e.g., SS*P*SS, SS*M*SS, SS*S*SS, targets underscored) and to disregard the flanking stimuli (the distractors). It is well established that interference—as indexed by er-

E. Morsella, morsella@sfsu.edu

rors and response times (RTs)—depends on the nature of the distractors: Longer RTs are found when the distractors are associated with a response that is different from that associated with the target (*response interference* [RI]; e.g., SS**P**SS) than when the distractors are different in appearance but associated with the same response (*stimulus interference* [SI]; e.g., SS**M**SS). Of the three examples above, the shortest RTs are found when the distractors are identical to the target (e.g., SS$\underline{S}$SS; C. W. Eriksen & Schultz, 1979; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001).

It is obvious to the participant and to the experimenter alike that notable changes in subjective experience accompany each response on every trial, rendering the task qualitatively different from that of everyday perception–action mapping (e.g., flicking a light switch). Yet, less has been documented about the subjective aspects of such interference tasks than about their behavioral and neural aspects.[4]

If these subjective effects are systematic and reliable, which kinds of cognitive interference lead to the strongest modulations in subjective experience, and which lead to little or no modulation? In agreement with recent views (Gazzaley & D'Esposito, 2007; van Veen & Carter, 2006), we believe that answering such questions is essential for understanding the dynamics of more hot (Metcalfe & Mischel, 1999) kinds of conflict, such as those involving self-control and motivation (e.g., approach–approach conflicts; Livnat & Pippenger, 2006; Miller, 1959). In short, the cognitive dynamics underlying the subjective effects found in interference paradigms may reveal principles about the fundamental tuggings and pullings and ups and downs of the human emotional experience (Morsella, 2005). In addition, we believe that subjective data can illuminate aspects of cognitive processing that may not be revealed in traditional dependent measures, such as RTs and error rates.

Hence, in a series of experiments, we examined the subjective effects from different kinds of cognitive interference in variants of the Stroop and flanker tasks.

In addition to documenting the trial-by-trial subjective effects from the Stroop and flanker tasks (an important corpus of data in its own right), we took the opportunity to examine the hypothesis (Bargh & Morsella, 2008; Vygotsky, 1962) that internalized actions such as subvocalizations should feature the same subjective dynamics as externalized actions. If so, this would rule out the hypothesis that the subjective effects of these tasks stem only from proprioceptively detected conflict at the level of effectors (McGuigan, 1966; see also Pickering & Garrod, in press). Considering that RI from flankers can lead to subthreshold muscular activations (e.g., Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; cf. Morsella & Krauss, 2005), there is the possibility that subjective effects may be constituted in part by proprioception of activation in effector systems (cf. Wegner, 2002). Showing that covert and overt actions are similar with respect to subjective effects would illuminate a basic aspect of human action (itself an underexplored area of investigation; Morsella, 2009; Rosenbaum, 2005).

Overview of the Experiments

We designed a series of experiments to test the claim that subjective effects are systematic, measurable, reliable, and arise from cognitive interference in a principled fashion. Specifically, using vocal and subvocal versions of the classic Stroop task, in Experiments 1–3, we examined whether participants could reliably introspect on the subjective aspects (e.g., difficulty, competition, and control) of their responses at the level of the individual trial and whether similar effects are obtained for externalized and internalized actions.

EXPERIMENT 1
Are Subjective Urges Systematic?

Our primary goal in Experiment 1 was to establish that participants could reliably introspect the urge to err on each trial of a motionless, subvocal version of the Stroop task. We predicted that the participants would report the strongest urges to err for incongruent conditions and weaker urges to err for congruent conditions. Because the task involves an internalized form of action, finding this pattern of results would cast doubt on the hypothesis that the subjective effects arise from conflict occurring at the level of motor effectors (cf. Coles et al., 1985; McGuigan, 1966).

Method

Participants
Yale University students ($N = 15$) participated for class credit or $8.

Procedure
The participants were run individually. The session consisted of a block of trials in which the participants responded to Stroop stimuli subvocally. Each block consisted of 40 Stroop trials having 8 congruent (e.g., RED written in red), 16 incongruent (e.g., RED in blue), 8 control (e.g., HOUSE in green), and 8 neutral (e.g., XXXX in pink) stimuli, presented in random order. The eight colors used were correctly identified by all of the participants. In the incongruent condition, targets (colors) and distractors (words) were paired systematically (e.g., if RED was written in blue, BLUE was written in red). The participants were instructed as follows:

> In this task, you must respond to the words presented on the screen by naming the colors in which the words are written as fast and as accurately as you can, but you must name the colors only "in your head" and not aloud. Speaking in your mind but not aloud is called *subvocalizing*. For example, if the word FLOWER is presented in blue, you must think to yourself the color name "blue." As soon as you are done thinking the color name, you must press the space bar with your dominant hand. Pressing the space bar will allow you to proceed to the next trial.

It is important to emphasize that we never considered the RTs associated with pressing the space bar to be an informative or valid dependent measure. The participants performed this action only to indicate that they had finished responding subvocally, were paying attention to the task, and were ready to commence the next trial. In short, the motor aspect of the task in this and the subsequent subvocal experiments was not designed to render an accurate measure of the latency of cognitive processing or the onset of subvocalization. Rather, for subvocal tasks, our focus was on the nature of the reported subjective effects.

Each trial proceeded as follows. A ready prompt (question mark) appeared on-screen until the participants indicated that they were

ready to proceed by pressing the space bar. Thereafter, a fixation point (+) was shown at the center of the screen for 1,500 msec. It was followed by a blank screen (700 msec), after which a randomly selected Stroop stimulus appeared (in 48-point Helvetica) and remained on-screen until the space bar was depressed. After the response and 700 msec, the participants were asked, "How strong was the urge to make a mistake?" They responded on an 8-point scale, in which 1 signified *almost no urge*, and 8 signified *extremely strong urge*. Thereafter, the next trial began after 500 msec. For this and the following experiments, the stimuli were always presented in random order on a white background on a 43-cm Apple eMac computer monitor with a viewing distance of approximately 48 cm, and stimulus presentation was controlled by PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Results

Subjective Effects

As is illustrated in Figure 1, Stroop condition produced systematic effects on the urge to err [$F(3,42) = 33.679$, $p < .0001$, $\eta_p^2 = .71$]; urges were strongest for the incongruent condition ($M = 4.01$, $SEM = 0.39$), followed by the control ($M = 2.38$, $SEM = 0.30$), neutral ($M = 1.59$, $SEM = 0.16$), and congruent ($M = 1.34$, $SEM = 0.14$) conditions. Fisher's post hoc least significant difference (PLSD) revealed that all differences among the conditions were significant ($ps < .01$), except for that between the neutral and congruent conditions ($p = .497$). Omitted responses and typing errors resulted in the loss of 1 (0.2%) of 600 ratings.

Supplementary Analysis

Again, for this and the subsequent subvocal experiments, the RTs associated with pressing the space bar were never considered to be a dependent measure of interest, for various reasons. Nevertheless, we did take the opportunity to analyze RTs associated with the space bar responses. Following Woodworth and Schlosberg (1954), RTs below 200 msec and above 2 sec were excluded from analysis, resulting in the loss of 78 (13%) out of 600 trials. We repeated this trimming procedure for all subsequent RT mea-

sures. See Appendix A for the mean proportion of trials removed per condition. Condition influenced the proportion of trimmed scores [$F(3,42) = 4.61$, $p < .01$, $\eta_p^2 = .24$]. However, Fisher's PLSD revealed that the only significant difference in these proportions was between the congruent and incongruent conditions ($p = .04$). That the participants were in fact responding to the stimuli subvocally was, to some extent, corroborated by a trend mirroring the Stroop effect that was manifested in the RTs associated with pressing the space bar; RTs were the longest for the incongruent condition ($M = 1,059.04$, $SEM = 86.27$), followed by the neutral ($M = 949.15$, $SEM = 92.52$), control ($M = 930.18$, $SEM = 85.20$), and congruent ($M = 905.29$, $SEM = 83.85$) conditions [$F(3,39) = 4.068$, $p = .0132$, $\eta_p^2 = .24$]. However, Fisher's PLSD revealed that there were no significant differences among conditions ($ps > .05$).

Only for 3 of the 15 participants were there significant ($ps < .05$) correlations between RTs and the urge to err (mean $r = .16$, Fisher's r to z, $p > .05$), suggesting that the participants did not make their judgments solely on the basis of observations of their own RTs. For this and all subsequent analyses, the mean correlation was calculated by including ratings and RTs from all conditions. The small number of trials (e.g., $n = 8$) comprising each condition was inadequate for drawing strong conclusions regarding the relationship between RTs and the subjective measures in each condition.

Discussion

As was predicted, we found that the pattern of reported urges to err mirrored that of RT effects in the standard Stroop task. Because the task was subvocal, this finding casts doubt on the notion that subjective effects arise from interference at the level of the effector system. It seems that the urges result from a process that is more centralized and that features the subjective dynamics of externalized actions (Vygotsky, 1962).

EXPERIMENT 2
The Urge to Read During Vocal and Subvocal Stroop Tasks

On the basis of Cohen, Dunbar, and McClelland (1990), we believe that the urges to err reported for the incongruent Stroop condition in Experiment 1 stem in part from the conflict arising between the dominant word-naming and weaker color-naming plans, although one, of course, cannot rule out that the urges could simply reflect difficulty, effort, arousal, or a combination of the three. MacLeod and MacDonald (2000) suggested that, in the congruent condition, participants often read the stimulus word inadvertently. With respect to this condition, in which word reading and color naming lead to the same response, MacLeod and MacDonald stated, "The experimenter (perhaps the participant as well) cannot discriminate which dimension gave rise to the response on a given congruent trial" (p. 386). MacLeod and MacDonald then reviewed substantial evidence from diverse sources that
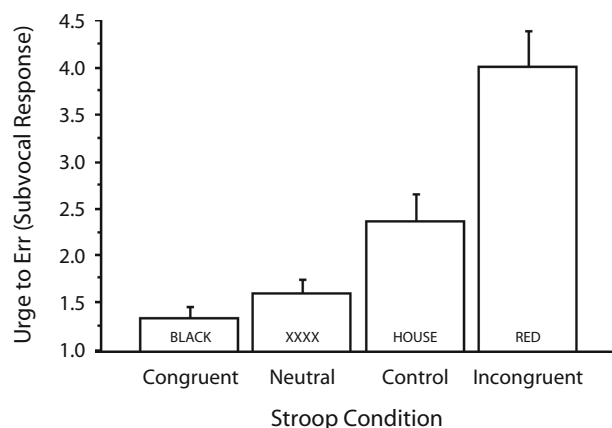

Figure 1. Mean urges to err on a subvocal version of the Stroop task (Experiment 1). Error bars indicate one standard error of the mean.

participants often do read the stimulus word inadvertently in the congruent condition. However, it remains unknown whether the participants are aware of this phenomenon.

Method

Participants

Yale University students ($N = 112$) participated for class credit or $8.

Procedure

The participants were run individually. The session consisted of two blocks of Stroop trials. In one block, the participants responded to Stroop stimuli aloud (vocal condition), and in the other they responded subvocally (subvocal condition). Block order was randomized across participants to minimize order effects (54 participants began with the vocal condition, and 58 began with the subvocal condition). Each block consisted of 40 Stroop trials having 8 congruent, 16 incongruent, 8 control, and 8 neutral stimuli presented in random order. For the vocal condition, the participants were instructed as follows:

> In this task, you must respond to the words presented on the screen by naming aloud as fast as you can the colors in which the words are written. For example, if the word FLOWER is written in blue, you must say "blue" as fast as you can. The microphone will record your response and measure your response time.

Vocal responses were detected by a microphone (Model 33-3014; Radio Shack, Fort Worth, TX) connected to a PsyScope buttonbox (Model 2.02; New Micros, Dallas, TX). For the subvocal condition, the participants received the instructions from Experiment 1. Each trial proceeded exactly as did those in Experiment 1, except that the

participants were now asked, "How strong was the urge to read the word?" They responded on an 8-point scale, in which 1 signified *almost no urge* and 8 signified *extremely strong urge*.

Results

Vocal Condition

Subjective effects. As is illustrated in Figure 2, Stroop condition produced effects on the urge to read [$F(3,333) = 180.497, p < .0001, \eta_p^2 = .62$]; urges were strongest for the incongruent condition ($M = 5.87, SEM = 0.11$), followed by the congruent ($M = 5.20, SEM = 0.20$), control ($M = 4.46, SEM = 0.14$), and neutral ($M = 2.18, SEM = 0.13$) conditions (Fisher's PLSD, $ps < .01$). Omitted responses and typing errors resulted in the loss of 166 (3.7%) of 4,480 ratings.

RTs. RT trimming resulted in the loss of 157 (3.5%) out of 4,480 vocal trials. See Appendix A for the mean proportion of trials removed per condition. Condition did not influence the proportion of trimmed scores [$F(3,333) = 2.85, p > .05$]. We replicated the Stroop RT effect; RTs were longest for the incongruent condition ($M = 859.00, SEM = 13.98$), followed by the control ($M = 808.24, SEM = 13.82$), neutral ($M = 719.67, SEM = 11.85$), and congruent ($M = 698.58, SEM = 11.66$) conditions [$F(3,333) = 134.085, p < .0001, \eta_p^2 = .55$]. Fisher's PLSD revealed that all differences among conditions were significant ($ps < .01$), except for that between the neutral and congruent conditions ($p = .25$).
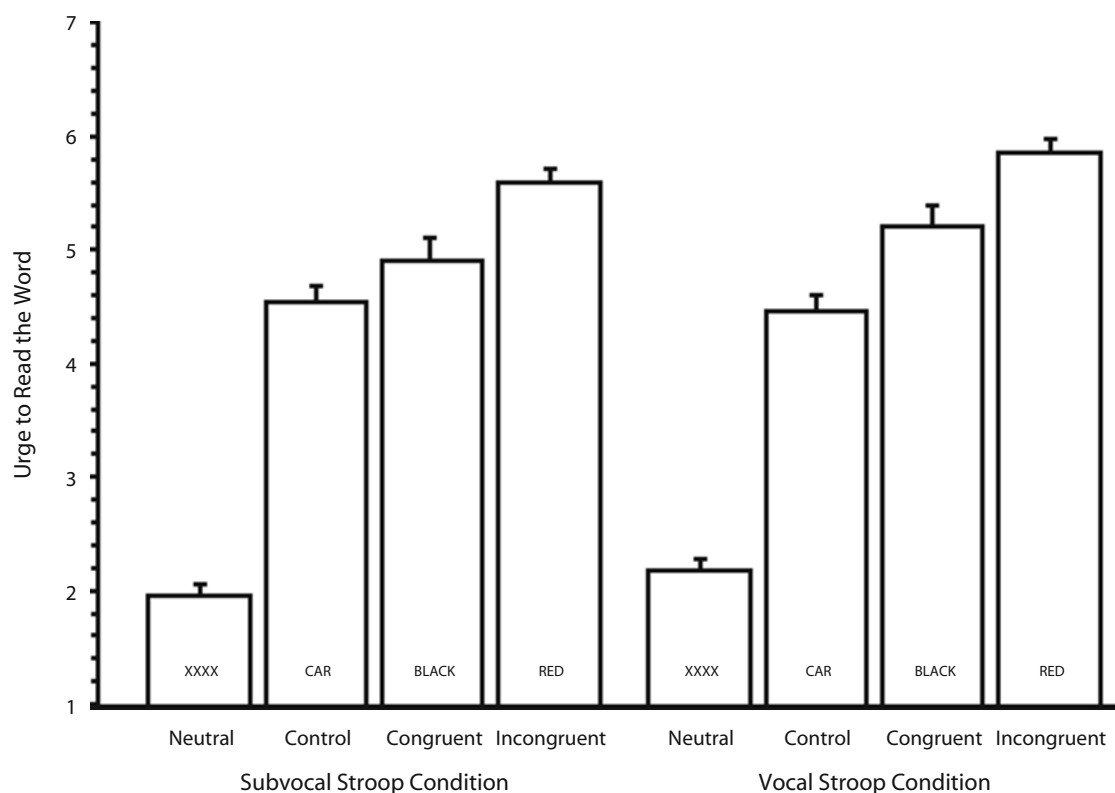

Figure 2. Mean urges to read stimulus words in subvocal and vocal versions of the Stroop task (Experiment 2). Error bars indicate one standard error of the mean.

Only for 6 of the 112 participants were there significant ($ps < .05$) correlations between RTs and the urge to read (mean $r = .01$, Fisher's r to z, $p > .05$), suggesting that the participants did not make their judgments solely on the basis of observations of their own RTs.

Subvocal Condition

Subjective effects. Stroop condition produced clear effects on the urge to read [$F(3,333) = 189.246$, $p < .0001$, $\eta_p^2 = .63$]; urges were strongest for the incongruent condition ($M = 5.60$, $SEM = 0.13$), followed by the congruent ($M = 4.92$, $SEM = 0.21$), control ($M = 4.54$, $SEM = 0.16$), and neutral ($M = 1.95$, $SEM = 0.12$) conditions (see Figure 2). Fisher's PLSD revealed that all of the differences among the means were significant ($ps < .01$), except for that between the control and congruent conditions ($p = .25$). Omitted responses and typing errors resulted in the loss of 267 (6.0%) of 4,480 ratings.

Supplementary analysis. As one would expect, the Stroop RT effect was again not clearly manifested in the RTs associated with pressing the space bar. The mean RTs for pressing the space bar were comparable in the incongruent ($M = 907.62$, $SEM = 34.83$) and control ($M = 907.63$, $SEM = 36.54$) conditions, but were shorter for the congruent ($M = 821.43$, $SEM = 31.50$) and neutral ($M = 805.78$, $SEM = 31.92$) conditions [$F(3,324) = 16.499$, $p < .0001$, $\eta_p^2 = .13$]. Fisher's PLSD revealed that the only significant differences among means were that between the control and neutral conditions and that between the incongruent and neutral conditions ($ps < .05$). Our trimming procedures resulted in the loss of 659 (14.7%) out of 4,480 subvocal trials. See Appendix A for the mean proportion of trials removed per condition. Condition influenced the proportion of trimmed scores [$F(3,99) = 3.194$, $p < .05$, $\eta_p^2 = .25$]. Fisher's PLSD revealed significant differences in these proportions between the incongruent condition and each of the other conditions ($ps < .05$). Only 16 of the 112 participants had significant ($ps < .05$) correlations between RTs and the urge to read (mean $r = .07$, Fisher's r to z, $p > .05$), suggesting, again, that the participants did not make their judgments solely on the basis of observations of their motor RTs.

General analysis. An omnibus within-subjects ANOVA revealed that the vocal and subvocal conditions led to similar patterns of results with respect to the urge to read the stimuli (Figure 2). The urge to read was judged to be stronger in the vocal than in the subvocal condition [$F(1,111) = 4.562$, $p = .0349$, $\eta_p^2 = .04$], but Fisher's PLSD revealed that this difference was unreliable ($p = .1130$). There was a significant main effect of Stroop condition [$F(3,333) = 217.863$, $p < .0001$, $\eta_p^2 = .66$; Fisher's PLSD, $ps < .001$]. There was also a subtle interaction between Stroop condition and task (vocal vs. subvocal) [$F(3,333) = 3.359$, $p = .0191$, $\eta_p^2 = .03$], an unpredicted interaction that will require further investigation.

Discussion

We present, for the first time, the participants' urges to read in the Stroop task. The participants reported the strongest urges to read the stimuli when responding in the incongruent condition and reported less of an urge to read when responding in the congruent condition, although it is known that reading tendencies may actually be equal in the two conditions (MacLeod & MacDonald, 2000). This may support the interpretation that the participants underestimated their urge to read words in the congruent condition.[5] This reduction in the urge to read in the congruent condition may be an instance of *double blindness*—the notion that one has diminished awareness that two distinct cognitive processes have taken place when those processes lead to identical action plans (Morsella, 2005; Morsella, Gray, Krieger, & Bargh, 2009). Of course, this experiment cannot rule out the alternative hypothesis that color naming actually suppresses the cognitive processes involved in word reading, but this hypothesis is inconsistent with the findings of MacLeod and MacDonald.

EXPERIMENT 3
Subjective Dimensions
of Cognitive Interference

To further examine the subjective aspects of cognitive interference, in Experiment 3, we had participants perform the same vocal and subvocal Stroop tasks, but they now answered additional questions following each trial. Comprising our urge to err measure, the first question was, "How strong was the urge to make a mistake?" Comprising our perception of control measure, the second question was, "How much personal control did you feel when saying the name of the color?" (for the vocal condition) or "How much personal control did you feel when thinking the name of the color?" (for the subvocal condition). The last question was, "How strong was the thought of a competing response (e.g., the thought of the word name)?" This question served as our measure of perception of competition.

We predicted that, for both the vocal and the subvocal versions of the task, when RI is low or absent during the Stroop task (as in the neutral and congruent conditions), urges to err and perceptions of competition will tend to be low, whereas perceptions of control will tend to be high; when RI is high (as in the incongruent condition), urges to err and perceptions of competition will tend to be high, whereas perceptions of control will tend to be low. Such a finding would illuminate the nature of the subjective effects associated with interference tasks and would also suggest that the participants can introspect not just about task difficulty, but about a richer set of subjective dimensions. Of course, such a finding cannot rule out the possibility that the judgments simply reflect overall difficulty or folk knowledge regarding how one should make judgments while experiencing cognitive conflict/effort in psychological experiments (see the General Discussion section).

Method

Participants

Yale University students ($N = 35$) participated for class credit or $8.

Procedure

The procedures were identical to those of Experiment 2, except that the participants were told that they would have to answer three questions following each Stroop trial. Each of the three questions was separated from the next by a span of 700 msec following the participant's response.

Results

All the data from one session were excluded from the analysis, because the participant failed to follow instructions. Of the remaining data, omitted responses and typing errors resulted in the loss of 414 (5.1%) of 8,160 ratings. With task (vocal vs. subvocal) and Stroop condition as within-subjects factors, we conducted omnibus ANOVAs for each of the three dependent measures (urge to err, control, and competition). There was no main effect of task ($ps > .05$), no interaction between task and Stroop condition ($ps > .05$), and a main effect only of Stroop condition ($ps < .0001$), the nature of which is elaborated below for each task.

Vocal Condition

RTs. Trimming resulted in the loss of 44 (3.2%) of 1,360 RT data points from the vocal task. See Appendix A for the mean proportion of trials removed per condition.

Condition influenced the proportion of trimmed scores [$F(3,99) = 3.194$, $p < .05$, $\eta_p^2 = .08$]. Fisher's PLSD revealed a significant difference in these proportions between the neutral and incongruent conditions ($p < .05$). We replicated the Stroop effect; RTs were longest for the incongruent condition ($M = 895.02$, $SEM = 22.34$), followed by the control ($M = 805.21$, $SEM = 23.43$), neutral ($M = 719.15$, $SEM = 18.84$), and congruent ($M = 716.39$, $SEM = 23.41$) conditions [$F(3,99) = 45.792$, $p < .0001$, $\eta_p^2 = .58$]. As in Experiment 2, Fisher's PLSD revealed that all of the differences among conditions were significant ($ps < .01$), except for that between the congruent and neutral conditions ($p = .93$).

Urge to err. As is illustrated in Figure 3, Stroop condition produced systematic effects on the urge to err [$F(3,99) = 58.453$, $p < .0001$, $\eta_p^2 = .64$]; urges were greatest for the incongruent condition ($M = 4.41$, $SEM = 0.28$), followed by the control ($M = 3.21$, $SEM = 0.30$), neutral ($M = 2.15$, $SEM = 0.25$), and congruent ($M = 2.04$, $SEM = 0.22$) conditions. Fisher's PLSD revealed that all of the differences among conditions were significant ($ps < .01$), except for that between the neutral and congruent conditions ($p = .78$). Twenty-seven of the 34 participants had significant ($ps < .05$) positive correlations between RTs and the urge to err (mean $r = .53$, Fish-
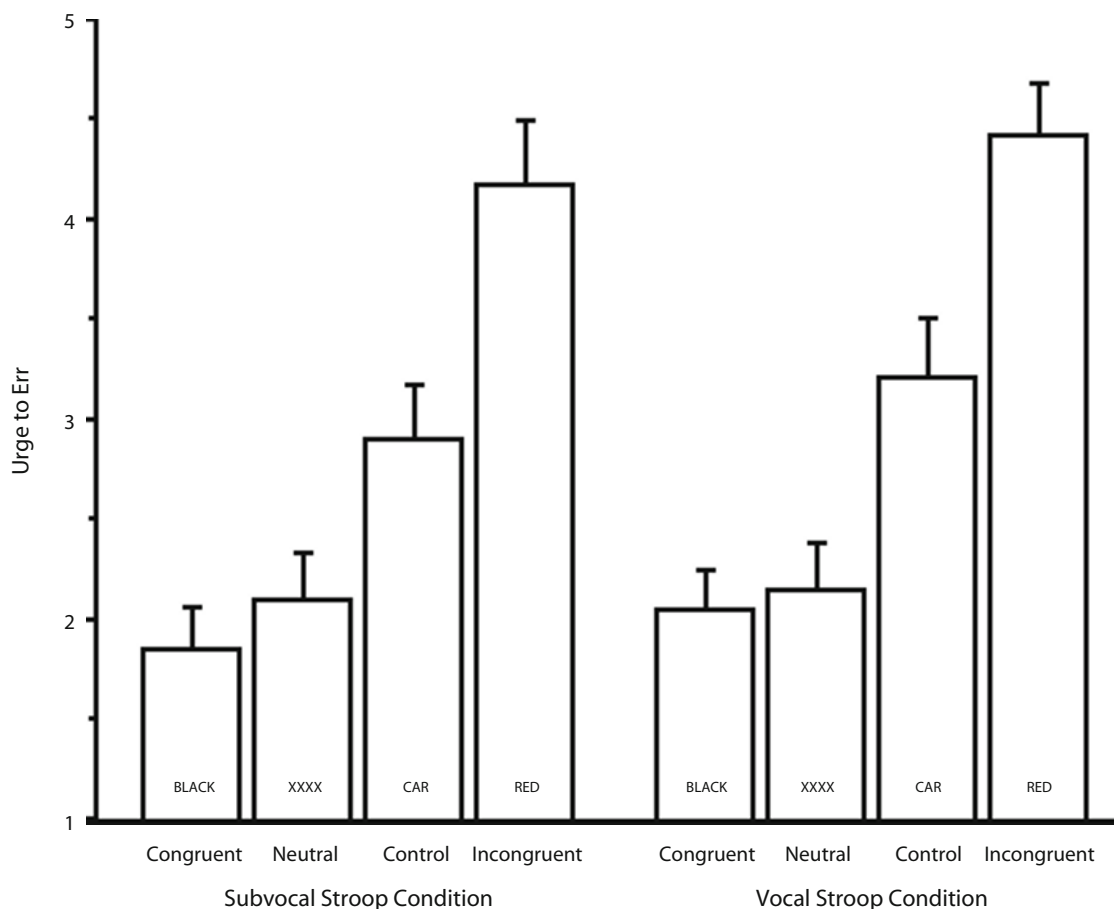

Figure 3. Mean urges to err in subvocal and vocal versions of the Stroop task (Experiment 3). Error bars indicate one standard error of the mean.

er's r to z, $p < .05$), suggesting that the participants may have made their judgments on the basis of observations of their own vocal RTs.

Perceptions of control. Stroop condition produced effects on the reported perception of control [$F(3,99) = 14.353$, $p < .0001$, $\eta_p^2 = .30$]; perception of control was greatest for the neutral condition ($M = 6.52$, $SEM = 0.23$), followed by the congruent ($M = 6.37$, $SEM = 0.24$), control ($M = 5.68$, $SEM = 0.24$), and incongruent ($M = 5.19$, $SEM = 0.26$) conditions (see Figure 4). Fisher's PLSD revealed that all of the differences among conditions were significant (ps $< .05$), except for the difference between the neutral and congruent conditions ($p = .68$) and that between the incongruent and control conditions ($p = .16$). Twenty-four of the 34 participants had significant (ps $< .05$) negative correlations between RTs and perceptions of control (mean $r = -.34$, Fisher's r to z, $p < .05$), suggesting that the participants may have made their judgments on the basis of observing their own RTs.

Perceptions of competition. Stroop condition produced effects on the perception of a competing response [$F(3,96) = 40.078$, $p < .0001$, $\eta_p^2 = .56$]; competition was greatest for the incongruent condition ($M = 5.08$, $SEM = 0.27$), followed by the control ($M = 4.06$, $SEM = 0.33$), congruent ($M = 2.46$, $SEM = 0.32$), and neutral ($M = 2.33$, $SEM = 0.30$) conditions (see Figure 5). Fisher's PLSD revealed that all of the differences among condi-

tions were significant (ps $< .01$), except for that between the neutral and congruent conditions ($p = .77$). Twenty-six of the 34 participants had significant (ps $< .05$) positive correlations between RTs and perceptions of competition (mean $r = .45$, Fisher's r to z, $p < .05$), suggesting that the participants may have made their judgments on the basis of observations of their own RTs.

Subvocal Condition

Urge to err. As in Experiment 1, Stroop condition did produce clear effects on the urge to err [$F(3,99) = 38.214$, $p < .0001$, $\eta_p^2 = .54$]; urges were strongest for the incongruent condition ($M = 4.17$, $SEM = 0.33$), followed by the control ($M = 2.90$, $SEM = 0.29$), neutral ($M = 2.10$, $SEM = 0.25$), and congruent ($M = 1.85$, $SEM = 0.22$) conditions (see Figure 3). Fisher's PLSD revealed that all of the differences among conditions were significant (ps $< .05$), except for that between the neutral and congruent conditions ($p = .52$). Fourteen of the 34 participants had significant (ps $< .05$) correlations between RTs and the urge to err (mean $r = .29$, Fisher's r to z, $p = .07$).

Perceptions of control. Stroop condition produced effects on the reported perception of control [$F(3,99) = 17.896$, $p < .0001$, $\eta_p^2 = .35$]; perceptions of control were greatest for the neutral condition ($M = 6.46$, $SEM = 0.27$), followed by the congruent ($M = 6.42$, $SEM = 0.29$), con-
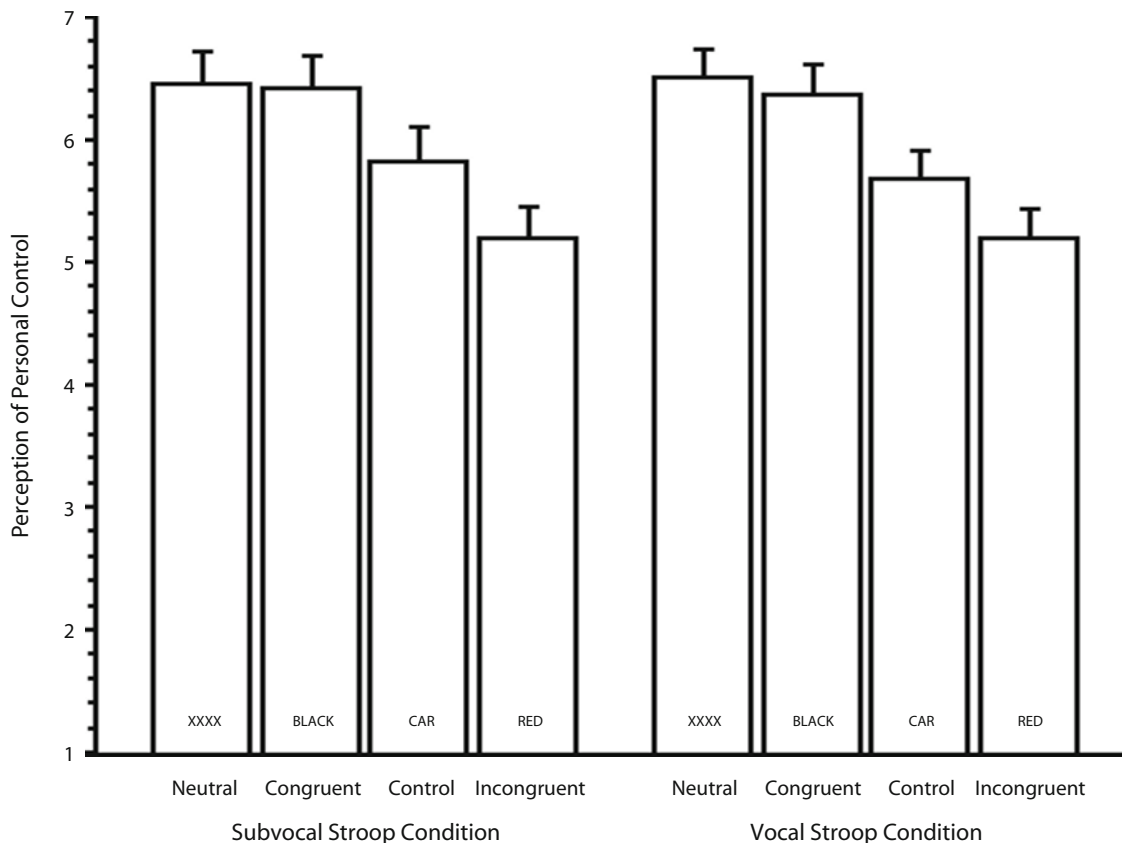

Figure 4. Mean perceptions of control in subvocal and vocal versions of the Stroop task (Experiment 3). Error bars indicate one standard error of the mean.
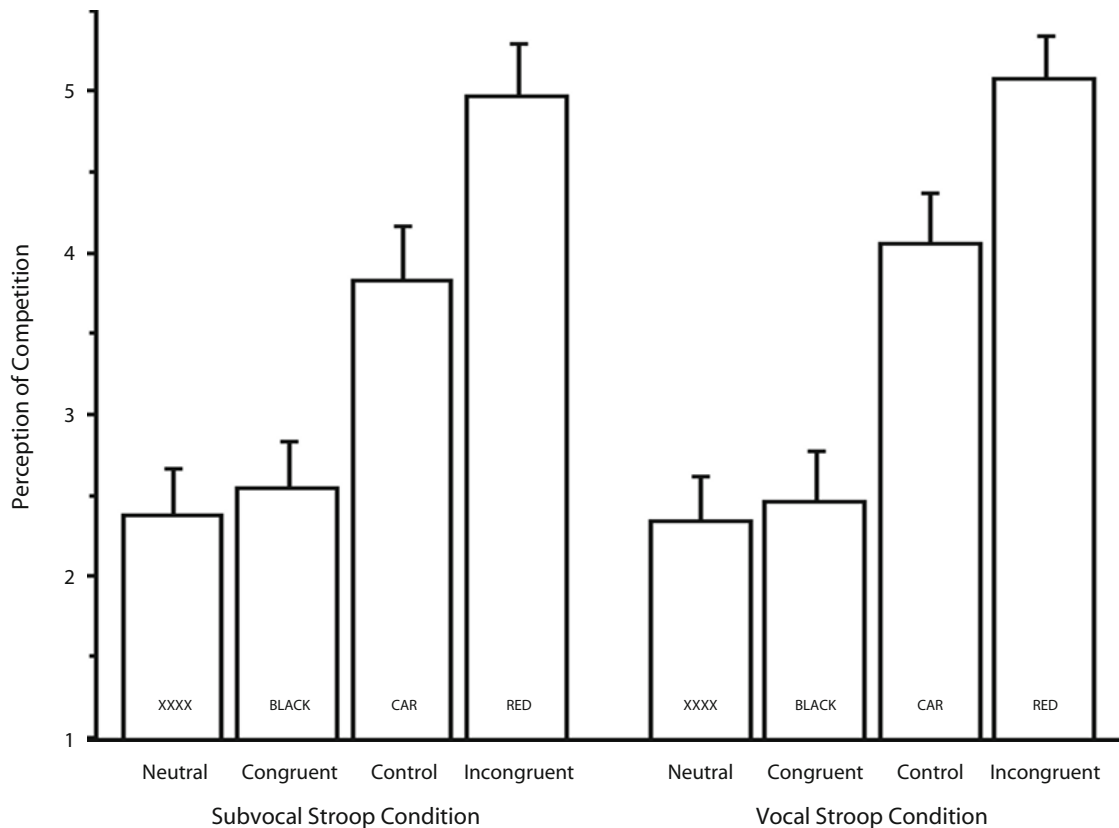
Figure 5. Mean perceptions of competition in subvocal and vocal versions of the Stroop task (Experiment 3). Error bars indicate one standard error of the mean.

trol ($M = 5.82$, $SEM = 0.30$), and incongruent ($M = 5.19$, $SEM = 0.28$) conditions (see Figure 4). Fisher's PLSD revealed that only the difference between the congruent and incongruent conditions and that between the neutral and incongruent conditions were significant ($ps < .05$). Twelve of the 34 participants had significant ($ps < .05$) negative correlations between RTs and perceptions of control (mean $r = -.25$, Fisher's r to z, $p > .05$).

Perceptions of competition. Stroop condition produced effects on the reported perception of a competing response [$F(3,96) = 37.267$, $p < .0001$, $\eta_p^2 = .54$]; perceptions of competition were greatest for the incongruent condition ($M = 4.97$, $SEM = 0.33$), followed by the control ($M = 3.83$, $SEM = 0.34$), congruent ($M = 2.55$, $SEM = 0.31$), and neutral ($M = 2.39$, $SEM = 0.29$) conditions (see Figure 5). Fisher's PLSD revealed that all of the differences among conditions were significant ($ps < .05$), except for that between the neutral and congruent conditions ($p = .73$). Only for 9 of the 34 participants were there significant ($ps < .05$) correlations between RTs and perceptions of a competing response (mean $r = .24$, Fisher's r to z, $p > .05$), suggesting that the participants did not make their judgments solely on the basis of observations of their own RTs.

Supplementary analysis. For the subvocal task, trimming resulted in the loss of 211 (15.5%) of 1,360 RT data points. See Appendix A for the mean proportion of trials removed per condition. Condition influenced the proportion of trimmed scores [$F(3,99) = 17.408$, $p < .05$, $\eta_p^2 = .34$]. Fisher's PLSD revealed significant differences in these proportions between the incongruent condition and each of the other conditions ($p < .05$). Regarding RTs, a trend mirroring the Stroop effect was manifested in the RTs associated with pressing the space bar; RTs were longest for the incongruent condition ($M = 920.84$, $SEM = 56.35$), followed by the control ($M = 860.65$, $SEM = 60.08$), neutral ($M = 801.88$, $SEM = 62.55$), and congruent ($M = 787.46$, $SEM = 54.50$) conditions [$F(3,96) = 7.144$, $p = .0002$, $\eta_p^2 = .18$]. However, Fisher's PLSD revealed no significant differences among conditions ($ps > .05$).

Discussion

The vocal and subvocal versions of the Stroop task again lead to similar patterns of subjective effects: When interference was low or absent during the Stroop task (as in the neutral and congruent conditions), urges to err and perceptions of competition tended to be low, whereas perceptions of control tended to be high; when interference was high (as in the incongruent condition), urges to err and perceptions of competition tended to be high, whereas perceptions of control tended to be low (see the General Discussion section for limitations of this approach).

EXPERIMENT 4A
Subjective Aspects of RI Versus SI

In Experiments 1–3, the perceived level of difficulty of the task alone could have guided the participants' self-reports, including those about perceptions of control and competition. To begin to illuminate this issue, in Experiment 4A, we examined the subjective aspects of the flanker task mentioned in the introduction, a task that has historically been used to tease apart the subjective effects of SI and RI (see van Veen et al., 2001). (The inherent limitations of such a dissociation are discussed in the General Discussion section.) As one would expect, urges to err are greater for flanker conditions involving RI than for conditions involving SI (Morsella, Rigby, & Gazzaley, 2009). Perhaps this explains why the participants of the present studies assigned stronger urges to the incongruent condition than to the congruent condition.

A Subjective Localizer Task

To have a better idea of whether the subjective effects from our experiments were driven in part by RI, in a novel paradigm we first had participants rate their urge to err while performing the Stroop task and then had them introspect the same *thing* while experiencing the conditions (SI and RI) of the flanker task. In this way, the Stroop task served as a sort of within-subjects localizer task for the subjective dimensions of interest (i.e., RI involving incompatible action plans). Following the Stroop task, the participants were told that, when estimating their urge to err, what they were "looking [for] inside their minds and measuring" was a psychological state known as *activity*, and that they would be asked to measure this kind of activity later in a different task. The rationale of the approach is that, if the subjective modulations associated with the Stroop task were driven to some extent by RI, the participants should report a greater degree of such modulations for flanker RI than for SI.

To not bias the participants, we defined the nondescript concept of *activity* only in terms of their experience. Thus, the participants learned to introspect not the general tendency to err on a task, but a specific urge or feeling that happens to be associated with increased urges to err in the Stroop task. No participant had difficulty understanding the concept of *activity*. It is important to note that we could just as well have called this dimension of interest something as arbitrary as *H5* or *Wundt Energy*, because the construct was defined only by the participant's own experience. We selected the term *activity* only because it is unbiased and intuitive.

Our primary motivation for introspection training was that we wanted to be as certain as possible that the participants were introspecting the same "thing" during both the flanker and Stroop tasks. Normally, this is difficult to establish, because identical ratings could emerge from the measurement of distinct subjective dimensions. For example, on an 8-point scale, the participants could judge a game of chess and the act of holding one's breath to be of comparable difficulty, even though this does not imply that the participants had measured the same dimension(s) in each task. To circumvent this problem, for the flanker task, we instructed the participants to home in on what they measured during the Stroop task. Independent support for the notion that the participants introspect the same dimension of interest for both tasks stems from a neuroimaging variant of this introspection training paradigm in which activity ratings were proportional to the degree of activation in brain regions that were common for two different interference tasks (Kang, Morsella, Shamosh, Bargh, & Gray, 2008).

Method

Participants
Yale University students ($N = 30$) participated for class credit or $8.

Procedure
The participants were run individually in two training phases and a test phase. Introspection training consisted of 24 Stroop trials having eight congruent (e.g., RED written in red), eight incongruent (e.g., RED in blue), and eight control (e.g., HOUSE in green) stimuli in random order. After the response and 700 msec, the participants were asked, "How strong was the urge to make a mistake?" They responded on an 8-point scale, in which 1 signified *almost no urge*, and 8 signified *extremely strong urge*. (We did not collect data for training in this experiment but did collect them for identical training sessions in Experiments 4B and 4C.) After introspection training, and following the exact procedures of van Veen et al. (2001), the participants were trained to press specified computer keys when presented with certain letter targets (48-point Helvetica): When presented with S or M, they pressed a key (occupying the "4" position on the number pad of the keyboard) with their right index finger; when presented with a P or H, they pressed the adjacent ("5") key with their right middle finger. To make target keys perceptually salient, the "4" and "5" keys were replaced with blackened keys, which were the only blackened keys on the white keyboard. Targets were presented in random order in the center of the screen and occupied less than 2 cm^2. For training (32 trials), the participants were told that accuracy was more important than speed.

After training, the participants were told that the remainder of the experiment (96 trials) would involve a similar task and that they should continue to respond to the shape in the center of the screen (the target), although they must now also disregard whatever stimuli appear peripherally (the distractors). The participants were now encouraged to respond as quickly and as accurately as possible and to avoid anticipations. Following van Veen et al. (2001), distractors horizontally flanked the target in the center of the screen. In the *identical* condition, targets were flanked by distractors that were identical to it (e.g., SSSSS or HHHHH; 48 trials, 12 replications for each letter). In the SI condition, targets and distractors were associated with the same response but were different letters (e.g., SSMSS or HHPHH; 24 trials, 6 replications of each possible combination). In the RI condition, targets and distractors were associated with different responses (e.g., SSPSS or MMHMM; 24 trials, 6 replications of each possible combination). Each trial began with a warning prompt (question mark) and beep, which preceded the randomly selected stimulus display by 1,300 msec. After responding, the participants rated their perceived level of activity on an 8-point scale, in which 1 signified *almost no urge* and 8 signified *extremely strong urge*.

Results

Errors and RTs
The data from two sessions were excluded from analysis because the participants failed to follow the instruc-

tions. Of the remaining data, incorrect responses to targets resulted in the loss of 125 (4.7%) out of 2,688 data points. See Appendix B for the mean proportion of errors per condition. As was expected, there was an effect of condition on error rate [$F(2,54) = 11.209$, $p <$ .0001, $\eta_p^2 = .30$]. Fisher's PLSD revealed significant differences between the error rates of the SI and RI conditions and between those of the SI and *identical* conditions ($ps < .05$). RT trimming resulted in the loss of 32 (1.2%) of 2,688 RT data points. Condition did not influence the proportion of trimmed scores [$F(2,54) = 0.912$, $p > .40$]. In this and subsequent analyses, errors were removed from the RT analysis. We replicated the findings of C. W. Eriksen and Schultz (1979) and van Veen et al. (2001). There was a main RT effect of condition [$F(2,54) = 23.749$, $p < .0001$, $\eta_p^2 = .47$], and the RI condition produced longer RTs ($M = 734.01$, $SEM = 31.78$) than did the SI ($M = 672.51$, $SEM = 33.52$) and *identical* ($M = 643.19$, $SEM = 29.29$) conditions. Planned comparisons revealed that all of these means were significantly different from each other ($ps_{paired} < .05$). The same pattern of results was obtained when we trimmed only RTs that were more than 2.5 *SD*s from each participant's mean score in each condition [$F(2,54) = 23.564$, $p < .0001$, $\eta_p^2 = .47$].

Subjective Activity

Typing errors and omissions resulted in the loss of 4 (0.1%) of 2,688 trials. As is clear in Figure 6, there were significant differences in subjective activity among the three conditions [$F(2,54) = 61.652$, $p = .0001$, $\eta_p^2 = .70$]. Planned comparisons revealed that more activity was found for the RI condition ($M = 3.17$, $SEM = 0.21$) than for the SI ($M = 2.21$, $SEM = 0.17$) and *identical* ($M = 1.85$, $SEM = 0.17$) conditions. All differences were significant ($ps_{paired} < .0001$). Twenty-six of the 28 participants had significant ($ps < .05$) positive correlations between RT and activity (mean $r = .50$, Fisher's r to z, $p < .05$), suggesting that the participants may have made their judgments on the basis of RTs.
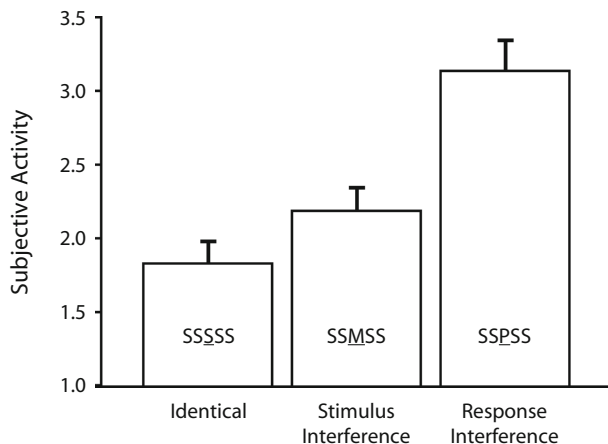

Figure 6. Mean subjective activity as a function of flanker condition (Experiment 4A).

Discussion

Note that, in this experiment, the size of the subjective effect was larger than that of the behavioral RT effect, which underscores the importance of using subjective data to illuminate aspects of cognitive processing that may not be revealed in standard dependent measures such as RT and error rates. We found that more subjective activity accompanied the RI condition than the SI and *identical* conditions. In our next experiment, we replicated this effect with a different class of stimuli and included additional conditions, which allowed for more precise appreciation of the influence of different kinds of interference on subjective experience.

EXPERIMENT 4B
Replication and Extension of Experiment 4A

In Experiment 4B, we replicated and extended Experiment 4A by including a weak RI condition, which could be construed as falling between the SI and RI conditions with respect to the amount of RI that it generates. In this condition, distractors were not part of the response set. Although no responses had been learned toward these objects in the laboratory, it was assumed that, as environmental stimuli, they would still elicit action plans (e.g., exploratory behavior, such as attending and orienting to them; Tinbergen, 1952). Moreover, these distractors should induce greater interference and subjective effects than those of the SI condition, because they are not associated with the correct response. We took the opportunity to replicate the procedures of Experiment 4A using shape stimuli instead of letter stimuli in order to weaken the strength of distractors in the weak RI condition, because it is well known that orthographic stimuli are already strongly linked to automatic action plans (see Roelofs, Meyer, & Levelt, 1995).

Method

Participants

Yale University students ($N = 22$) participated for class credit or $8.

Procedure

The procedure was identical to that of Experiment 4A, except that the participants responded to shapes instead of letters, and in addition to the three standard conditions of RI, SI, and *identical*, which served as our primary conditions of interest, we included a weak RI condition featuring distractors (wavy lines and objects resembling lightning bolts) that were not part of the response set and presented each an equal number of times (six) with each kind of target, totaling 48 trials. For the sake of comparison, an *alone* condition presented targets by themselves (24 trials, six replications per shape).

During training, the participants pressed a key (occupying the "4" position on the number pad of the keyboard) with their right index finger when presented with a circle or a square; when presented with a triangle or plus sign, they pressed the adjacent ("5") key with their right middle finger. After training, the participants were told that the remainder of the experiment (168 trials) would involve a similar task and that they should continue to respond to the shape in the center of the screen (the target), although they must now also disregard whatever stimuli appear peripherally (the distractors). The participants were now encouraged to respond as quickly and as accurately
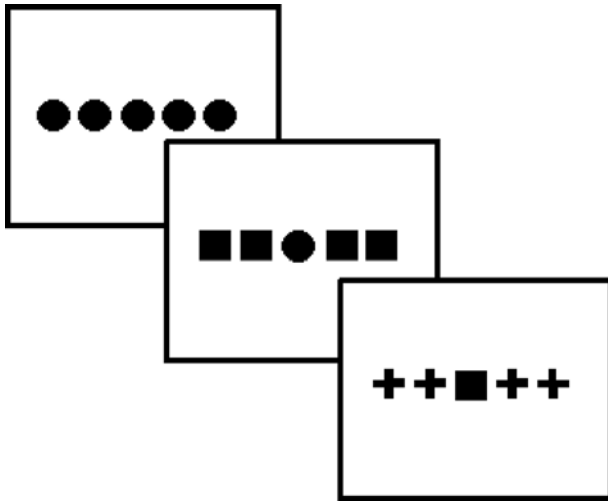
Figure 7. Sample distractor conditions (Experiment 4B), from top to bottom: *identical*, stimulus interference, and response interference. Not drawn to scale.

as possible and to avoid anticipations. The distractors horizontally flanked the target in the center of the screen (Figure 7), occupying a rectangular region less than 10 cm in length and 2 cm in height, so that the stimuli would fit well within the participant's visual field. In the *identical* condition, targets were flanked by distractors that were identical to it (24 trials, six replications for each shape). In the SI condition, targets and distractors were associated with the same response but comprised different shapes (e.g., a square flanked by circles; 24 trials, six replications of each possible combination). In the RI condition, targets and distractors were associated with different responses (e.g., a square flanked by triangles; 48 trials, six replications of each possible combination). Each trial began with a warning prompt (question mark) and beep, which preceded the stimulus display by 1,300 msec. After responding, the participants rated their perceived level of activity on an 8-point scale, in which 1 signified *almost no urge* and 8 signified *extremely strong urge*.

Results

Introspection Training

Trimming resulted in the loss of 6 (1.1%) of 528 RT data points. We replicated the Stroop effect; RTs were longest for the incongruent condition, followed by the control and congruent conditions [$F(2,42) = 40.906, p < .0001, \eta_p^2 = .66$]. Stroop condition produced analogous effects on the urge to err [$F(2,42) = 40.627, p < .0001, \eta_p^2 = .66$]; urges were greatest for the incongruent condition ($M = 4.40, SEM = 0.28$), followed by the control ($M = 3.26, SEM = 0.26$) and congruent ($M = 1.73, SEM = 0.16$) conditions (Fisher's PLSD, $ps < .01$). Typing errors resulted in the loss of 2 (0.4%) of the 528 ratings. Fourteen of the 22 participants had significant ($ps < .05$) positive correlations between RT and activity (mean $r = .50$, Fisher's r to $z, p < .05$), suggesting that the participants may have made their judgments on the basis of RTs.

Flanker Task

Errors and RTs. Errors ($n = 169$) and trials on which no responses were made (28 trials) resulted in the loss of 197 (5.3%) of 3,696 data points. See Appendix B for the mean error rate per condition. Condition did not influence error rates [$F(4,84) = 0.601, p > .50$]. RT trimming resulted in the loss of 38 (1.0%) of 3,696 trials. Condition did not influence the proportion of trimmed trials [$F(4,84) = 1.14, p > .34$]. The RT data were analyzed in a within-subjects design ANOVA, with environment as a five-level factor (*alone*, *identical*, SI, RI, weak RI). The mean RT across all conditions was 672.77 msec ($SEM = 22.06$). As is shown in Table 1, and replicating previous findings (C. W. Eriksen & Schultz, 1979; van Veen et al., 2001), there was a main effect of condition [$F(4,84) = 5.307$, $p < .001, \eta_p^2 = .20$]. As in Experiment 4A, planned comparisons revealed that the RI condition produced greater RTs than did the SI and *identical* conditions ($ps_{paired} < .05$). Comparisons also revealed that the only nonsignificant differences ($ps > .05$) in RTs were between the RI and weak RI, SI and weak RI, SI and *alone*, *identical* and weak RI, and SI and *identical* conditions (difference = 18.14 msec, $p = .199$). Not having obtained a significant difference between the SI and *identical* conditions—traditionally a small but reliable effect that has been shown to range from 15 to 20 msec (cf. van Veen et al., 2001)—may simply reflect excessive noise from the peculiarities of our multiple conditions and stimuli. The same general pattern of results was obtained when only RTs that were more than 2.5 SDs from each participant's mean score in each condition were trimmed [$F(4,84) = 3.850, p = .006, \eta_p^2 = .15$].

Subjective Effects

The ratings from one session were excluded because the participant failed to follow the instructions. Of the remaining data, typing errors and omissions resulted in the loss of 39 (1.1%) of 3,528 ratings.

Principal results. As is illustrated in Figure 8, reported activity increased as RI increased, with activity being highest for the RI condition, followed by the weak RI, SI, *identical*, and *alone* conditions. Statistically, there were significant differences in activity among the six conditions [$F(4,80) = 12.274, p = .0001, \eta_p^2 = .38$]. Replicating the subjective effects of Experiment 4A, planned comparisons revealed that more activity was found for the RI condition ($M = 2.92, SEM = 0.14$) than for the SI ($M = 2.47, SEM = 0.13$) and *identical* ($M = 2.26, SEM = 0.15$) conditions ($ps_{paired} < .01$). Fisher's PLSD revealed that all the differences among these three conditions were significant ($ps < .05$), except for that between the *identical* and SI conditions ($p = .32$).

Table 1
Mean Response Times (RTs, in Milliseconds)
As a Function of Distractor Environment in Experiment 4B

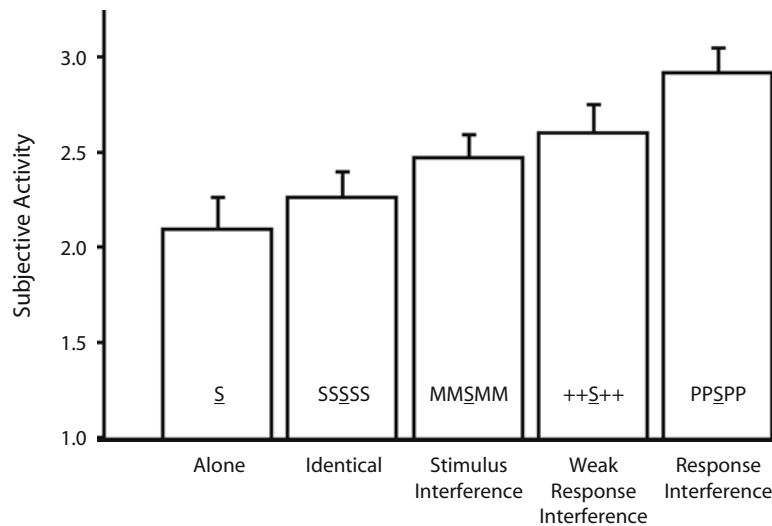| Distractor Environment	RT	
	M	*SEM*
Alone	640.46	26.53
Identical	648.14	24.69
Stimulus interference	666.28	22.68
Weak response interference	681.49	25.21
Response interference	695.70	20.47

Figure 8. Mean subjective activity as a function of flanker condition (Experiment 4B), exemplified by stimuli from the introduction (not actual experimental stimuli).

Secondary results. Planned comparisons revealed that more activity was reported for the RI than for the weak RI condition ($M = 2.60$, $SEM = 0.16$) ($p_{paired} < .01$) but that the weak RI and SI conditions led to comparable activity (difference = 13 msec, $p_{paired} = .17$). Fisher's PLSD revealed that the only nonsignificant ($ps > .05$) post hoc differences of reported activity were between the *alone* and *identical*, *alone* and SI, *identical* and SI, *identical* and weak RI, and SI and weak RI conditions. The same pattern of results was obtained when we removed the trials of the weak RI condition in which novel distractors were presented for the first time and were presumably the most distracting.

For each participant, there was a significant ($ps < .05$) positive correlation between RT and activity (mean $r = .495$, Fisher's r to z, $p < .05$), suggesting, again, that the participants may have made their judgments on the basis of RTs. After the experiment, 8 of the 21 participants reported that they did not know why the most difficult flanker conditions were so difficult, and only 4 participants surmised that task difficulty was somehow due to the actions associated with distractors. The remaining 9 participants provided other kinds of explanations, such as that task difficulty was due to the visual complexity of the distractors.

Discussion

We successfully replicated Experiment 4A with non-orthographic stimuli and with additional conditions that differed in the amount of RI that they elicit. The RI condition produced more subjective activity than did the weak RI and SI conditions, as was predicted by the hypothesis that subjective modulations are driven primarily by RI. Regarding our secondary results, it seems that responding to targets in the presence of distractors that are not in the response set is associated with high subjective modulations, relative to, say, that of the *identical* condition. It is

reasonable to propose that this is because the distractors in the weak RI condition elicited exploratory action tendencies that needed to be suppressed in order for the participants to perform the task successfully or because they were not associated with the correct response. Further investigation is necessary to better understand the relationship between these subjective modulations and different forms of interference.

Experiments 4A and 4B were not designed to rule out the alternative hypothesis that, in making their judgments, the participants did not actually introspect their subjective states but simply monitored their RTs or speed of processing. Conceptually, it is difficult to imagine how such confounding cognitive dynamics could be eradicated from these tasks. Both the properties of speed of processing and of fluency of processing will always be involved in cognitive interference tasks, and participants will always have direct or indirect access to these features (Winkielman, Schwarz, Fazendeiro, & Reber, 2003). Yet this alternative hypothesis seems less likely, given the time scales involved in our trials—time scales in which it is probably difficult for people to introspect and monitor their own RTs (Libet, 2004; but cf. Corallo, Sackur, Dehaene, & Sigman, 2008). For example, the average RT difference between the SI and RI conditions was on the order of a mere 30 msec (a difference that would be challenging to introspect; Buzsáki, 2006; Libet, 2004), yet the participants reported different degrees of subjective activity for these conditions. (Even the greatest mean RT difference among conditions was on the order of a mere 55 msec.) In addition, in Experiments 1 and 2, RTs did not always covary with the magnitude of subjective effects. In general, it may be that urges and RTs are distinct consequences of cognitive interference, but that it is difficult, if not impossible, to separate the two. Observing one's RTs could influence judgments regarding urges; given the difficulty of introspecting RTs at this time scale, perhaps urges too could inform judgments about RTs.

Nevertheless, to weaken the potential influence of overt RTs on judgments, the participants in Experiment 4C were instructed to respond to targets at the same time, after hearing a beep that always sounded 1,200 msec after stimulus presentation. The results of piloting ($N = 4$) and of previous research (C. W. Eriksen & Schultz, 1979) suggest that such a delay is sufficient to eliminate flanker RT effects.

EXPERIMENT 4C
Reducing the Potential Influence of Overt RTs on Judgments

Method

Participants

Yale University students ($N = 17$) participated for class credit or $8.

Procedure

The procedure was identical to that of Experiment 4B, except that the participants were instructed to prepare to respond as soon as the targets appeared but to withhold responding until hearing a beep, which always sounded 1,200 msec after stimulus presentation. Specifically, they were told the following:

> Prepare to respond only to the shape in the center of the screen and, while you are doing so, look inside your head and introspect how much "activity" you feel as you prepare to respond. But respond as fast and as accurately as possible only once you hear the beep. Report the "activity" that you experienced as soon as you saw the shapes and prepared to respond, although you withheld the response until you heard the beep.

We did not regard the RTs associated with this delayed response to be an informative dependent measure with respect to cognitive processing. Our primary goal was to obtain subjective data while diminishing the potential influence that overt RTs had on judgments.

Results

Introspection Training

RT trimming resulted in the loss of 6 (1.5%) of 408 trials. Again, we replicated the Stroop effect [$F(2,32) = 19.93$, $p < .0001$], and condition produced analogous effects on ratings [$F(2,32) = 55.61$, $p < .0001$, $\eta_p^2 = .78$]; urges were greatest for the incongruent condition ($M = 4.10$, $SEM = 0.30$), followed by the control ($M = 2.49$, $SEM = 0.27$) and congruent ($M = 1.36$, $SEM = 0.10$) conditions (Fisher's PLSD, $ps < .01$). Typing errors (1% of the data set) were excluded from the analysis. Fifteen of the 17 participants had significant ($ps < .05$) positive correlations between RTs and the urge to err (mean $r = .61$, Fisher's r to z, $p < .05$).

Flanker Task

Errors and RTs. The RT data from one session were excluded from the analysis because the participant buttonpressed before the beep on every trial. Of the remaining data, errors resulted in the loss of 58 (2.2%) of 2,688 data points. See Appendix B for the mean error rate per condition. Condition did not influence error rates [$F(4,64) = 0.516$, $p > .72$]. Delaying a response and executing it upon hearing an anticipated beep is far easier than responding to the targets of the previous experiments. Thus,

we now removed RTs below 100 msec and above 1 sec. This resulted in the loss of 256 (9.5%) of 2,688 RT data points, which was expected, given that the task involved the unnatural delay of a response and required the participants to pay close attention to the auditory cue. Condition did influence the proportion of trimmed trials [$F(4,60) = 4.005$, $p < .05$], but Fisher's PLSD revealed that none of these differences was significant ($ps > .05$). (The same pattern of results was obtained with the previous trimming procedure and with RTs below 100 msec and above 2 sec removed, although either procedure led to a substantially greater loss of data.) The same pattern of results was obtained with all of the RT data included or with only RTs that were greater than 2.5 SDs from each participant's mean score in each condition trimmed [$F(2,56) = 1.096$, $p > .35$, $\eta_p^2 = .07$].

As was anticipated (C. W. Eriksen & Schultz, 1979), by our delaying participants' responses, traditional flanker RT effects were effectively eliminated: The RI ($M = 374.53$, $SEM = 30.71$), SI ($M = 375.25$, $SEM = 28.98$), and *identical* ($M = 363.66$, $SEM = 31.71$) conditions did not yield significant differences in RTs [$F(2,30) = 0.58$, $p > .58$, $\eta_p^2 = .03$] (planned comparison, $ps_{paired} > .44$). Even when all of the conditions were included, there was no main effect of condition [$F(4,60) = 1.937$, $p > .10$, $\eta_p^2 = .11$]. Nine of the 16 participants had significant ($ps < .05$) positive correlations between RTs and activity (mean $r = .21$, Fisher's r to z, $p < .05$).

Principal results. The present data replicated the general pattern of results of Experiments 4A and 4B: There were significant differences in activity among the six conditions [$F(4,60) = 9.568$, $p = .0001$, $\eta_p^2 = .39$], and more activity was found for the RI condition ($M = 2.56$, $SEM = 0.30$) than for the SI ($M = 2.21$, $SEM = 0.22$) and *identical* ($M = 1.67$, $SEM = 0.18$) conditions ($ps_{paired} < .05$). Omitted responses and typing errors resulted in the loss of 20 (0.7%) of 2,688 ratings.

Secondary results. Regarding the *alone* and weak RI conditions, additional planned comparisons revealed that the only nonsignificant differences in activity were between the RI and the weak RI conditions ($M = 2.24$, $SEM = 0.26$), and between weak RI ($M = 2.24$, $SEM = 0.26$) and the SI ($M = 2.21$, $SEM = 0.22$) conditions. Again, the least reported activity was for the *alone* condition ($M = 1.45$, $SEM = 0.17$). After the experiment, 10 of the 17 participants reported that they did not know why the most difficult conditions were so difficult, and only 4 participants reported that task difficulty was somehow due to the actions associated with distractors. The remaining 3 participants provided other kinds of explanations, such as that the responses were influenced by the phonological similarity between the targets and the distractors.

GENERAL DISCUSSION

Because of recent developments (e.g., Mayr, 2004; Mayr et al., 2003; Mulert et al., 2005; Rosen et al., 2007), the *Zeitgeist* has arrived to investigate the fleeting "urges," "tendencies," and "inclinations" that thou-

sands of laboratory participants have experienced when naming the color blue when it was presented on the word RED or when performing other kinds of interference tasks (e.g., flanker and Simon tasks; Simon, Hinrichs, & Craft, 1970). In our initial enquiry to catalog and understand the nature of these elusive subjective phenomena, we demonstrated that participants appear to be able to introspect the subjective aspects (e.g., perceptions of difficulty, control, and competition) of responding in interference tasks and that they can do this reliably on a trial-by-trial basis. Stronger subjective effects were systematically associated with experimental conditions featuring high levels of RI. Specifically, when RI was low or absent, urges to err and perceptions of competition tended to be low, whereas perceptions of control tended to be high; when RI was high, urges to err and perceptions of competition tended to be high, whereas perceptions of control tended to be low. This observation was further corroborated by the results from Experiments 4A–4C, which were designed to tease apart the subjective effects of SI and RI. Theories on cognitive control, effortful processing, and conscious processing will have to account for this catalog of basic, reliable findings.

More generally, these new data corroborate the notion that similar effects are obtained for externalized and internalized actions (Bargh & Morsella, 2008; Vygotsky, 1962) and that these subjective effects do not stem only from conflict at the level of effectors, although effectors are often engaged in subthreshold (imperceptible) ways during cognitive interference (Coles et al., 1985). In our experiments, it seems that the locus of the subjective effects involved a central process and that action execution was unnecessary for these effects. In addition, we have introduced some initial, tentative evidence for the phenomenon of double blindness. Perhaps double blindness features more strongly in the congruent conditions of countermanding tasks such as the antisaccade task (Curtis & D'Esposito, 2009). In general, the notion of double blindness is consistent with the view that one is conscious only of the outputs of processes, not of the processes themselves (Jackendoff, 1990; Lashley, 1951).

Limitations of the Present Approach

With these findings, one must be careful about making claims regarding the subjective effects of RI versus other kinds of interference. At this stage of understanding, it is difficult—if not impossible—to eradicate the influence of processing speed, processing fluency, or a general sense of effort (or a combination thereof) on the judgments made by participants. As with other introspective measures, it is challenging to verify what the participants introspect at the moment that they make their judgment. Self-reports are far from infallible, even if they occur just seconds after the relevant conscious experience (Block, 2007).

Regarding speed, along with the observation that RTs did not always correlate with reported urges, the positive results of Experiment 4C suggest that the subjective effects obtained in our experiments may not have been just artifacts of the participants' making their judgments on the

basis of observed RTs. Moreover, it is not obvious how simply making one's judgments on the basis of RTs could lead to the rich kinds of subjective effects reported in Experiment 3, in which the participants yielded systematic introspections regarding their perceptions of difficulty, control, and competition. Then again, the participants may have made all of their judgments on the basis of an overall sense of difficulty, which may be introspected directly or indirectly from inferences about RT performance. If the former, one must then ask what it is about interference in the Stroop task that is so effortful and engenders changes in subjective experience. More generally, what is meant by *effort*? It is known that subjective effort is a complex construct that—like subjective pain (for a review, see Grahek, 2007)—is not linked to physiological processes in any straightforward manner (Bartley & Chute, 1947; Baumeister & Vohs, 2004; Botvinick, 2007; Kahneman, 1973; Preston & Wegner, 2009; Rosen et al., 2007). There is a whole world of literature demonstrating discrepancies between objective and subjective indices of cognitive effort. For example, participants may be oblivious to the occurrence of neural events that are metabolically or computationally costly, or they may conflate motivational states with subjective fatigue (Bartley & Chute, 1947; Baumeister & Vohs, 2004; Kahneman, 1973; Preston & Wegner, 2009). (For a case in which normal Stroop performance is dissociated from a subjective sense of effort, see Naccache et al., 2005.) Perhaps cognitive effort in tasks such as the Stroop paradigm stems from one's (to use a homuncular description) having to experience incompatible action plans (Cohen et al., 1990; Morsella, 2005) or to suppress a prepotent response (Baumeister, Gailliot, & Tice, 2009). Regarding suppression, we ascribe to the nonhomuncular and anticentral executive perspective that conflict stems not from representations conflicting with an internal supervisor-like system (e.g., Norman & Shallice, 1980) but from the strength of competition among representations (Curtis & D'Esposito, 2009; Morsella, 2005).

A related limitation of these studies, which is perhaps inherent in all flanker paradigms (Coles et al., 1985; B. A. Eriksen & Eriksen, 1974; C. W. Eriksen & Schultz, 1979; van Veen et al., 2001), is that RI still features some SI, possibly rendering it more complicated than the latter. This confound alone could lead to the kinds of subjective effects reported above. Unfortunately, a flanker-like paradigm that can induce RI without also invoking SI has yet to be developed. Hypothetically, this could be instantiated by having targets and distractors be perceptually identical but somehow cue different responses, if such a scenario is possible. However, there is evidence that RI is qualitatively distinct from SI and that its behavioral and subjective effects are not simply the outcome of increased difficulty. In a neuroimaging study, van Veen et al. demonstrated that, although both RI and SI are associated with differences in performance, the former is the condition that most activates the anterior cingulate cortex, a brain region located on the medial surface of the frontal lobe that is interconnected with many motor areas and is believed to be involved in cognitive monitoring (cf. Botvinick, 2007;

Botvinick et al., 2001; Brown & Braver, 2005). Consistent with the idea that the incompatibility of plans is what primarily drives the subjective effects in interference paradigms, it has been shown that, independent of suppression or SI/perceptual interference, and on the basis of a priori theoretical predictions, merely sustaining incompatible intentions (e.g., to point left and right) leads to subjective effects that are greater than those associated with sustaining compatible intentions (e.g., to point left and utter a word; Kang et al., 2008; Morsella, Gray, et al., 2009).

Because of how little is known about the nature of cognitive/subjective effort (but see Sanders, 1983) and because of the limitations inherent in all introspection paradigms, we cannot rule out that our pattern of subjective effects (e.g., perceptions of control and competition) stem from only a sense of difficulty, which could be deduced by participants directly or indirectly. Similarly, at this stage of understanding, we cannot rule out that judgments were based on self-observations involving RT performance or on folk beliefs regarding how one should comport oneself in an experiment about cognitive control. Perhaps the participants made their ratings on the basis of heuristics such as, "if the Stroop trial is incongruent, then I will report 6 as the rating." Although this cannot be fully ruled out by the present experiments, this alternative seems unlikely, given that the participants' ratings tended to vary across trials within each condition. For instance, for incongruent Stroop trials, the first 8 ratings from a participant selected at random from Experiment 1 were 1, 5, 3, 1, 5, 4, 7, and 7 (the mean SDs for each condition of Experiment 1 were 0.37, 0.98, 0.66, and 1.61 for the congruent, control, neutral, and incongruent conditions, respectively). Of course, it may well be that the participants used a more sophisticated and nuanced heuristic when engendering the present pattern of results.

Future investigations on cognitive effort and control will certainly be needed to qualify the kinds of conclusions that can be drawn from this initial project.

Observations to Spur Future Investigation

During training in Experiments 4A–4C, the participants were capable of introspecting aspects of cognitive processing, despite the fleetingness of color naming, an act lasting less than 1 sec. For all of the flanker tasks, less than 23% of the participants were capable of discerning the source of these subjective effects, which is consistent with the view that one can be conscious of tendencies (e.g., urges and inclinations), but not necessarily of the factors engendering such tendencies (Baker, Piper, McCarthy, Majeskie, & Fiore, 2004). Consistent with findings from metacognition and social cognition research (Metcalfe, Funnell, & Gazzaniga, 1995; Nisbett & Wilson, 1977; Roser & Gazzaniga, 2004), we found in our informal post-session interviews that the participants tended to provide varied accounts about the source of their judgments.

Our pattern of results is consistent with tenets of Sanders's (1983) cognitive energetic model. In the model, controlled processes such as response selection and top-down attentional control are energy/resource dependent,

whereas automatic processes such as stimulus preprocessing are not. In the model, different stages of processing rely on distinct energetical resources. Levels of arousal are most influential with respect to the input stages of processing (e.g., feature extraction), and levels of activation are most influential with respect to the response stages of processing. (Because stimulus preprocessing is automatic, it does not require a separate energetical resource.) For a given task, there is an optimal level of arousal and of activation (Fischer, Langner, Birbaumer, & Brocke, 2008; Kahneman, 1973). To reach this optimal level, effort can be used to adjust arousal and activation. If effort is overloaded or fails to implement the necessary energetical adjustments, stress arises. Consistent with our finding that strong changes in consciousness (e.g., the urge to err) accompany interference that targets response selection, effort at the response choice stage is construed as being intimately associated with conscious processing and with the conflict type of stress (Sanders, 1983, p. 81).

More generally, our pattern of results is consistent with the observation that conflicts occurring at perceptual levels of processing (e.g., intersensory conflicts, as with ventriloquism) are not as subjectively taxing as those occurring at response selection levels of processing, whether in approach–avoidance conflicts (Livnat & Pippenger, 2006; Miller, 1959) or the delay of gratification (Metcalfe & Mischel, 1999; Morsella, 2005). Figuratively speaking, people tend not to experience any mental strife while watching a ventriloquist or being subjected to the McGurk effect[6] (McGurk & MacDonald, 1976), but such is apparently not the case when they perform the Stroop task or exert self-control (Baumeister & Vohs, 2004).

According to supramodular interaction theory (Morsella, 2005), these findings can be explained by the hypothesis that people are most likely to be conscious of conflicts involving competition for control of the skeletal–muscle system, because the primary function of consciousness is to integrate such incompatible skeletomotor intentions. From this standpoint, conscious conflicts stem from incompatible skeletomotor intentions, such as when one suppresses a prepotent response, diets, suppresses emotions, holds one's breath while underwater, or inhibits a prepotent response in a laboratory interference paradigm (Morsella, 2005). From this standpoint, regarding the conflicts occurring at the different stages of processing, consciousness is required to integrate information at the response-selection end of processing.

Accordingly, incompatible skeletomotor intentions (e.g., to point right and left, to eat and not eat, to inhale and not inhale) produce strong, systematic changes in consciousness. For example, in a paradigm similar to that of Experiments 4A–4C, in which the participants were trained to introspect conflict-related aspects of cognition during an interference task and then to introspect the same "thing" while sustaining compatible intentions (e.g., pointing left with a given finger and vibrating that finger) and incompatible intentions (e.g., to point left and right with the same finger), the participants reported stronger systematic changes in subjective experience when sustaining incompatible intentions

than when sustaining compatible skeletomotor intentions, even though the participants were always in a motionless state (Kang et al., 2008; Morsella, Gray, et al., 2009).

In conclusion, we believe that the limitations of this initial enquiry into the nature of the fleeting subjective effects that accompany cognitive interference will be remedied easily by future investigation. Theories on cognitive control and conscious processing will have to account for these reliable and replicable patterns of results. More generally, we believe that, just as RTs can reveal aspects of cognitive processing that may not be detectable through less subtle behavioral measures (e.g., response accuracy), measures of the subjective aspects of processing may illuminate features of cognitive processing that are undetectable in standard behavioral and psychophysiological measures. (For example, in Experiment 4A, the subjective effect was larger than the behavioral RT effect.) We hope that, with the present paradigms and by targeting brain regions involved in cognitive control (cf. Kang et al., 2008), future researchers may identify the neural correlates of these subjective phenomena and assess their role in negative affect (e.g., stress and anxiety) and failures of self-regulation, where disharmony from strong forms of response conflict (as in the delay of gratification) seems to play a critical role (Baumeister & Vohs, 2004; Metcalfe & Mischel, 1999).

REFERENCES

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., & Fiore, M. C. (2004). Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review*, **111**, 33-51.

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science*, **3**, 73-79.

Bartley, S. H., & Chute, E. (1947). *Fatigue and impairment in man*. New York: McGraw-Hill.

Baumeister, R. F., Gailliot, M. T., & Tice, D. M. (2009). Free willpower: A limited resource theory of volition, choice, and self-regulation. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 487-508). New York: Oxford University Press.

Baumeister, R. F., & Vohs, K. D. (Eds.) (2004). *Handbook of self-regulation: Research, theory, and applications*. New York: Guilford.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral & Brain Sciences*, **30**, 481-548.

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, **7**, 356-366.

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.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, **307**, 1118-1121.

Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

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

Cohen, J. [D.], MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 529-553.

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection: Distorted subjective time during the dual-task bottleneck. *Psychological Science*, **19**, 1110-1117.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, **6**, 119-126.

Curtis, C. E., & D'Esposito, M. (2009). The inhibition of unwanted actions. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 72-97). New York: Oxford University Press.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143-149.

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

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness & Cognition*, **9**, 288-307.

Fischer, T., Langner, R., Birbaumer, N., & Brocke, B. (2008). Arousal and attention: Self-chosen stimulation optimizes cortical excitability and minimizes compensatory effort. *Journal of Cognitive Neuroscience*, **20**, 1443-1453.

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, **8**, 1298-1300.

Gazzaley, A., & D'Esposito, M. (2007). Unifying prefrontal cortex function: Executive control, neural networks and top-down modulation. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 187-206). New York: Guilford.

Grahek, N. (2007). *Feeling pain and being in pain* (2nd ed.). Cambridge, MA: MIT Press.

Gray, J. A. (1995). The contents of consciousness: A neuropsychological conjecture. *Behavioral & Brain Sciences*, **18**, 659-676.

Jackendoff, R. S. (1990). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kang, Y. N., Morsella, E., Shamosh, N. A., Bargh, J. A., & Gray, J. R. (2008). *The essence of subjective conflict during self-control: Neural correlates of sustaining incompatible intentions*. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, San Francisco.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112-146). New York: Wiley.

Libet, B. (2004). *Mind time: The temporal factor in consciousness*. Cambridge, MA: Harvard University Press.

Livnat, A., & Pippenger, N. (2006). An optimal brain can be composed of conflicting agents. *Proceedings of the National Academy of Sciences*, **103**, 3198-3202.

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**, 383-391.

Mayr, U. (2004). Conflict, consciousness, and control. *Trends in Cognitive Sciences*, **8**, 145-148.

MAYR, U., AWH, E., & LAUREY, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, **6**, 450-452.

MCGUIGAN, F. J. (ED.) (1966). *Thinking: Studies of covert language processes*. New York: Appleton-Century-Crofts.

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

METCALFE, J., FUNNELL, M., & GAZZANIGA, M. S. (1995). Right hemisphere memory superiority: Studies of a split-brain patient. *Psychological Science*, **6**, 157-165.

METCALFE, J., & MISCHEL, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, **106**, 3-19.

MILLER, N. E. (1959). Liberalization of basic S–R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science. Study 1: Conceptual and systematic. Vol. 2: General systematic formulations, learning, and special processes* (pp. 196-292). New York: McGraw-Hill.

MORSELLA, E. (2005). The function of phenomenal states: Supramodular interaction theory. *Psychological Review*, **112**, 1000-1021.

MORSELLA, E. (2009). The mechanisms of human action: Introduction and background. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 1-32). New York: Oxford University Press.

MORSELLA, E., GRAY, J. R., KRIEGER, S. C., & BARGH, J. A. (2009). The essence of conscious conflict: Subjective effects of sustaining incompatible intentions. *Emotion*, **9**, 717-728.

MORSELLA, E., & KRAUSS, R. M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture–speech theories. *Journal of Psycholinguistic Research*, **34**, 415-427.

MORSELLA, E., RIGBY, T., & GAZZALEY, A. (2009). *A working-memory version of the flanker task yields traditional perceptual- and response-interference effects*. Unpublished manuscript, San Francisco State University.

MULERT, C., MENZINGER, E., LEICHT, G., POGARELL, O., & HEGERL, U. (2005). Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *International Journal of Psychophysiology*, **56**, 65-80.

NACCACHE, L., DEHAENE, S., COHEN, L., HABERT, M.-O., GUICHART-GOMEZ, E., GALANAUD, D., & WILLER, J.-C. (2005). Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, **43**, 1318-1328.

NAGEL, T. (1974). What is it like to be a bat? *Philosophical Review*, **83**, 435-450.

NISBETT, R. E., & WILSON, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, **84**, 231-259.

NORMAN, D. A., & SHALLICE, T. (1980). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1-18). New York: Plenum.

PICKERING, M., & GARROD, S. (in press). Prediction and embodiment in dialogue. *European Journal of Social Psychology*.

PRESTON, J., & WEGNER, D. M. (2009). Elbow grease: When action feels like work. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 569-586). New York: Oxford University Press.

ROELOFS, A., MEYER, A. S., & LEVELT, W. J. M. (1995). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 246-251.

ROSEN, Z. V., MCGUIRE, J., & BOTVINICK, M. M. (2007). *Is mental effort aversive? Some behavioral and psychophysiological evidence*. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, New York.

ROSENBAUM, D. A. (2005). The Cinderella of psychology: The neglect of motor control in the science of mental life and behavior. *American Psychologist*, **60**, 308-317.

ROSER, M., & GAZZANIGA, M. S. (2004). Automatic brains—Interpretive minds. *Current Directions in Psychological Science*, **13**, 56-59.

SANDERS, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, **53**, 61-97.

SIMON, J. R., HINRICHS, J. V., & CRAFT, J. L. (1970). Auditory S–R compatibility: Reaction time as a function of ear–hand correspondence and ear–response-location correspondence. *Journal of Experimental Psychology*, **86**, 97-102.

STEVENS, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. *American Journal of Psychology*, **69**, 1-25.

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

TINBERGEN, N. (1952). Derived activities: Their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology*, **27**, 1-32.

VAN VEEN, V., & CARTER, C. S. (2006). Conflict and cognitive control in the brain. *Current Directions in Psychological Science*, **15**, 237-240.

VAN VEEN, V., COHEN, J. D., BOTVINICK, M. M., STENGER, V. A., & CARTER, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage*, **14**, 1302-1308.

VYGOTSKY, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Eds. & Trans.). Cambridge, MA: MIT Press.

WEGNER, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

WINKIELMAN, P., SCHWARZ, N., FAZENDEIRO, T. A., & REBER, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Mahwah, NJ: Erlbaum.

WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology* (Rev. ed.). New York: Holt, Rinehart & Winston.

NOTES

1. An organism is said to possess a subjective experience (the elusive phenomenon falling under the rubrics of consciousness, sentience, or basic awareness) if there is *something it is like* to be that organism—something it is like, for example, to be human and experience pain, breathlessness, or yellow afterimages (Nagel, 1974).

2. In this task, participants name the colors in which stimulus words are written. When the word and color are incongruent (e.g., RED presented in blue), response conflict leads to increased error rates and RTs (Cohen, Dunbar, & McClelland, 1990). When they are congruent (e.g., RED presented in red), there is little or no interference (for a review, see MacLeod & MacDonald, 2000).

3. One reason that these ineffable phenomena have remained underexplored may be because the basic relationship between nervous processes and subjective experience remains profoundly mysterious (Crick & Koch, 2003; Gray, 1995). Another reason may reflect the dominance of behaviorism in the first half of the twentieth century; behaviorism may have served as a healthy reaction to structuralism, which attempted to explain all operations in terms of conscious processes.

4. This, however, cannot be said for the whole or even most of experimental psychology, since the subjective aspects of cognitive processing have been examined systematically in research on psychophysics (e.g., Stevens, 1956), metacognitive processes (e.g., Fernandez-Duque, Baird, & Posner, 2000; Johnson & Raye, 1981), and memory and perceptual fluency (e.g., Jacoby, Kelley, & Dywan, 1989).

5. To obtain a sense of the baseline urge to read for the words used in our Stroop task, we used a procedure similar to that of Experiment 2: Another group of participants ($n = 15$) rated how strong their urge was to read the words when the words were presented in standard, uncolored (black) letters. In this passive viewing task, the participants pressed the space bar when presented with the stimulus and then rated their urge to read the word. The urges to read were comparable for the color and control words (color word, $M = 6.67$, $SEM = 0.47$; control word, $M = 6.68$, $SEM = 0.44$) [$t(14) = -0.015$, $p = .99$]. Of course, it is difficult to compare these urges with the (weaker) reported urges to read for the congruent condition of the Stroop task, because urges are probably context-sensitive and the two tasks differ in various respects.

6. For example, the McGurk effect (McGurk & MacDonald, 1976) involves a conflict between visual and auditory information: An observer views a speaker mouthing /ba/ while presented with the sound /ga/. Surprisingly, the observer is unaware of conflict, perceiving only /da/.

APPENDIX A
Mean Proportion of Response Trial Data Removed by
Trimming, As a Function of Experiment and Condition

Condition	Subvocal		Vocal	
	M	*SEM*	*M*	*SEM*
Experiment 1				
Congruent	.017	.009		
Neutral	.022	.007		
Control	.029	.014		
Incongruent	.068	.028		
Experiment 2				
Congruent	.017	.003	.002	.001
Neutral	.020	.004	.002	.001
Control	.027	.005	.002	.001
Incongruent	.058	.008	.004	.001
Experiment 3				
Congruent	.024	.006	.008	.002
Neutral	.027	.007	.002	.001
Control	.030	.009	.009	.003
Incongruent	.074	.015	.013	.002

APPENDIX B
Mean Proportion of Data Removed
From Experiments 4A–4C

Condition	Data Removed Due To			
	Trimming		Errors	
	M	*SEM*	*M*	*SEM*
Experiment 4A				
Identical	.009	.004	.047	.007
Stimulus interference	.012	.004	.021	.008
Response interference	.018	.008	.068	.012
Experiment 4B				
Alone	.008	.004	.051	.012
Identical	.010	.005	.042	.009
Stimulus interference	.013	.005	.037	.007
Weak response interference	.007	.003	.050	.007
Response interference	.014	.005	.046	.009
Experiment 4C				
Alone	.128	.026	.023	.008
Identical	.102	.025	.023	.008
Stimulus interference	.065	.014	.024	.009
Weak response interference	.082	.019	.014	.007
Response interference	.104	.023	.026	.013