

Working memory involvement in dual-task performance: Evidence from the backward compatibility effect

RAVID ELLENBOGEN AND NACHSHON MEIRAN
Ben-Gurion University of the Negev, Beer-Sheva, Israel

In three experiments, the authors supported the hypothesis that parallel response activation seen in dual-task performance results from holding Task 2 rules in working memory (WM) while performing Task 1. To this end, the authors used the backward compatibility effect (BCE; quicker primary responses when the Task 2 response is compatible with codes of Task 1) as a marker for parallel response activation and manipulated WM load. Increasing the number of primary task rules from two to four did not modulate BCE, replicating Hommel and Eglau (2002), but a higher load condition, involving six primary task rules, reduced the BCE to nonsignificant levels. Experiment 3 further showed that WM is loaded by rules associating abstract stimulus categories to responses, and not by rules that associate individual stimuli to responses (S–R rules).

People often carry out more than one task at a time, such as talking on the phone while driving. Understanding this ability has become the focus of many recent investigations. In order to explore these issues, researchers use the dual-task paradigm, in which there is temporal overlap between a primary task (Task 1) and a secondary task (Task 2) processing. The performance in these situations is usually accompanied with costs: (1) Both tasks' responses are delayed and less accurate, as compared with a single-task performance level, and (2) only secondary task performance impairment increases as the temporal overlap between the tasks increases (for a review, see Pashler & Johnston, 1998). Three broad approaches have been proposed as possible accounts of dual-task costs: (1) models of structural bottleneck in which serial processing reflects structural limitations (e.g., Byrne & Anderson, 2001; Pashler, 1994, 1998), (2) models suggesting that sequential processing is strategic (e.g., Logan & Gordon, 2001; Meyer & Kieras, 1997a, 1997b), and (3) models that propose capacity sharing between tasks during response selection, together with a change in allocation from Task 1 to Task 2 (e.g., Navon & Miller, 2002; Tombu & Jolicoeur, 2003). As will be explained in the General Discussion section, these theories differ from one another with respect to the putative involvement of working memory (WM) in dual-task performance. In the present study, our aim was to explore the involvement of WM in dual-task performance. Specifically, we focused on how WM load affects the backward compatibility effect (BCE).

The Backward Compatibility Effect

The BCE has been observed in many studies (Caesens, Hommel, Reynvoet, & van der Goten, 2004; Hom-

mel, 1998; Hommel & Eglau, 2002; Lien & Proctor, 2000; Lien, Schweickert, & Proctor, 2003; Miller, 2006; Miller & Alderton, 2006; Watter & Logan, 2006; see also Logan & Delheimer, 2001; Logan & Gordon, 2001; Logan & Schulkind, 2000). These studies have shown that when two tasks overlap in time, processing of the primary task is affected by the compatibility relationship between the secondary task response (R2) and either the primary task response (R1) or the primary task stimulus (S1). These effects are known as the R2–R1 and R2–S1 BCEs, respectively. For instance, Hommel (1998) presented red or green letters (H or S) to participants. In his Experiment 2, R1 involved a left or a right keypress in response to the letter color (red or green), and R2 involved saying “red” or “green” in response to letter identity (H and S). In this experiment, the compatibility relationship was between S1 (stimulus color) and R2 (color-naming response to the letter identity). The results showed that R1 was quicker and more accurate when S1 and R2 were compatible than when they were incompatible.

The Present Hypothesis

According to our hypothesis, task rules are represented in limited capacity WM. When the capacity is not exhausted, it can be shared between the two tasks, and participants hold Task 2 rules in WM while executing Task 1, in order to achieve high preparedness for Task 2 (the second task to be performed). The rules that are held in WM lead to autonomous response activation (Cohen-Kadosh & Meiran, 2007). Because activation is gradual and not all-or-none in nature, Task 2 rules can be partly active while Task 1 is executed, giving rise to the BCE. Our hypothesis accords with modern versions of the *prepared reflex*

R. Ellenbogen, ravidel@bgu.ac.il

metaphor originally suggested by Woodworth (1938; see also Ach, 1910/2006; Hommel, 2000; Logan, 1978) and is most consistent with capacity-sharing theories (e.g., Navon & Miller, 2002; Tombu & Jolicœur, 2003).

A very similar hypothesis was considered by Hommel (1998). According to his *transient-link model*, multiple task rules can be held and implemented in WM, and these rules are used for parallel response activation. From this viewpoint, increasing WM load should impair the degree to which task rules are prepared and should lead to a smaller BCE. The alternative model suggested by Hommel (1998) was the *direct-link model*. According to this model, repeatedly applying the task rules leads to the formation of memory traces outside WM. These direct stimulus–response (S–R) links bypass the central bottleneck and activate responses in parallel. However, the latter alternative is less likely according to recent findings (Hübner & Lehle, 2007), which have shown that BCE is smaller when participants are encouraged to perform serial processing rather than parallel processing.

To decide between these models, Hommel and Eglau (2002) manipulated WM load by varying the number of primary task rules. In their low-load condition, there were two Task 1 rules and two Task 2 rules, whereas in the high-load condition the number of Task 1 rules was increased to four. Because the BCE was unaffected by this manipulation, Hommel and Eglau endorsed the direct-link model.

We suggest that Hommel and Eglau's (2002) rejection of the transient-link model may have been premature, because their WM manipulation was not sufficiently strong to show an effect. This argument is reasonable, since previous studies in which WM load manipulations have been used have often shown that the effect of this manipulation is not monotonic but shows up only when extreme loads are used (see, e.g., Baddeley & Hitch's [1974] classic study). Accordingly, a higher load on WM might reduce the BCE and support the transient-link model.

To test our hypothesis, we adhered closely to the design and analytic approaches employed by Hommel and Eglau (2002) in their Experiment 2. Such adherence was important in order to ensure that any discrepancy between our results and theirs did not indicate a replication failure. There was an additional reason for preferring this method of loading WM. The literature suggests that WM is divided into subsystems (see Baddeley, 2003; Cowan, 1998), which implies that no certainty exists that loading one subsystem (e.g., such as remembering irrelevant digits) will impact the other systems (those presumably holding the task rules). To ensure as much as possible that the system in which task rules are held was loaded, we needed to load WM with task rules (but see Logan, 1978). The fact that an increase in the number of Task 1 rules prolongs response selection is not problematic from our perspective, because slowing response selection should increase, rather than decrease, the BCE. The reason is that with prolonged response selection, there is greater opportunity for interference to take place.

Three levels of WM load were compared in the present work: (1) the *low-load* condition with two color categories for Task 1 (two Task 1 rules), (2) the *high-load* condition with four color categories for Task 1 (four Task 1 rules), and (3) the *higher load* condition with six color categories

for Task 1 (six Task 1 rules). The prediction of the direct-link model is that the BCE will be unaffected by WM load. In contrast, if response activation proceeds via task rules held in WM, as suggested by the transient-link model, the BCE is predicted to be reduced in high and/or higher WM load conditions.

In Experiment 1, we replicated Hommel and Eglau's (2002) Experiment 2. We used their two Task 1 rules (low-load condition) and four Task 1 rules (high-load condition) and replicated their results. As was expected, the BCE was unaffected in the high-load condition with four rules, as compared with a low-load condition with two Task 1 rules. In Experiment 2, we used the higher load condition with six Task 1 rules instead of just four Task 1 rules, with the same two Task 1 rules as in the low-load condition. In this experiment, the BCE in the higher load condition was reduced to a nonsignificant level, as would be expected from the transient-link model. Experiment 3 was run in order to rule out an alternative explanation, as well as to characterize the nature of the task rules that are held in WM and give rise to the BCE. For reasons that will become clearer later on, we refer to category–response (C–R) rules, which link the stimulus category (such as “red”) to the response.

EXPERIMENT 1

This experiment was conducted as a replication of Hommel and Eglau's (2002) Experiment 2. The number of Task 1 rules was manipulated between participants to minimize potential carryover effects.

Method

Participants. Twenty-six undergraduate students from Ben-Gurion University of the Negev participated in this experiment. The participants were assigned to one of the two loading conditions according to the order of entry.

Apparatus and Stimuli. The stimuli were presented on an IBM-PC clone with a 17-in. monitor controlled by software written in E-Prime (Schneider, Eschman, & Zuccolotto, 2002). A white asterisk served as a fixation mark. The stimuli for the primary and the secondary tasks were the uppercase letters H and S (S2), presented in red or green color (S1) in the low-load condition or presented in red, green, yellow, or blue color (S1) in the high-load condition. All the stimuli were presented at the center of a black screen and subtended approximately $0.40^\circ \times 0.30^\circ$ of visual angle from a viewing distance of about 60 cm. Manual responses in the low-load condition (only red and green stimuli) were performed by pressing the *f* (left) and *h* (right) keys of a QWERTY keyboard with the index fingers of both hands. The manual responses in the high-load condition (four colors) were performed by pressing the *f* and *h* keys with the index fingers of both hands to the red or green stimuli (as in the low-load condition) and, in addition, pressing *d* and *j* with the middle fingers of both hands to the blue or yellow stimuli. Vocal responses to letter identity consisted of saying the words “adom” and “yarok,” the Hebrew words for “red” and “green,” respectively, in both load conditions. A microphone collected the vocal responses for the reaction time (RT) analysis, and the experimenter who sat behind the participant entered the correctness scoring of these responses during the intertrial interval (1 sec). Errors were followed by a beep feedback consisting of a 400-Hz tone presented for 50 msec.

Design and Procedure. The experiment consisted of a single session, consisting of one practice block and four experimental blocks. Each block was composed of five replications of each of the four or eight combinations of letter identity and color, randomly

intermixed. Here, each block consisted of 20 trials in the low-load condition (2 colors × 2 letters × 5 replications) or 40 trials in the high-load condition (4 colors × 2 letters × 5 replications), for a total of 80 or 160 experimental trials, respectively.

The procedure for any given trial was as follows. After an intertrial interval of 1 sec, the fixation mark appeared for 1 sec, followed by a blank interval of 250 msec. Then the stimulus, a colored H or S, was presented until both responses were made or until 2,500 msec had elapsed. Manual responses were assigned to stimulus color (Task 1), and vocal responses to letter identity (Task 2). Accordingly, the participants first made a manual response to stimulus color and then made a vocal response to the letter. They were required to perform the two responses in strict serial order as quickly as possible while maintaining high accuracy.

Results and Discussion

The data were treated exactly as in Hommel and Eglau’s (2002) article, since we wanted to ensure that any discrepancy between our results and theirs would not be due to differential procedural or analytical approaches. Data associated with middle finger responses were excluded, because these were available only for the high-load group. Therefore, all the analyses are based exclusively on the data from the two index finger responses and, hence, on conditions

with the stimulus colors red and green. We excluded premature Task 1 responses (RT1 < 150 msec, 0.13%), Task 1 response omissions (RT1 > 2 sec, 2.27%),¹ and R1 and/or R2 error trials (5.45%). Responses in opposite order were excluded from the RT analysis and were considered as errors. For the remaining data, mean RT and the proportion of errors (PE) were computed for each participant according to response (primary or secondary), compatibility (compatible, a red stimulus with a secondary response of “red,” or a green stimulus with a secondary response of “green”; incompatible, a red stimulus with a secondary response of “green,” or a green stimulus with a secondary response of “red”) and alternatives (two or four Task 1 rules). The group mean RTs and PEs for Task 1 are presented in Figure 1A.

An alpha level of .05 was used for all the statistical tests. The ANOVA on RT, with response and compatibility as within-participants independent variables and alternatives as a between-participants independent variable, yielded three significant main effects: response [$F(1,24) = 733.56, MS_e = 5,855.88$], alternatives [$F(1,24) = 10.54, MS_e = 76,024.47$], and compatibility [$F(1,24) = 34.58, MS_e =$

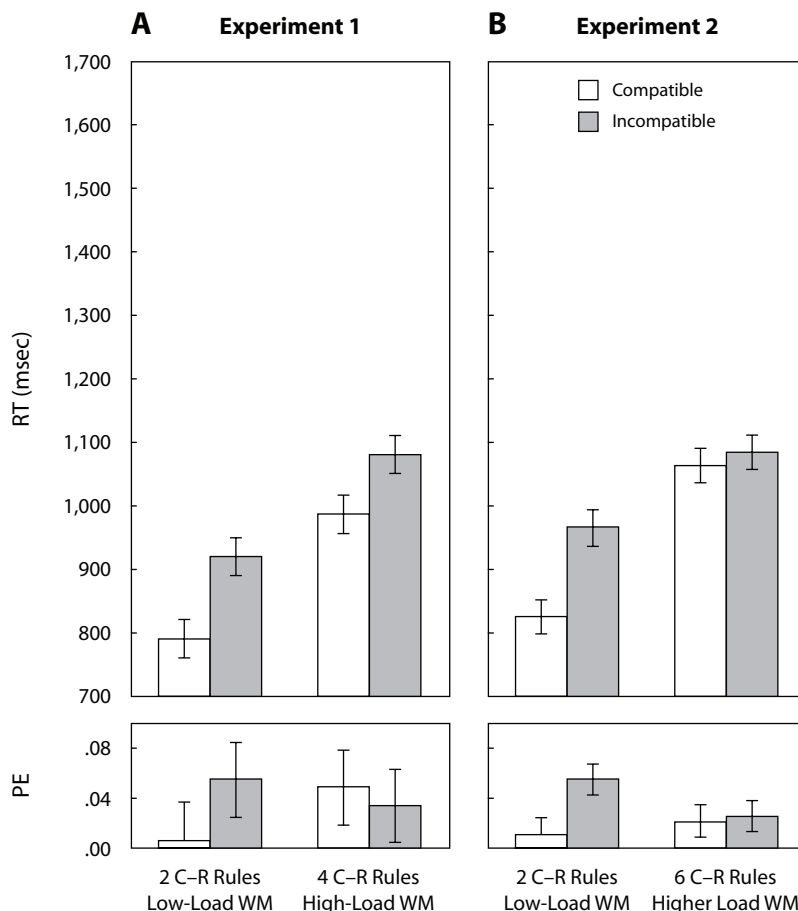


Figure 1. Mean reaction times (RTs) and proportions of errors (PEs) for the primary task according to working memory (WM) load in Experiment 1 (A) and Experiment 2 (B). The data come from the two index finger responses—that is, only from compatible and incompatible trials. Error bars indicate a 95% confidence interval computed for the contrast examining the compatibility effect. C, category; R, response.

11,523.47], indicating slower responses with secondary (1,351 msec) versus primary (945 msec) responses, with four (1,236 msec) versus two (1,060 msec) C1–R1 rules and with an incompatible (1,210 msec) versus compatible (1,086 msec) relationship between C1 and R2. The effect of alternatives validates our WM load manipulation. Also, the interaction between response and compatibility was significant [$F(1,24) = 6.01$, $MS_e = 619.14$], indicating larger compatibility effects on secondary (136 msec) versus primary (112 msec) responses. More important, the interaction between alternatives and compatibility was far from significance, as was the three-way interaction (both F s < 1).

A parallel ANOVA on the PE data indicated a main effect for compatibility [$F(1,24) = 5.83$, $MS_e = .001$] and a three-way interaction [$F(1,24) = 8.49$, $MS_e = .001$]. Follow-up planned contrasts showed that, for the primary responses, there was a significant simple interaction between compatibility and alternatives [$F(1,24) = 7.99$, $MS_e = .002$]. There was a significant simple–simple BCE in the low-load condition [$F(1,24) = 9.37$, $MS_e = .002$], but not in the high-load condition ($F < 1$). In contrast, for the secondary response, the simple interaction between compatibility and alternatives was nonsignificant ($F < 1$).

This experiment replicates Hommel and Eglau's (2002) RT results. Although the BCE we observed (around 140 msec) was considerably larger than that observed by Hommel and Eglau (56 msec), increasing the WM load from two to four Task 1 rules affected RT in the expected direction but did not influence the RT-BCE. However, the PE-BCE was absent for the high-load condition, already hinting that WM load is influential.²

EXPERIMENT 2

This experiment was identical to Experiment 1, except that we used a higher WM load condition of six Task 1 rules, rather than four.

Method

Participants. Twenty-six undergraduate students with attributes similar to those of the participants in Experiment 1 took part in this experiment.

Apparatus, Stimuli, and Procedure. The only difference, relative to Experiment 1, was that the high-load condition involved six Task 1 rules (higher load condition). The stimuli in the higher load condition were presented in red, green, yellow, blue, white, or pink color (S1). The manual responses in this condition were performed by pressing the *f* and *h* keys with the index fingers to indicate that the stimulus was red or green, the adjacent *d* and *j* keys to indicate that the stimulus was yellow or blue (using the middle fingers), and the *k* and *s* keys to indicate that the stimulus was white or pink (using the ring fingers). Here, each block consisted of 20 trials in the low-load condition (2 colors \times 2 letters \times 5 replications) or 60 trials in the higher load condition (6 colors \times 2 letters \times 5 replications), for a total of 80 or 240 experimental trials, respectively.

Results and Discussion

The data were treated as in Experiment 1. As before, all the analyses are based exclusively on the two index finger responses and, hence, on conditions with stimulus colors red and green. As before, premature Task 1 responses (RT1 < 150 msec, 0.32%), Task 1 response omissions

(RT1 > 2 sec, 0.82%; see note 1), and R1 and/or R2 error trials (3.59%) were excluded. Responses performed in the opposite order (R2–R1) were excluded from the RT analyses and were considered as errors. The group mean RT and PE for Task 1 are presented in Figure 1B.

The ANOVA on RT, with response and compatibility as within-participants independent variables and alternatives as a between-participants independent variable, yielded three significant main effects: response [$F(1,24) = 599.22$, $MS_e = 6,838.60$], alternatives [$F(1,24) = 9.12$, $MS_e = 128,342.80$], and compatibility [$F(1,24) = 18.71$, $MS_e = 9,739.50$], indicating slower responses with secondary (1,384 msec) versus primary (987 msec) responses, with six (1,291 msec) versus two (1,079 msec) C1–R1 rules, and with an incompatible (1,227 msec) versus a compatible (1,143 msec) relationship between C1 and R2. The effect of alternatives validates our WM load manipulation. More important, the interaction between alternatives and compatibility was significant [$F(1,24) = 12.85$, $MS_e = 9,739.50$]. There was a significant simple compatibility effect of 153 msec in the low-load condition [$F(1,24) = 31.29$, $MS_e = 9,739.50$] and a nonsignificant simple compatibility effect of 14 msec in the higher load condition ($F < 1$). Also, the three-way interaction was significant [$F(1,24) = 4.31$, $MS_e = 514.40$]. There was a significant simple interaction between alternatives and compatibility for both responses [$F(1,24) = 10.06$ and 14.41 , $MS_e = 4,678.40$ and $5,575.50$, for primary and secondary responses, respectively]. In the primary response, the BCE was reduced from 142 msec in the low-load condition [$F(1,24) = 27.99$, $MS_e = 4,678.40$] to 22 msec in the higher load condition ($F < 1$). In the secondary response, the compatibility effect was reduced from 164 msec in the low-load condition [$F(1,24) = 31.46$, $MS_e = 5,575.50$] to 7 msec in the higher load condition ($F < 1$).

The PE analysis indicated a significant main effect of compatibility [$F(1,24) = 14.49$, $MS_e = .001$] and a significant three-way interaction [$F(1,24) = 6.42$, $MS_e = .001$]. Follow-up planned contrasts showed that, for the primary responses, there was a simple interaction between compatibility and alternatives [$F(1,24) = 5.35$, $MS_e = .001$]. It resulted from a significant simple–simple BCE in the low-load condition [$F(1,24) = 12.87$, $MS_e = .001$] and a nonsignificant simple–simple BCE in the higher load condition ($F < 1$). In contrast, for the secondary response, the simple interaction between compatibility and alternatives was nonsignificant ($F < 1$).

The present results, showing that BCE was reduced to nonsignificant levels by higher WM load, support the notion that holding multiple task rules in WM results in parallel response activation.

EXPERIMENT 3

In the two previous experiments, we loaded WM by increasing the number of primary task rules but, concomitantly, we also increased the number of Task 1 stimuli. It is therefore conceivable that increasing the number of Task 1 stimuli was responsible for the eliminated BCE, rather than the WM load being responsible, as we have argued. One possibility is that increasing the number of

primary task stimuli prolonged the stimulus discrimination stage of Task 1 and postponed the response activation stage (Hommel, 1998), thereby reducing the temporal overlap between the two tasks. Another possibility is that, in this particular paradigm, Task 1 stimuli prime Task 2 responses (Lien, Ruthruff, Hsieh, & Yu, 2007) and this priming becomes weaker when fewer of the Task 1 stimuli overlap with Task 2 responses, as took place in the high-load and higher load conditions.

Aside from ruling out this alternative explanation, Experiment 3 was also run in order to clarify the nature of the task rules that give rise to the BCE. One hypothesis, the *S-R rule hypothesis*, states that the rules link each stimulus directly to its response. Another hypothesis, the *C-R rule hypothesis*, is that the stimuli are linked to abstract response categories and these response categories are linked, in turn, to the responses (see Logan, 1990; Meiran & Kessler, 2008; Pashler & Baylis, 1991; Schuch & Koch, 2003). The load manipulations that were used in Experiments 1 and 2 could not discriminate between these options because each concrete stimulus (color) was linked to a unique response category. Nonetheless, the literature strongly favors the C-R rule hypothesis. For example, Logan and Schulkind (2000) used parity and magnitude judgments of eight digits. If each digit was paired with the response through a separate link, as the S-R hypothesis states, the load was such that no BCE should have been observed (due to the enormous WM load: 8 S-R rules, in this case), but this effect was found nonetheless. Moreover, the participants in Logan and Gordon's (2001) Experiment 3 were asked to classify 96 pictures and 96 words into two categories. If each of the 192 stimuli was linked to its response through a unique link, this would most certainly exceed the capacity of WM by any estimate. The C-R hypothesis can account for these results by assuming that each stimulus was linked to its category through overlearned associations in long-term memory (LTM) and that WM held only the novel mapping of categories to responses.

There was an additional, more technical motivation for running Experiment 3, having to do with the fact that the participants were not assigned to experiment (1 vs. 2) on a random basis, and our primary finding concerns the difference between these two experiments. There was a hint that participant differences contributed to the effects in the lack of significant difference in the mean level of performance between the high-load condition (Experiment 1) and the higher load condition (Experiment 2), although a trend was found. Such a lack of significant difference is unexpected, given our claim that the WM load was influenced.

In this experiment, we included the three previous conditions (low, high, and higher WM load) and included an additional condition. In the low-load condition (henceforth, 2C2S, for 2 categories and 2 stimuli), the participants had two C1-R1 rules with two colors. In the high-load condition (4C4S), the participants had four C1-R1 rules with four colors. In the higher load condition (6C6S), the participants had six C1-R1 rules with six colors. In all these groups, the number of C-R rules was confounded with the number of stimuli (and potential S-R rules). The added condition,

2C6S, had only two C1-R1 rules involving six stimuli that were the same as those used in the higher load condition with two C-R rules that were the same as those used in the low-load condition. To create the task, the six colors in this condition were mapped to two categories: pink-related colors (red, white, and pink) and green-related colors (yellow, blue, and green). All the groups had the same secondary task (letter identity) requiring a vocal response, the words "pink" and "green" in Hebrew.

We predicted that we would replicate the results of Experiments 1 and 2 in the first three groups, this time also with the expected effects of WM load on mean RT that were missing before, presumably due to participant differences. Moreover, the last group enabled us to determine whether the number of C-R rules or the number of S-R rules (and stimuli) is critical. If the number of C-R rules is critical, the 2C6S group should produce results similar to those of the 2C2S group. If the number of stimuli (and S-R rules) matters most, the 2C6S condition should produce results that are most similar to those for the 6C6S condition.

Moreover, if the BCE resulted from the Task 2 rules that were held in WM, RT₂ should be more strongly delayed by primary task performance when Task 2 rules are not held in WM—that is, when BCE is not observed—since Task 2 rules need to be uploaded to WM in serial after Task 1 rules are uploaded. Specifically, we expected R₂ to be slower in the 6C6S condition relative to the remaining conditions and slower in condition 4C4S relative to the 2C2S and 2C6S conditions, which would not differ from one another. For the same reason, we also expected bigger RT differences between Task 1 and Task 2 responses—that is, longer interresponse intervals (IRIs)—in the 6C6S condition relative to the remaining conditions.

Method

Participants. Forty undergraduate students from Ben-Gurion University of the Negev participated in this experiment. The participants were assigned to one of the four loading conditions according to the order of entry.

Apparatus, Stimuli, and Procedure. This experiment was similar to Experiments 1 and 2, with the following exceptions. First, the vocal responses were changed to "varod" and "yarok" (the Hebrew words for "pink" and "green," respectively). The stimuli in the 2C2S condition were presented in pink or green color, the stimuli in the 4C4S condition were presented in pink, green, yellow, or blue color, and the stimuli in the 6C6S and 2C6S conditions were presented in pink, green, yellow, blue, white, or red color. The manual responses in the 2C2S and 2C6S conditions were pressing the *f* (left) and *h* (right) keys with the index fingers of both hands. The manual responses in the 4C4S condition were pressing the *f* and *h* keys with both hands in response to the pink or green stimuli (as in the low-load conditions) and, in addition, pressing *d* and *j* in response to the blue or yellow stimuli with the middle fingers of both hands. Finally, manual responses in the 6C6S condition were pressing the *f* and *h* keys with the index fingers to indicate that the stimulus was pink or green, the adjacent *d* and *j* keys to indicate that the stimulus was yellow or blue (using the middle fingers), and the *k* and *s* keys to indicate that the stimulus was white or red (using the ring fingers). In this experiment, each block was composed of 20 trials in the 2C2S condition (2 colors × 2 letters × 5 replications), 40 trials in the 4C4S condition (4 colors × 2 letters × 5 replications), and 60 trials in the 6C6S and 2C6S conditions (6 colors × 2 letters × 5 replications).³

Results and Discussion

As in the previous experiments, all the analyses are based exclusively on the two index finger responses (pink or pink-related colors vs. green or green-related colors). As previously, premature Task 1 responses ($RT_1 < 150$ msec, 0.71%), Task 1 response omissions ($RT_1 > 2$ sec, 1.80%; see note 1), R1 and/or R2 error trials (4.58%), and trials with potential response grouping ($IRI < 100$ msec, 15.83%) were excluded from data analyses. As before, responses in the opposite order were excluded from the RT analyses and were considered as errors. Mean RT and PE for Task 1 are presented in Figure 2.

The ANOVA on RT₁ yielded two significant main effects: group [$F(3,36) = 16.18$, $MS_e = 752,042.03$] and compatibility [$F(1,36) = 27.4$, $MS_e = 6,716.5$]. Responses were slower with an increasing number of C–R alternatives (878, 1,131, 1,315, and 975 msec for the 2C2S, 4C4S, 6C6S, and 2C6S conditions, respectively) and with an incompatible (1,118 msec) versus compatible (1,022 msec) relationship between C1 and R2. Planned comparisons revealed that all the group differences were significant except for the differences between Groups 2C2S and 2C6S [$F(1,36) = 1.34$, n.s.]. The significant

differences include the comparison between the 2C2S and 4C4S conditions [$F(1,36) = 13.74$], between the 4C4S and 2C6S conditions [$F(1,36) = 6.48$], between the 2C2S and 6C6S conditions [$F(1,36) = 40.89$], between the 6C6S and 2C6S conditions [$F(1,36) = 27.39$], and between the 4C4S and 6C6S conditions [$F(1,36) = 7.22$]. For all these comparisons, $MS_e = 4,646.98$. These results support the C–R hypothesis, which states that WM is loaded by C–R rules and not by S–R rules.

No less important, the interaction between group and compatibility was significant [$F(3,36) = 3.35$, $MS_e = 6,716.50$]. Follow-up planned comparisons showed significant simple BCEs of 176, 100, and 90 msec in the 2C2S, 4C4S, and 2C6S conditions, respectively [$F(1,36) = 23.06$, 8.21, and 6.02, respectively; all MS_e s = 6,716.50], and a nonsignificant simple BCE of 14 msec in the 6C6S condition ($F < 1$). Planned interaction contrasts compared the BCE between groups. This contrast was nonsignificant for the comparisons between the 2C2S and 4C4S conditions, between the 2C2S and 2C6S conditions, and between the 2C2S and 2C6S conditions [$F(1,36) = 2.76$, 1.87, and 0.08; n.s.]. Also, this contrast was significant for the comparisons between the 2C2S and 6C6S conditions between the 4C4S

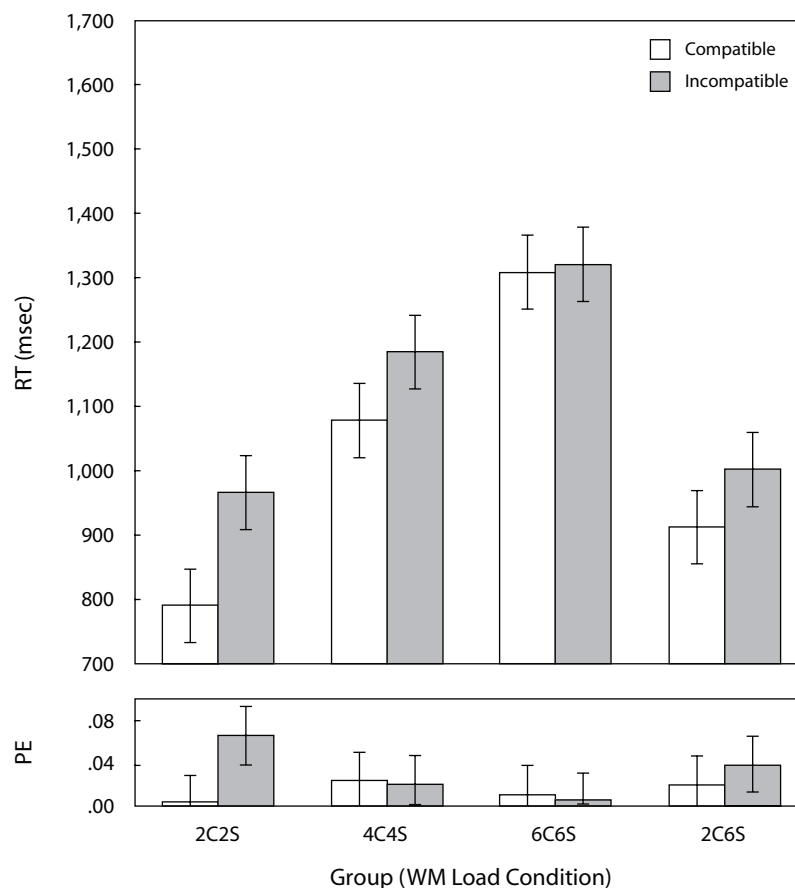


Figure 2. Mean reaction times (RTs) and proportions of errors (PEs) for the primary task according to group (working memory [WM] load) in Experiment 3. The data for all groups of WM condition come from the two index finger responses—that is, only from compatible and incompatible trials. Error bars indicate a 95% confidence interval computed for the contrast examining the compatibility effect.

and 6C6S conditions [$F(1,36) = 9.94$ and 3.18 , respectively] and was marginally significant for the comparison between the 2C6S and 6C6S conditions [$F(1,36) = 2.22$, $p = .083$; all $MS_{e,s} = 6,716.50$].

The PE analysis indicated a significant main effect of compatibility [$F(1,36) = 4.54$, $MS_e = .03$] and a significant interaction between compatibility and group [$F(3,36) = 3.45$, $MS_e = .02$]. Follow-up planned comparisons showed a simple BCE that was significant solely in the 2C2S condition [$F(1,36) = 13.53$, $MS_e = .0001$] and nonsignificant in the 4C4S, 6C6S, and 2C6S conditions [$F(1,36) = 0.03, 0.12$, and 1.23 , respectively; n.s.].

We also tested our predictions on RT2 to support our claims concerning Task 2 readiness in WM. The ANOVA revealed a significant effect for group [$F(3,36) = 22.44$, $MS_e = 49,923.75$]. As was expected, R2 was slower in the 6C6S conditions relative to the 4C4S, 2C2S, and 2C6S conditions [$F(1,36) = 16.16, 51.61$, and 48.25 , respectively; $MS_e = 49,923.75$], slower in the 4C4S condition than in the 2C2S and 2C6S conditions [$F(1,36) = 10.01$ and 8.56 , respectively; $MS_e = 49,923.75$], and nonsignificantly different in the 2C2S and 2C6S conditions ($F < 1$). These results support our hypothesis that Task 2 readiness is achieved by holding Task 2 C–R rules activated in WM.

The final comparisons concern the unique Task 2 slowing. According to response selection bottleneck theory (Pashler, 1994, 1998) and response deferment theory (Byrne & Anderson, 2001), R1 prolongation propagates into RT2 because the Task 2 processing stage at or after the bottleneck must wait until the processors are freed from Task 1 processing. If the group effect on RT2 merely reflects the group effects on RT1, one would predict that RT2 prolongation would be equivalent for all groups. Contrary to this expectation, we found a greater degree of R2 prolongation in the 6C6S condition, relative to the other conditions (see Figure 3). This result suggests that Task 2 was especially prolonged in this condition, in line with our reasoning that Task 2 rules were present in WM while Task 1 was performed in all the conditions except the 6C6S condition. Specifically, when RT2 and RT1 were entered into a single ANOVA with response (1 vs. 2) and group as independent variables, the interaction was statistically significant [$F(3,36) = 3.98$, $MS_e = 8,184.92$]. We conducted a series of follow-up planned interaction contrasts between contrasts comparing group pairs and response. All the comparisons not involving Group 6C6S yielded nonsignificant differences [$F(1,36) = 0.52, 2.36$, and 0.66 for the comparisons between Groups 2C2S and 4C4S, between Groups 2C2S and 2C6S, and between Groups 4C4S and 2C6S, respectively; all $MS_{e,s} = 8,184.92$]. In contrast, all the comparisons involving Group 6C6S were significant [$F(1,36) = 3.14, 6.22$, and 10.96 for the comparisons with Groups 2C2S, 4C4S, and 2C6S, respectively; all $MS_{e,s} = 8,184.92$].

As in Experiments 1 and 2, we found that the BCE was reduced to nonsignificant levels in the higher load (6C6S) condition. We also refuted the two alternative explanations concerning prolonged stimulus discrimination and priming of Task 2 rules by Task 1 stimuli, because

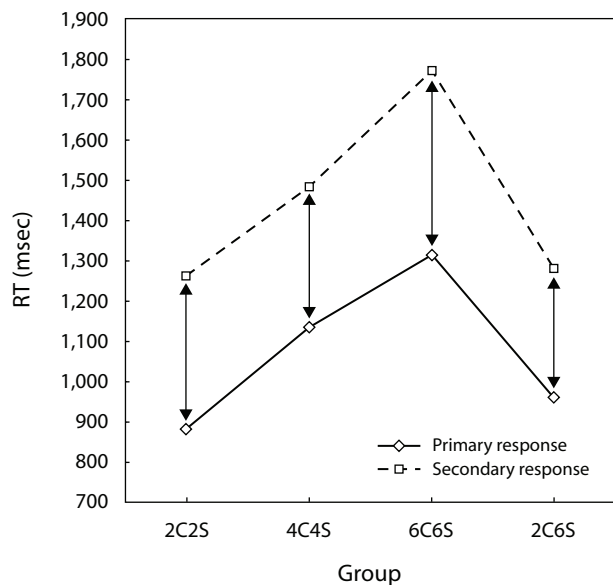


Figure 3. Mean reaction times (RTs) for primary and secondary tasks according to group (working memory [WM] load) in Experiment 3.

the 2C6S condition yielded a significant BCE in spite of being equivalent to the 6C6S condition, which was identical to it in these respects and yielded a negligible and nonsignificant BCE. Finally and most important, the present results clearly show that WM holds C–R rules and not S–R rules, because the number of C–R rules determined the overall RT and the BCE, and not the number of stimuli or S–R rules. The present results support again the notion that holding multiple rules for both tasks in WM results in parallel response activation. When WM is loaded by many Task 1 rules, Task 2 rules cannot be held in preparation in WM while Task 1 is executed, and Task 2 responses are prolonged while the BCE is eliminated.

PRACTICE EFFECTS

Because the direct link model assumes that repeatedly used task rules are held in long(er) term memory, this model predicts that the BCE will increase with practice. We therefore examined how practice (block) affected the BCE. As can be seen in Figures 4A, 4B, and 4C, there was a trend for an increase in the BCE only in the low-load condition. To examine whether the increase in BCE was significant, we computed a series of planned simple-interaction contrasts between compatibility and the contrasts that compared adjacent blocks only in the low-load conditions—that is, in the low-load condition of Experiments 1 and 2 and in Conditions 2C2S and 2C6S in Experiment 3. In all the experiments, all the comparisons were nonsignificant. However, this might be due to the relatively small increases in the effect from block to block. Therefore, we ran a series of similar contrasts that compared the first and the last blocks separately in each experiment. Here, the BCE was significantly larger in the last block, relative to the first block, in

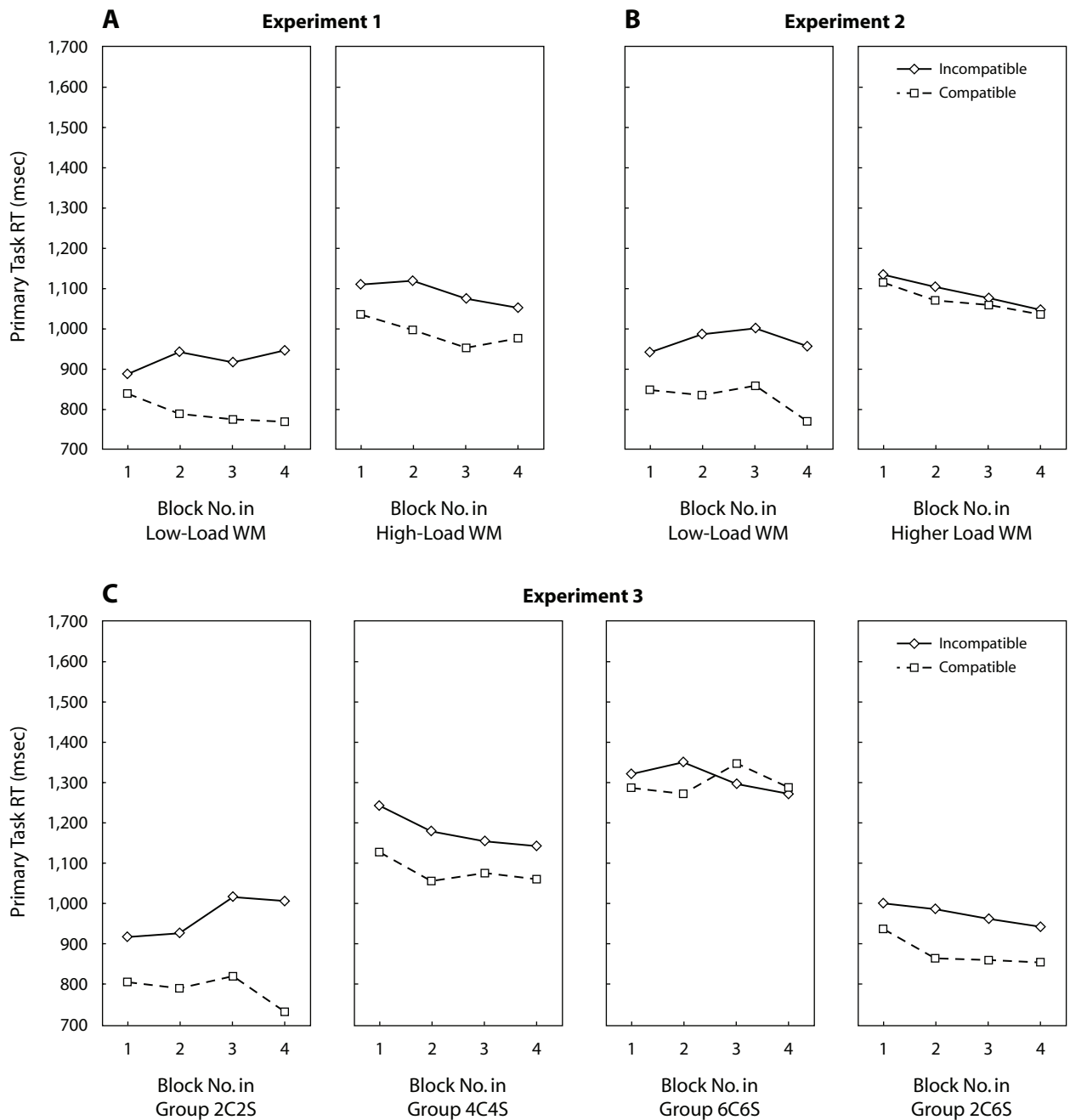


Figure 4. Mean reaction times (RTs) for primary task according to block, compatibility, and working memory (WM) load in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C).

the low-load WM condition in Experiment 1 [$F(1,24) = 6.18$, $MS_e = 8,059.19$], in the low-load WM condition in Experiment 2 [$F(1,24) = 4.48$, $MS_e = 6,443.02$], and in the 2C2S condition in Experiment 3 [$F(1,24) = 10.90$, $MS_e = 6,141.84$]. In contrast, this comparison was nonsignificant in the 2C6S condition in Experiment 3 ($F < 1$).

Two alternative (not necessarily mutually exclusive) explanations are offered for the results above. First, after practice, responses can be activated by direct S-R links outside WM, in addition to being activated from rules held

in WM. Furthermore, the high and the higher WM loads may have prevented retrieving these direct links from LTM. Alternatively, the direct links are held in rule complexes, so that when the rule complex is retrieved, it supports parallel response activation. However, for these complexes to be placed in LTM, they must first be simultaneously held in WM (see Jiménez & Méndez, 1999, for a related idea). Accordingly, in the high-load and higher load conditions, it was impossible to hold all the rules simultaneously active, which prevented the formation of rule complexes.

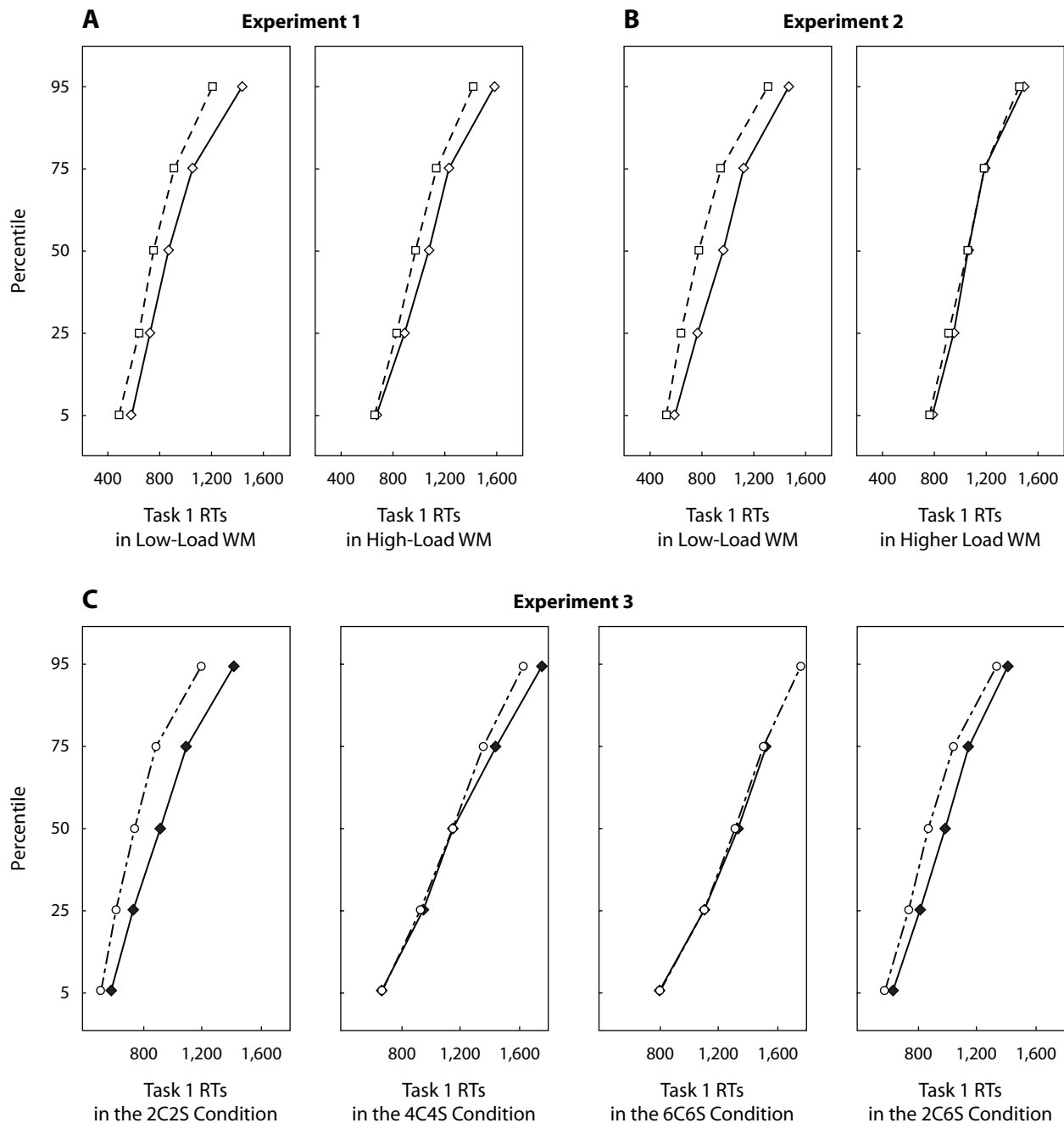


Figure 5. Primary task mean reaction time (RT) percentiles as a function of working memory (WM) load and compatibility in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C).

DISTRIBUTIONAL ANALYSES

Our results show that the BCEs are eliminated in conditions with longer RTs. One could argue that it is response prolongation per se that caused the elimination of the BCE, and not the WM load. This account is unlikely to be correct because Hommel (1998) observed increasing BCEs with longer RTs in all of his experiments. We nonetheless wanted to rule out this account of our own data. In order to address this problem, we performed a Vincentiz-

ing procedure (Ratcliff, 1979) on the RT1 data. For each participant and cell in the design, the 5th, 25th, 50th, 75th, and 95th RT percentiles were computed. As can be seen in Figure 5, the BCE was absent throughout the distribution in the higher load conditions of Experiments 2 and 3. Moreover, in conditions in which the BCE was present in the means, it increased, rather than decreased, with relative response speed, replicating the results in Hommel (1998). These results fully rule out slowing as an account for our core finding.

GENERAL DISCUSSION

In the present article, we examined the hypothesis that the BCE results from holding Task 2 rules in WM while performing Task 1. To this end, we followed up on an earlier article by Hommel and Eglau (2002), who rejected this hypothesis. In Experiment 1, we were able to replicate their findings by showing that the BCE in RT was statistically the same for conditions involving two and four task rules. This was critical in showing that the Experiment 2 results were not merely a replication failure. Using a higher WM load condition with six task rules in Experiment 2 reduced the BCE to nonsignificant levels. Finally, in Experiment 3, we replicated these results in a single experiment, ruled out some alternative explanations, and most important, showed that WM holds C–R rules, rather than S–R rules. These results suggest that the effect of WM load may not be monotonic and that it shows up only when extreme loads are used, so that this limited capacity system is pushed to its limits.

Our results confirm our hypothesis and suggest that Task 2 readiness is achieved by holding its rules in WM. This idea is in line with the *prepared reflex* metaphor suggested by Woodworth (1938) and revived by Hommel (2000) and Logan (1978; see also Kunde, Kiesel, & Hoffmann, 2003). In modern terms (see Cohen-Kadosh & Meiran, 2007), the idea of a prepared reflex is that, once the task rules (or *trigger conditions*) are placed in WM, stimulus information is translated more or less autonomously into the corresponding response activation. The readiness of the rules expedites response activation, but serial response selection still needs to be done (Hommel, 1998).

An interesting implication of our (confirmed) hypothesis is that Task 2 rules can be activated while Task 1 is performed, as long as there is enough WM capacity to hold both tasks sets activated. If so, a major challenge in dual-task situations is ensuring that Task 2 will not be executed before Task 1, in line with recent theorizing on dual-task performance emphasizing the contribution of task order control to performance (Logan & Gordon, 2001; Luria & Meiran, 2003, 2005, 2006; Meyer & Kieras, 1997a, 1997b).

Implications for Dual-Task Theories

The present results challenge some current theories of dual-task performance. Basically, the existence of the BCE challenges theories assuming that response activation is serial (Byrne & Anderson, 2001) and theories assuming that there is only serial response selection (e.g., Pashler, 1984; Ruthruff, Pashler, & Hazeltine, 2003; Ruthruff, Pashler, & Klaassen, 2001). This claim, however, is not new (cf. Miller & Alderton, 2006; Watter & Logan, 2006) and not entirely conclusive, because response activation, and not response selection, may be carried out in parallel (e.g., Hommel, 1998). The novel aspect of our results is that the BCE results from rules that operate inside WM. This finding is not predicted by theories that assume that task rules reside outside WM (Meyer & Kieras, 1997a, 1997b) or theories that assume that the BCE originates in the task execution modules, which also reside outside WM (Logan, 2002; Logan & Gordon, 2001). Although the result is not predicted by

these theories, it does not refute the theories. The reason is that the extreme WM load manipulation may have signaled increased control demands that, in turn, resulted in a shift to a more serial processing (Logan & Gordon, 2001; Luria & Meiran, 2005). Alternatively, Logan and Gordon's model could potentially explain our effects by the increased load on the number of bias parameters that are simultaneously held in WM, although this remains to be shown.

We suggest that the present results are most naturally explained by capacity-sharing theories (Navon & Miller, 2002; Tombu & Jolicoeur, 2003), in which a fixed, limited capacity for response selection is shared between the tasks. It is assumed that the capacity is devoted to the activation of task rules within WM. From this point of view, when WM is not overloaded, there is sufficient spare capacity to activate Task 2 rules while Task 1 is performed. As a result, Task 2 responses are activated (but not selected and prepared yet), giving rise to the BCE. When the WM load is overloaded, there is no spare capacity, and the entire capacity is devoted to the activation of Task 1 rules when this task is performed. As a result, the BCE is eliminated in this condition.

Methodological Implications for Testing WM Load Effects

As in previous studies (see, e.g., Baddeley & Hitch's [1974] classic study), we showed that the influence of WM load on performance may not be monotonic and may show up only with extreme loads. Therefore, in order to observe performance decrements due to WM load, one needs to push the system to its limits.

AUTHOR NOTE

This research was supported by a research grant from the Israel Science Foundation, given to N.M. We thank Eric Ruthruff, Bernhard Hommel, Gordon Logan, and Roy Luria for helpful suggestions. Correspondence concerning this article should be sent to R. Ellenbogen, Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: ravidel@bgu.ac.il).

REFERENCES

- ACH, N. (2006). *On volition* (T. Herz, Trans.). Retrieved September, 2006, from University of Konstanz, Cognitive Psychology Web site: www.uni-konstanz.de/kogpsych/ach.htm. (Original work published 1910)
- BADDELEY, A. [D.] (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- BYRNE, M. D., & ANDERSON, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review*, *108*, 847-869.
- CAESSENS, B., HOMMEL, B., REYNVOET, B., & VAN DER GOTEN, K. (2004). Backward-compatibility effects with irrelevant stimulus-response overlap: The case of the SNARC effect. *Journal of General Psychology*, *131*, 411-425.
- COHEN-KADOSH, O., & MEIRAN, N. (2007). The representation of instructions in working memory leads to autonomous response activation: Evidence from the first trials in the flanker paradigm. *Quarterly Journal of Experimental Psychology*, *60*, 1140-1154.
- COWAN, N. (1998). Evidence against the global speed of processing theory of working memory. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (p. 1211). Mahwah, NJ: Erlbaum.

- HOMMEL, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1368–1384.
- HOMMEL, B. (2000). The prepared reflex: Automaticity and control in stimulus–response translation. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 247–273). Cambridge, MA: MIT Press.
- HOMMEL, B., & EGLAU, B. (2002). Control of stimulus–response translation in dual-task performance. *Psychological Research*, **66**, 260–273.
- HÜBNER, R., & LEHLE, C. (2007). Strategies of flanker coprocessing in single and dual tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 103–123.
- JIMÉNEZ, L., & MÉNDEZ, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 236–259.
- KUNDE, W., KIESEL, A., & HOFFMANN, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, **88**, 223–242.
- LIEN, M.-C., & PROCTOR, R. W. (2000). Multiple spatial correspondence effects on dual-task performance. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1260–1280.
- LIEN, M.-C., RUTHRUFF, E., HSIEH, S., & YU, Y.-T. (2007). Parallel central processing between tasks: Evidence from lateralized readiness potentials. *Psychonomic Bulletin & Review*, **14**, 133–141.
- LIEN, M.-C., SCHWEICKERT, R., & PROCTOR, R. W. (2003). Task switching and response correspondence in the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 692–712.
- LOGAN, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, **107**, 32–63.
- LOGAN, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, **22**, 1–35.
- LOGAN, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, **109**, 376–400.
- LOGAN, G. D., & DELHEIMER, J. A. (2001). Parallel memory retrieval in dual-task situations: II. Episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 668–685.
- LOGAN, G. D., & GORDON, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, **108**, 393–434.
- LOGAN, G. D., & SCHULKIND, M. D. (2000). Parallel memory retrieval in dual-task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1072–1090.
- LURIA, R., & MEIRAN, N. (2003). Online order control in the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 556–574.
- LURIA, R., & MEIRAN, N. (2005). Increased control demand results in serial processing: Evidence from dual-task performance. *Psychological Science*, **16**, 833–840.
- LURIA, R., & MEIRAN, N. (2006). Dual route for subtask order control: Evidence from the psychological refractory paradigm. *Quarterly Journal of Experimental Psychology*, **59**, 720–744.
- MEIRAN, N., & KESSLER, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, **34**, 137–157.
- MEYER, D. E., & KIERAS, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, **104**, 3–65.
- MEYER, D. E., & KIERAS, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, **104**, 749–791.
- MILLER, J. (2006). Backward crosstalk effects in psychological refractory period paradigms: Effects of second-task response types on first-task response latencies. *Psychological Research*, **70**, 484–493.
- MILLER, J., & ALDERTON, M. (2006). Backward response-level crosstalk in the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 149–165.
- NAVON, D., & MILLER, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, **44**, 193–251.
- PASHLER, H. [E.] (1984). Evidence against late selection: Stimulus quality effects in previewed displays. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 429–448.
- PASHLER, H. [E.] (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220–244.
- PASHLER, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.
- PASHLER, H. [E.], & BAYLIS, G. (1991). Procedural learning: 1. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 20–32.
- PASHLER, H. [E.], & JOHNSTON, J. C. (1998). Attentional limitations in dual-task performance. In H. [E.] Pashler (Ed.), *Attention* (pp. 155–189). Hove, U.K.: Psychology Press.
- RATCLIFF, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.
- RUTHRUFF, E., PASHLER, H. E., & HAZELTINE, E. (2003). Dual-task interference with equal task emphasis: Graded capacity sharing or central postponement? *Perception & Psychophysics*, **65**, 801–816.
- RUTHRUFF, E., PASHLER, H. E., & KLAASSEN, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, **8**, 73–80.
- SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- SCHUCH, S., & KOCH, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 92–105.
- TOMBU, M., & JOLICŒUR, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 3–18.
- WATTER, S., & LOGAN, G. D. (2006). Parallel response selection in dual-task situations. *Perception & Psychophysics*, **68**, 254–277.
- WOODWORTH, R. S. (1938). *Experimental psychology*. New York: Holt.

NOTES

1. This cutoff is relatively aggressive and potentially influences factor effects. To check for this possibility, we ran all the analyses without this cutoff, meaning that the longest RT could be 2.5 sec (the response deadline that we used). The pattern of significant effects was not influenced by this change in the analyses in any of the experiments.

2. For exploratory reasons, we took advantage of the fact that there were S1 colors not having any compatibility relationship and used these responses as a baseline against which the compatible and the incompatible conditions were compared. These analyses were conducted separately for each experiment. We are aware of the fact that this comparison is problematic because condition is confounded with both color and responding finger.

In Experiment 1, trials with blue or yellow stimuli served as the baseline, whereas blue, yellow, white, and pink stimuli served as the baseline in Experiment 2. In all the experiments, baseline RTs were shorter than both incompatible and compatible RTs. Therefore, we adopted a stringent criterion and compared the baseline RTs with the (quicker) compatible RTs. The mean RTs were 987 msec (compatible) versus 929 msec (baseline) in Experiment 1 and 1,066 msec versus 983 msec, respectively, in Experiment 2 [$F(1,12) = 3.99$ and 8.78 , $MS_e = 5,594.94$ and $5,158.91$, respectively]. The results of this analysis are surprising in that the baseline did not fall between the compatible and the incompatible conditions, as one might expect. A potential explanation for this result is that there was less code conflict in the neutral trials because the color codes were different from the response codes of Task 2. Such code conflict should affect performance, according to Logan and Gordon (2001).

3. The number of trials per block differed between the conditions for the reasons specified in the Method section. In order to show that this difference was not a contributing factor to our primary finding, we conducted an analysis on the first 20 trials of the blocks, thus equating the conditions in this respect. This analysis produced the same pattern of significant results as did the core analysis. Specifically, the planned comparisons with an alpha of .05 showed significant BCEs of 179, 98, and 102 msec in the 2C2S, 4C4S, and 2C6S conditions, respectively [$F(1,29) = 14.15$, 3.82 , and 3.58 , respectively] and a nonsignificant BCE of -58 msec in the 6C6S condition ($F < 1$; all MS_e s = $45,626.9$).