

Dual-task backward compatibility effects are episodically mediated

Maria Giammarco¹ · Sandra J. Thomson² · Scott Watter¹

Published online: 16 November 2015
© The Psychonomic Society, Inc. 2015

Abstract In dual-task performance, the backward compatibility effect (BCE; faster Task 1 reaction time when Task 1 and Task 2 responses are compatible) is thought to represent automatic activation of Task 2 response information in parallel with attended Task 1 performance. Work by Hommel and Eglau (*Psychological Research*, 66, 260–273, 2002) has suggested the BCE relies on stimulus-response learning in long-term memory. Subsequent work by Ellenbogen and Meiran (*Memory and Cognition*, 36, 968–978, 2008), however, proposed that the BCE is mediated by Task 2 rules held in working memory (WM) during Task 1 performance. The present study aimed to dissociate these two theoretical claims. In Experiment 1, we assessed the effects of prior single-task practice with Task 1 or Task 2 of a subsequent dual-task paradigm. Where the WM-mediated model predicts both BCE and overall reaction time improvement relative to prior task practice, an episodic learning model makes divergent predictions for BCE based on the context specificity of prior Task 2 learning. Results showed a close fit with episodic predictions and contradicted WM model predictions. Experiment 2 examined the finer grained timecourse of BCE over initial

development, subsequent interference of this initial learning on BCE development with new conflicting Task 2 response mappings, and finally reestablishment of BCE in the original dual task. Data again showed close agreement with long-term learning predictions. We argue in favor of an episodic account of the BCE, and consider implications of WM and episodic mechanisms of automatic response activation on other aspects of dual-task performance.

Keywords Psychological refractory period · Dual-task processing · Backward response compatibility · Divided attention · Episodic memory · Working memory · Practice effects

It is well established that performance on a given task is impaired when it is performed concurrently with a second task compared to when it is performed on its own (Pashler, 1994; Telford, 1931; Welford, 1952). The psychological refractory period (PRP) paradigm has become a popular method for studying the fine-grained timecourse of dual-task performance. Here, the stimuli for two tasks are presented, separated in time by a variable stimulus onset asynchrony (SOA), and response times for both tasks are evaluated. In a wide range of studies, Task 1 is seen to be relatively unaffected by SOA, but Task 2 reaction times (RT) increase as SOA decreases and tasks overlap more in time. In cognitive psychology, this dual-task interference has classically been thought to be the result of a structural bottleneck in central information processing, typically identified as response selection, that prevents two tasks from being fluently processed at the same time (Pashler, 1994; Pashler & Johnston, 1998).

Many studies support the general notion of a strict bottleneck in central response selection. For example, studies have demonstrated that dual-task interference persists after

✉ Maria Giammarco
mgiammarco13@gmail.com

✉ Sandra J. Thomson
sthomson@stu.ca

✉ Scott Watter
watter@mcmaster.ca

¹ Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

² Department of Psychology, St. Thomas University, 51 Dineen Drive, Fredericton, New Brunswick E3B 5G3, Canada

substantial amounts of dual-task training (Ruthruff, Johnston, & Van Selst, 2001; Tombu & Jolicoeur, 2004; Van Selst, Ruthruff, & Johnston, 1999), when participants are instructed to place equal emphasis on both tasks (Ruthruff, Pashler, & Hazeltine, 2003; Ruthruff, Pashler, & Klaassen, 2001), and even after strong incentives for parallel performance (Ruthruff, Johnston, & Remington, 2009).

The backward compatibility effect

A number of studies have shown evidence that despite persistent overall Task 2 RT costs, response information for Task 2 can be generated in parallel with focused Task 1 performance in these dual-task situations (Ellenbogen & Meiran, 2008; Hommel, 1998; Hommel & Eglau, 2002; Lien & Proctor, 2000; Logan, Miller, & Strayer, 2011; Miller, 2006; Miller & Alderton, 2006; Thomson & Watter, 2013; Thomson, Watter, & Finkelshtein, 2010; Watter & Logan, 2006). These effects are observed as compatibility effects on Task 1 reaction time when meaningful correspondences exist between the Task 2 response and the Task 1 response or between the Task 2 response and the Task 1 stimulus. These effects have collectively become known as “backward compatibility effects” (BCEs) and have been suggested by a number of authors to influence Task 1 central response selection processing rather than later postbottleneck motor processing in Task 1 (e.g., Hommel, 1998; Thomson & Watter, 2013; Thomson et al., 2010). BCEs are most reliably observed in dual-task RT data at very short SOAs, where tasks overlap sufficiently to allow Task 2 response information to be generated in time to produce compatibility effects on Task 1 response selection.

The mechanism responsible for generating these backward compatibility effects has been the topic of a number of studies and is the focus of our present paper. In the first major investigation of BCEs, Hommel (1998) suggested two alternative mechanisms for Task 2 response activation in parallel with attended Task 1 performance. In the *transient-link* model, multiple S-R mapping rules are held online in working memory (WM), and these rules facilitate the automatic retrieval or activation of responses relative to presented Task 2 stimuli. In contrast, in the *direct-link* model, experience performing the task produces learning of S-R associations in long-term memory that automatically retrieve or activate responses relative to presented Task 2 stimuli. Hommel (1998) speculated that the transient-link model may be more likely, given BCEs were observed to develop relatively quickly with practice.

Hommel and Eglau (2002) subsequently performed a more thorough comparison of these alternative mechanisms by testing the effect of WM load on the BCE. Hommel and Eglau (2002) manipulated WM load by varying the number of task rules for a given task, and found consistent BCEs despite this varying WM load. In addition to a lack of WM load effects,

they demonstrated that once BCEs had developed in dual-task performance, they persisted when participants performed only Task 1 of the dual-task pair. In this situation, Task 2 response information was activated automatically in the presence of the Task 2 stimulus, even though participants no longer performed or intended to perform Task 2. Hommel and Eglau (2002) also demonstrated that previously learned Task 2 S-R associations could interfere with subsequent S-R mappings with a new Task 2 when these relationships overlapped or competed. From this set of findings, they made a strong case for a direct-link long-term memory account of backward compatibility effects. Hommel and Eglau (2002) noted that either a traditional S-R learning model account (strengthening of direct, permanent S-R associations; e.g., Thorndike, 1927) or an episodic memory account (accumulation of multiple traces of S-R episodes over trials; e.g., Logan, 1988) were both potential alternative models for a direct-link account of BCEs.

More recently, however, Ellenbogen and Meiran (2008) argued that backward compatibility effects are mediated via a WM-dependent transient-link-style model after all. Ellenbogen and Meiran concentrated on the WM capacity findings of Hommel and Eglau (2002) and suggested that this earlier work may have failed to detect an influence of WM because participants' WM capacity was not sufficiently taxed. Ellenbogen and Meiran first closely replicated Hommel and Eglau's Experiment 2, with the same null effects on BCEs, and then demonstrated that they could completely abolish BCEs by placing more extreme demands on WM. Across several experiments, Ellenbogen and Meiran demonstrated that when a large number of category-to-response mapping rules were required for Task 1, BCEs were not observed in their dual-task data. They attributed this finding to participants being unable to instantiate Task 2 mapping rules in WM along with the Task 1 rules during Task 1 performance in this high-demand condition, leading to a lack of transient-link-mediated Task 2 response activation.

Working memory versus episodic mechanisms of the BCE

Our goal in the present study was to critically examine direct-link versus transient-link accounts of backward compatibility effects. To achieve this, we need to carefully consider what predictions these alternative models make for BCEs and for related observable aspects of dual-task behavior. We consider three distinct models here: a WM-mediated transient-link model, as described by Ellenbogen and Meiran (2008); a simple S-R associative learning direct-link model; and an episodic direct-link model.

Ellenbogen and Meiran (2008) have described a WM-mediated transient-link model, and it is their particular version of this model we consider here. The critical assumption of

WM-mediated BCE is that Task 2 rules are instantiated in WM along with Task 1 rules during attended Task 1 performance, mediating automatic Task 2 response activation to produce BCEs. This model claims that, in dual-task performance, WM capacity is dedicated to rule representation; so long as WM capacity is not overloaded by representing rules for Task 1, participants will simultaneously represent Task 2 rules in WM during Task 1 performance in order to be better prepared for eventual Task 2 performance (Ellenbogen & Meiran, 2008). This model predicts potentially abrupt appearance and disappearance of BCEs, depending on WM instantiation of Task 2 rules. When participants stop performing one set of tasks and begin performing a new dual task combination (with different S-R mappings), the WM-mediated model predicts that BCEs should rapidly appear in the new task situation, to the extent that participants are able to adequately understand, represent, and perform the task. This model also predicts that if participants are not actively representing rules from a prior Task 2 in WM, then prior Task 2 S-R relationships should not show interference effects via current task BCEs.

We can distinguish two general kinds of direct-link models – a traditional learning account, where particular S-R associations are learned and strengthened over experience with multiple trials, and an episodic account, where multiple instances or episodes of S-R bindings are accrued over multiple trials. Both kinds of direct-link models predict increasing BCEs with continued practice of Task 2. A traditional learning account makes the simple prediction of increasing BCE dependent on the amount of prior Task 2 practice, generally independent of the context of prior Task 2 learning. The episodic version of a direct-link model predicts similar effects of practice given a consistent task context. In both models, BCEs are produced when Task 2 response information is automatically activated from a presented Task 2 stimulus, in parallel with attended Task 1 performance. These direct-link models do not depend on available WM capacity for Task 2 rule representation to elicit automatic S-R activation for Task 2 to produce BCEs. The acquisition of these S-R associations, however, does rely on WM indirectly, via eventual attended Task 2 performance within the dual-task setting. These direct-link models distinguish the acquisition of S-R associations for Task 2 (learned during deliberate attended performance of Task 2) from the automatic activation of Task 2 S-R associations when the Task 2 stimulus is present in parallel with deliberate Task 1 performance, observed as BCEs.

The episodic direct-link model

An episodic direct-link model makes the same predictions as a traditional direct-link learning model with practice within a consistent task context, but it also predicts that the context of prior Task 2 performance and learning will influence how

subsequent Task 2 automaticity, and hence BCEs, are expressed. More detailed and nuanced views of this kind of mechanism have been developed over the past decade, all of which point increasingly to the important role of contextual specificity in encoding and later retrieval for many aspects of speeded choice performance and related attentional and cognitive control components of this behavior (e.g., Crump & Milliken, 2009; Crump, Vaquero, & Milliken, 2008; Hommel, 2007, 2009; Jacoby, Lindsay, & Hessels, 2003; Waszak, Hommel, & Allport, 2003; see Hommel, 2004, for a brief overview). BCEs in the episodic model are assumed to reflect activation of response information from prior contextually sensitive Task 2 S-R episodes, in parallel with attended Task 1 performance. Similar to the traditional learning direct-link model described previously, the acquisition and learning of these Task 2 S-R episodes most likely comes from attended performance of Task 2 within the general dual-task or experimental context and not in parallel with attended Task 1 performance (see Jiménez & Méndez, 1999, for related work on learning of implicit associations in attended versus unattended situations). Learning of separate S-R episodes within a contextually sensitive long-term memory system should produce BCEs that are themselves sensitive to the context in which previous S-R episodes are acquired (with respect to how well context of prior learning matches the context in which BCEs may be expressed).

We can consider our predictions for an episodic direct-link model based on Hommel's (2007, 2009) general event file framework and the relative disruption to performance that occurs with a partial match of prior episodic content to current performance, versus a complete match or a complete mismatch. If a current dual-task context is consistent with the earlier context in which prior Task 2 S-R learning occurred (a complete match), automatic Task 2 S-R translation should be elicited and will drive BCEs. If there is no prior Task 2 experience, then there is no episodic support for a new Task 2 (a complete mismatch) – here we predict no BCEs initially, with the development of Task 2 automaticity and thus BCEs over subsequent dual-task practice. If there has been prior Task 2 experience in a different context than the current dual-task situation, it is possible that the partial match (and partial mismatch) to the current dual-task context will not elicit substantial Task 2 S-R automaticity, leading to a relative lack of BCEs, despite prior Task 2 acquisition of these S-R episodes. In addition, the conflict arising from these partial contextual matches may additionally interfere with the development of Task 2 automaticity driving BCEs in the new dual-task context. Observing a relative lack or suppression of BCEs given prior Task 2 practice in a different task context, compared to BCEs from prior context-matched Task 2 practice or even the absence of prior Task 2 practice, is a unique prediction of an episodic direct-link model of the BCE.

The present study

In order to distinguish these issues, we need to look for experimental predictions that will dissociate transient-link and direct-link accounts of backward compatibility effects and, if possible, test the unique predictions of an episodic direct-link model. We suggest two possible dissociations, and tested each in separate experiments. In Experiment 1, we investigated the effect that prior single-task practice with Task 1 and/or Task 2 had on subsequent BCEs and overall RTs in dual-task performance. This experiment directly tests predictions of an episodic direct-link account by examining the effect of prior Task 2 practice context on subsequent dual-task performance. In Experiment 2, we more closely examined direct-link model predictions of how BCEs should develop over time. To anticipate our findings, Experiment 1 demonstrated a pattern of data that fits episodic predictions and is not well explained by a WM-mediated transient-link model or a traditional S-R learning direct-link model. Experiment 2 produced a detailed pattern of interference effects on BCEs predicted by direct-link models and reinforces the episodic account suggested by Experiment 1. Finally, in the General Discussion, we consider our findings along with experiments from Ellenbogen and Meiran (2008) and Hommel and Eglau (2002), and other theoretical constraints on dual-task performance in the presence of parallel response activation.

Experiment 1

In Experiment 1, we examined the influence of previous single-task practice with Task 1 and/or Task 2 on subsequent dual-task performance, with respect to the presence and timecourse of development of backward compatibility effects as well as overall RT performance. We defined four between-participants experimental groups, all of which performed the same dual task in the latter half of their experimental session. Each group performed a different combination of single tasks in alternating blocks prior to dual-task performance. In the Practice-Both group, participants practiced Task 1 and Task 2 of the upcoming dual-task paradigm as single tasks; in the Practice-T2 group, participants practiced Task 2 and a different filler task as single tasks; in the Practice-T1 group, participants practiced Task 1 and a different filler task as single tasks; and in the Practice-None group, participants practiced two different filler tasks prior to dual-task performance.

For a WM-mediated account, Ellenbogen and Meiran (2008) argued that BCEs should be present in suitable dual-task situations to the extent that participants have available capacity to represent both Task 1 and Task 2 rules in WM during attended performance for Task 1. We suggest that prior to dual-task performance, single task practice on either Task 1 or Task 2 of the dual-task should generally improve a

participant's ability to more quickly and/or more strongly instantiate rules in WM for that task. With prior practice on Task 2, participants should be better prepared to instantiate Task 2 rules in WM along with Task 1 rules during Task 1 performance in a dual-task setting, predicting more reliable or robust BCEs versus no prior Task 2 practice. Prior practice on Task 1 might also produce a larger BCE in dual-task performance, as practice representing Task 1 rules may allow more WM capacity to be allocated to represent Task 2 rules. These predictions of prior practice on BCE effects are in the same direction as more general predictions on overall Task 2 RT effects – we predict that a previously practiced single task will show a relative RT benefit under dual-task performance compared to having not practiced that task.

A traditional learning direct-link model makes the simple prediction that prior Task 2 single-task practice should help to develop Task 2 S-R learning and automaticity, which should produce BCEs in later dual-task performance. Prior single-task practice of Task 1 may facilitate later general dual-task performance, including, possibly, more rapid development of Task 2 automaticity during dual-task performance, due to reduced costs of Task 1 performance allowing greater capacity or focus on learning the new Task 2. These predictions are similar to those from the WM-mediated transient-link model, though they are made on the basis of prior learning and automaticity of Task 2 S-R associations and do not require WM instantiation of Task 2 rules in parallel with Task 1 performance in order to produce BCEs.

In contrast, an episodic model predicts that the contextual details of prior Task 2 practice will be important predictors of BCEs in dual-task performance. If participants have had prior Task 2 single-task practice, the degree to which dual-task performance incorporates a similar general-task context compared to prior Task 2 practice should be important. Here, we predict that the Practice-Both condition should produce strong BCEs, where dual-task performance incorporates both prior interleaved single tasks; in turn, we might expect relatively reduced or nonexistent BCEs in the Practice-T2 condition, where dual-task performance incorporates this Task 2 with a new task as Task 1, relative to prior interleaved single task practice. With no prior Task 2 practice, the episodic direct-link model predicts the development of BCEs over time with a novel Task 2 in dual-task performance. This should be the case, independent of whether or not Task 1 was practiced – Task 2 automaticity should develop over the course of dual-task performance, with no interference from prior contextually mismatched Task 2 S-R learning. This predicted pattern of BCEs is distinct from the pattern of general enhancement to overall Task 1 and Task 2 RT in dual-task performance that we expect given prior single-task practice, and it is distinct from the predicted pattern of BCEs under both the WM-mediated transient-link model and a traditional learning direct-link model.

Method

Participants

One hundred and three participants (87 females, mean age = 18.6 years) were recruited from the McMaster University undergraduate population. They were all enrolled in psychology courses and received partial course credit for their participation. This study was approved by McMaster's Research Ethics Board, and all participants gave written informed consent prior to beginning the experiment. All participants had normal or corrected to normal visual acuity, and normal color vision.

Apparatus and stimuli

Stimuli were presented either on a 19-in. ViewSonic Professional Series P95f+ CRT monitor controlled by a Dell Dimension 4600 computer or a 21.5-in. Samsung SyncMaster B2240 LCD monitor controlled by an HP Pro 3130 computer, using Presentation software (www.neurobs.com). The stimuli were identical in physical size across the two monitors. Participants were seated approximately 60 cm from the computer monitor, and their responses were collected using a standard keyboard, the mouse, or the thumb joystick of a gamepad, depending on the task being completed.

Four basic tasks were used. For the *shape* task (dual-task Task 1), the stimuli were line drawings of a star, a diamond, a circle, and a pentagon, filled in white. The height and width of each shape was approximately 1.25 degrees of visual angle. For the *color* task (dual-task Task 2), the stimuli were filled squares presented in orange, yellow, blue, or purple, with height and width of 1.25 degrees of visual angle. The stimuli for the *case* task (one of two filler tasks) were four letters from the English alphabet (A, E, G, and R), presented either in uppercase or lowercase Helvetica font, scaled to approximately 1.25 degrees of visual angle. Finally, in the *size* task (the second filler task), the stimuli were eight five-letter nouns, four of which referred to items that were larger than the computer monitor (bench, stove, piano, canoe), and four that referred to items smaller than the computer monitor (cigar, pearl, badge, spoon), scaled to an approximate height of 1.25 degrees of visual angle. All stimuli other than color patches were presented in white on a black background; color patches were solid color, not outlined in white, to emphasize the color value and not the square shape of the color patch.

Design

An outline of the experimental design and procedure is shown in Fig. 1. All participants completed a variable single-task phase followed by a common dual-task phase. In the single-task phase, participants practiced two different single tasks,

alternating between these two tasks twice in counterbalanced order before commencing the dual-task phase.

The tasks used in the single-task phase consisted of some combination of the component tasks used in the dual-task phase (Task 1, Task 2) and/or two filler tasks that were not encountered in the dual-task phase. For initial single-task practice, participants in the Practice-Both group practiced *shape* and *color* tasks (Task 1 and Task 2 of the dual-task paradigm, respectively). Participants in the Practice-T2 group initially practiced *color* (Task 2) and *case* (filler) tasks. Participants in the Practice-T1 group initially practiced *shape* (Task 1) and *case* (filler) tasks. Participants in the Practice-None group initially practiced *case* and *size* tasks (both filler). The dual-task phase incorporated the *shape* and *color* tasks as Task 1 and Task 2, respectively, of a typical PRP paradigm.

For *shape* (Task 1) and *color* (Task 2) in both the single- and dual-task phases, responses were collected using the number pad of a standard computer keyboard. Responses to the filler tasks were made using either gamepad joystick movements (*case* task) or mouse movements (*size* task) in order to minimize response mapping overlap with the tasks used in the dual-task phase.

Procedure

Single task In the *shape* task, participants performed shape discrimination, pressing one response key if the shape was either a star or a diamond and a different response key if the shape was a circle or a pentagon. Trials began with a fixation display consisting of two white dashes in the center of the screen against a black background, flanking the position where the shape would appear. After 500 ms, the fixation display was replaced with the shape stimulus for 1,000 ms, followed by a blank screen for 1,500 ms, until the next trial began. Participants responded by pressing the 1 or 2 key on the number pad of a standard computer keyboard using the index or middle finger of the right hand. Response mapping was counterbalanced across participants. In the *color* task, participants judged the color of a filled square stimulus by pressing one response key if the color was yellow or orange and another if the color was purple or blue. The trial sequence and response keys were identical to those in the shape task.

In the *case* task, participants judged whether the presented letter was in uppercase or lowercase. On every trial a letter was displayed in white in the center of the screen, next to a red square "cursor" (approximately 1 degree of visual angle) that moved with the joystick of a gamepad. The display also included a horizontal line presented 6.5 degrees of visual angle above and below the letter stimulus. Participants were instructed to move the joystick with their left thumb to push the cursor above the top line if the letter was uppercase, and to move it below the bottom line if it was lowercase. Participants then pushed a button on the gamepad with their right thumb to

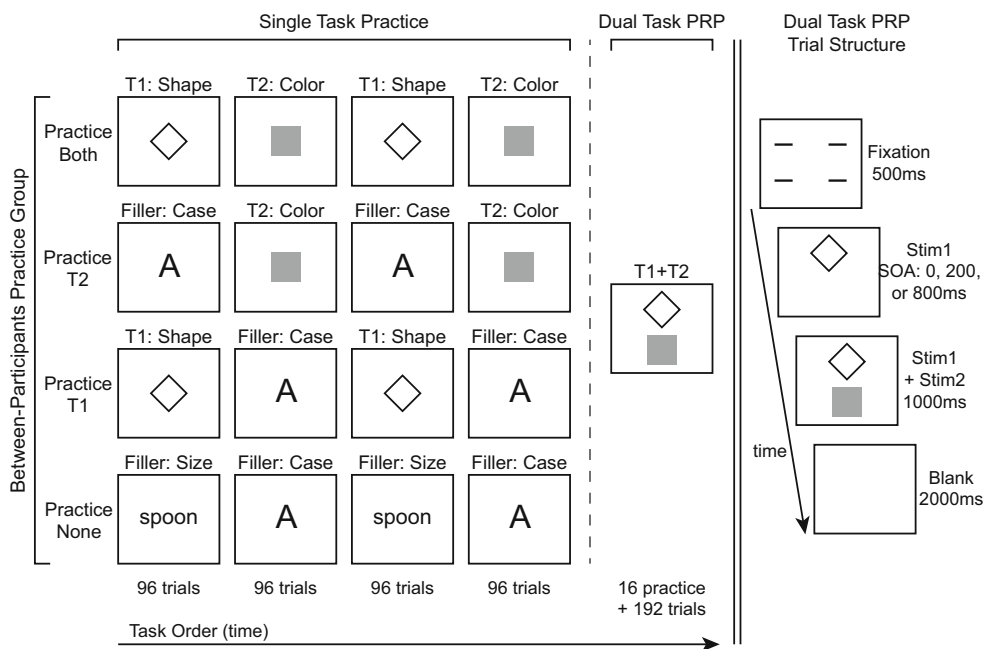


Fig. 1 Method and design for Experiment 1. Four separate groups of participants completed single-task practice with a range of tasks, potentially including Task 1 (T1) and/or Task 2 (T2) of an eventual dual-task psychological refractory period (PRP) paradigm common to all participants. Dual task PRP performance and backward compatibility

effects with the shape (T1) and color (T2) tasks were assessed relative to prior experience with these tasks as single tasks. Case judgements on letters and semantic size judgements on words were used as unrelated filler tasks in single-task practice. (SOA = stimulus onset asynchrony)

submit their response. The trial display remained on the screen until participants initiated a button-press response with the joystick at the top or bottom of the screen; button presses with the joystick cursor in any other location did not end the trial. At the end of each trial, participants were instructed to release the joystick so that it was in a neutral (central) position for the beginning of the following trial.

In the *size* task, participants judged whether a noun word referred to an object that was larger or smaller than the computer monitor. On every trial, a word was displayed in white on the center of the screen against a black background, with a white “+” symbol cursor immediately below it. Participants were instructed to use the mouse to move the cursor as far left on the screen as possible if the word referred to something smaller than the monitor, and to move it to the right edge of the screen if the word referred to something larger than the computer monitor. Once the mouse moved to the edge of the monitor, the selection was recorded and the trial ended.

Participants performed two blocks each of two single tasks in alternating and counterbalanced order. Each of the four blocks in the single-task phase consisted of 96 trials, with all stimuli presented an equal number of times within a block. Exclusively at the end of each block of 32 trials, participants’ average RT and accuracy were displayed, and they had the opportunity to rest before initiating the beginning of the next set of 32 trials. At the start of each block, participants were

informed and reminded of the task rules and response mappings for the task for that block.

Dual task All participants completed an identical dual-task phase, where Task 1 was the *shape* task and Task 2 was the *color* task, presented in a PRP dual-task paradigm. Every trial began with a fixation display for 500 ms, consisting of two rows of two dashes in the center of the screen, flanking the locations where the shape and color stimuli would appear. The Task 1 shape stimulus was presented vertically above the Task 2 color stimulus, separated by approximately 1.5 degrees of visual angle. The onset of the Task 1 and Task 2 stimuli were separated in time by an SOA of 0, 200, or 800 ms. Both stimuli were then displayed together for 1,000 ms, followed by a blank screen for 2,000 ms. Participants made separate responses to each task by pressing the 1 or 2 key on the number pad of a standard keyboard with their dominant hand. Responses were compatible if both tasks required the same manual response (the same key and thus finger for both tasks) and incompatible if each task required a different manual response. Response mapping was counterbalanced across tasks and participants and was consistent with response mappings in the single-task phase for individual participants. Participants were instructed to respond to both tasks as quickly and accurately as possible, but to place special emphasis on Task 1 and to make their response to it first before considering Task 2. Participants were provided with a note attached to the bottom

of the monitor, reminding them of the response mapping for both tasks.

The dual-task phase consisted of 16 initial practice trials that were not included in the analysis, and 192 experimental trials, made up of four iterations of the factorial combination of the four Task 1 (shape) stimuli, four Task 2 (color) stimuli, and three SOAs. These trials were divided into six blocks of 32 trials each. Participants received feedback about their overall accuracy and RT for Task 1 after every set of 32 trials and had the opportunity to rest before initiating the start of the next block. The entire experiment was completed in a single 1-hour session.

Data analysis

Our analyses focused on data from the dual-task phase. Mean reaction times for each condition were computed from trials where both Task 1 and Task 2 responses were correct. Participants' data were excluded from analysis if Task 1 accuracy was less than 85 % across the experiment or if the overall accuracy measure was less than 70 %, representing a per-task error rate of approximately 85 %. This criterion resulted in the elimination of seven participants' data, leaving 24, 26, 25, and 21 participants in the Practice-Both, Practice-T2, Practice-T1, and Practice-None groups, respectively. Trials with response latencies of less than 200 ms on either Task 1 or Task 2, or greater than 2,000 ms for Task 1 or 2,500 ms for Task 2, were excluded from analysis.

Analysis of RT data focused on the effects of single-task practice on (a) overall Task 1 and Task 2 reaction time performance in the PRP task, and (b) the presence and timecourse of development of backward compatibility effects, measured as response compatibility effects on Task 1 RT at the 0 ms SOA. We additionally examined the Task 1 response compatibility effects on 0 ms SOA trials across interresponse interval (IRI) quartiles to determine if compatibility effects were caused by response grouping and assessed error data to rule out speed-accuracy trade-off effects.

Initial inspection of RT data in all participant groups across 32-trial blocks revealed extremely variable RT performance in the first dual-task experimental block immediately following the single-task phase. Given our experimental and theoretical focus on distinguishing situations where backward compatibility effects are present versus absent, we were concerned that such variability might unduly bias us away from finding these effects when they may be present for some groups early in the timecourse of our data. To guard against this potential issue, and to improve our ability to discern between presence and absence of BCEs, we excluded this first dual-task block of 32 trials from analysis. We suggest this is a somewhat conservative approach, given that we are critically looking for differences in BCEs between conditions rather than simply trying to observe the effect in general.

Results and discussion

Effects of single-task practice on overall RT1 and RT2 performance

Mean reaction time data for Task 1 and Task 2 are presented in Fig. 2 for each practice group, separated by experimental half (first vs. second sets of 80 trials). To assess single-task practice effects on overall RT performance, we performed separate $3 \times 2 \times 2$ repeated measures ANOVAs on Task 1 and Task 2 RT data, with a within-subjects factor of SOA (0, 200, 800 ms) and between-subjects factors of prior Task 1 practice (yes, no) and prior Task 2 practice (yes, no), collapsing over response compatibility and experimental half. Subject was a random factor in this and all other ANOVAs reported in Experiments 1 and 2.

For Task 1 RT data, there was a strong main effect of prior Task 1 practice, $F(1, 92) = 8.613, p < .01$, with faster Task 1 RT in dual-task performance if Task 1 had been previously practiced as a single task. There was no observed influence of prior single task practice with Task 2, and no interaction of single task practice effects, $F_s < 0.5$. There was a marginal main effect of SOA, $F(2, 184) = 2.409, p = .093$, and a nonsignificant interaction of SOA with Task 1 and Task 2 single-task practice, $F(2, 184) = 2.066, p = .130$, possibly reflecting the influence of prior task experience being more detectable in short versus long SOA conditions with more or less task overlap and concurrent task demands. Other interactions were not significant, $F_s < 1.65$.

For Task 2 RT data, there was a strong main effect of SOA, $F(2, 184) = 1458.73, p < .001$, reflecting the PRP effect of increasing Task 2 RT with decreasing SOA. There was a strong interaction of SOA with prior Task 2 single-task practice, $F(2, 184) = 7.469, p < .001$, modifying a marginal main effect of prior Task 2 practice, $F(1, 92) = 3.652, p = .059$, reflecting faster Task 2 RT in dual-task performance if Task 2 had previously been practiced as a single task. There was also a main effect of prior Task 1 practice on Task 2 RT, $F(1, 92) = 6.196, p < .05$, with Task 2 RT faster in conditions with prior Task 1 single-task practice; however, this effect may be due to propagation of Task 1 RT effects onto Task 2 RT via PRP effects, most notably at earlier SOAs.

As a less biased and more direct assessment of prior task practice on Task 2 dual-task performance, we assessed single-task Task 1 and Task 2 practice effects on RT2 minus RT1 data at 0 and 200 ms SOAs. This measure eliminates the influence of duration differences in Task 1 that propagate onto RT2. At 0 ms SOA, adjusted RT2 was significantly faster with prior Task 2 single task practice, $F(1, 92) = 6.261, p < .05$, with no main effect or interaction of prior Task 1 practice, $F_s < 0.7$. Similarly at 200 ms SOA, adjusted RT2 was significantly faster with prior Task 2 single-task practice, $F(1, 92) =$

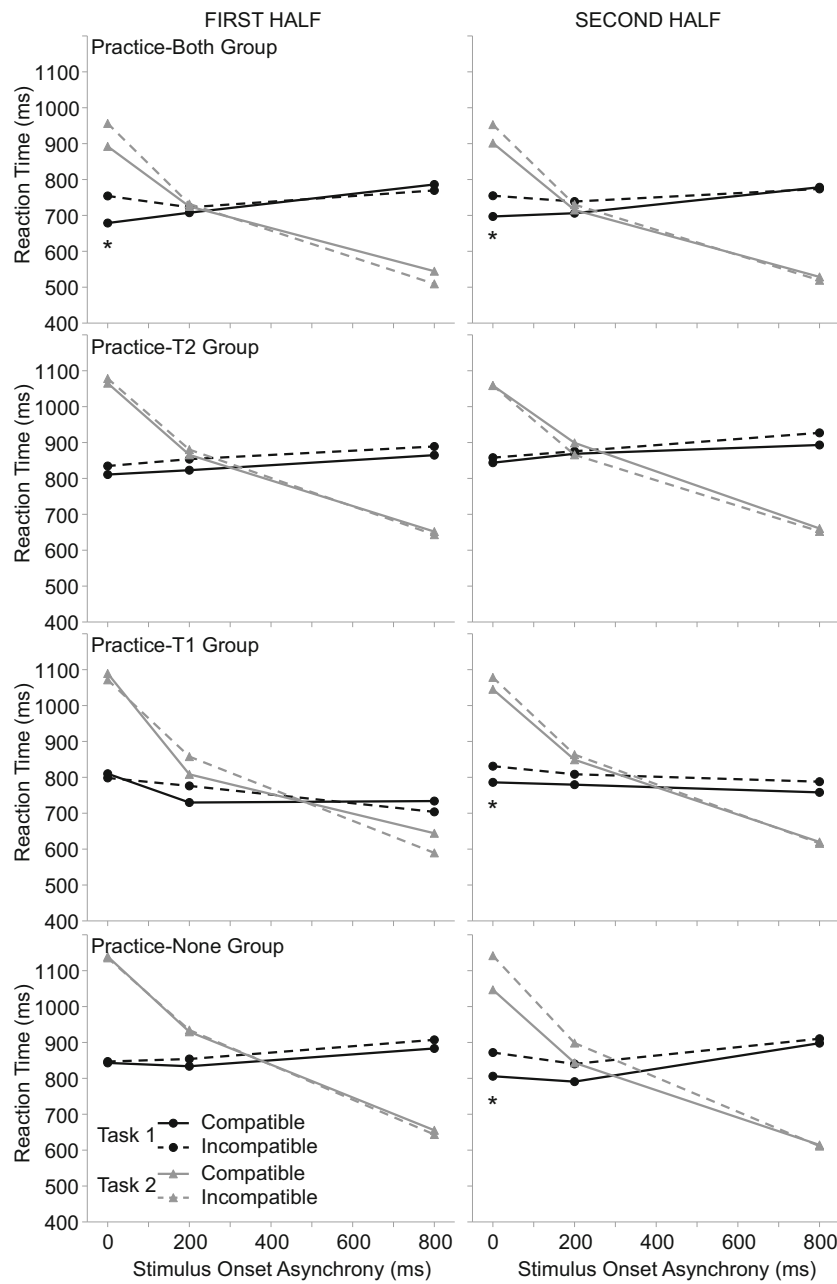


Fig. 2 Mean reaction time data for Experiment 1. Data are split by first and second halves and separated by between-subjects practice group. Each panel shows data for Task 1 and Task 2, divided by stimulus onset asynchrony (SOA) and response compatibility. Backward compatibility

effects (BCE) at 0 ms SOA are observed in both halves of the Practice-Both group, and in the second halves of Practice-T1 and Practice-None groups, indicated by an asterisk. No BCE was observed in the Practice-T2 group

7.794, $p < .01$, again with no influence of prior Task 1 single task practice, $F_s < 0.2$.

These results present a clear picture – practicing Task 1 or Task 2 as a single task improves subsequent dual-task performance for that particular task. For Task 2 dual-task performance, the influence of prior Task 1 practice is indirect, with changes in overall Task 2 RT due to savings or costs in Task 1 RT propagated onto Task 2 via the PRP effect. This pattern of practice effects is consistent with general expectations of RT

performance improvements with practice. We next assessed BCEs for effects of prior single-task practice.

Effects of single-task practice on BCE

We assessed backward compatibility effects here as the influence of Task 2 to Task 1 response compatibility relationships on Task 1 RT at the 0 ms SOA. In keeping with many prior studies, we anticipated that these effects should be best

observed at very short SOAs, where there is sufficient task overlap to allow Task 2 response activation to influence Task 1 response selection. A direct-link model predicts that BCEs will develop over time as participants acquire automaticity for Task 2. As such, we were interested to test for the presence or absence of BCEs over halves of our dual-task data (sets of 80 trials). An episodic direct-link model additionally predicts that this development of BCEs over dual-task performance should be sensitive to how Task 2 automaticity is expressed relative to prior Task 2 learning context. To test for these potential effects, we assessed the development of BCEs over dual-task experiment halves conditional on participants' prior experience with single-task practice with Task 1 and/or Task 2.

Mean BCE data (mean response-incompatible minus response-compatible Task 1 RT for 0 ms SOA trials) with 95 % confidence intervals are presented in Fig. 3, separated by practice group and experimental half (first vs. second sets of 80 trials). An initial omnibus $2 \times 2 \times 2 \times 2$ repeated measures ANOVA, with within-subject factors of response compatibility (compatible, incompatible) and experiment half (first, second) and between-subjects factors of prior Task 1 practice (yes, no) and prior Task 2 practice (yes, no), suggested a mixture of robust BCEs present in different conditions across our four experimental groups. In addition to a strong main effect of response compatibility, $F(1, 92) = 18.159, p < .001$, we observed a significant interaction of response compatibility, prior Task 1 practice and prior Task 2 practice, $F(1, 92) = 4.206, p < .05$, and of response compatibility, experiment half and prior Task 2 practice, $F(1, 92) = 7.780, p < .01$. Given these significant omnibus effects, we conducted more fine-grained analyses to better establish the pattern of BCEs present in this experiment.

Considering Fig. 3, the above omnibus interaction of response compatibility, experiment half, and prior Task 2 practice reflects differences in the pattern of BCEs over experiment halves, between conditions with prior Task 2 practice (Practice-Both and Practice-T2 groups) and with no prior Task 2 practice (Practice-T1 and Practice-None groups). First, considering conditions with prior Task 2 practice, we

observed an interaction of response compatibility and experiment group, $F(1,48) = 5.449, p < .05$, with no interaction of response compatibility and experiment half, and no interaction of these two factors with experiment group, $F_s < 0.6$. These data suggest substantially larger BCEs the Practice-Both group compared to the Practice-T2 group, with a consistent degree of BCEs across both halves of the experiment for both groups.

Considering the Practice-Both group alone, the prominent BCEs observed in both halves of our experiment were supported by a strong main effect of response compatibility, $F(1, 23) = 32.984, p < .001$, with no main effect or interaction with experiment half, $F_s < 0.7$. Considering the Practice-T2 group alone, we observed no evidence of a BCE in either half of the experiment. There was no main effect of response compatibility, $F(1, 25) = 1.324, p = .261$. There was a marginal main effect of experiment half, $F(1, 25) = 3.434, p = .076$, but this was not observed to interact with response compatibility, $F < 0.2$.

Next, considering conditions with no prior Task 2 practice (Practice-T1 and Practice-None groups), we observed an interaction of response compatibility and experiment half, $F(1, 44) = 9.813, p < .01$, with no interaction of experiment group with any factors, $F_s < 0.6$. These data suggest the pattern of increasing BCEs over halves of the experiment was similar for both groups.

Considering the Practice-T1 group alone, we observed no BCE initially, but a substantial BCE appeared to develop by the second half of the experiment. This observation was supported by the interaction of response compatibility and experiment half, $F(1, 24) = 4.601, p < .05$, with no main effects observed, $F_s < 0.8$. Subsequent t tests to examine this interaction showed no effect of response compatibility in the first half of the experiment, $t(24) = 0.468, p = .644$, but a significant BCE in the second half, $t(24) = 2.174, p < .05$. Considering the Practice-None group alone, data appeared similar to the Practice-T1 group, with no BCE in the first half of the experiment, but a substantial BCE in the second half. These observations were supported by an interaction of response compatibility and experiment half, $F(1, 20) = 5.348, p < .05$,

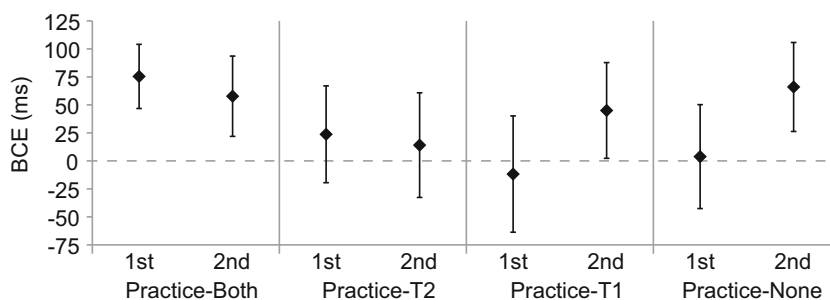


Fig. 3 Mean backward compatibility effect (BCE) data for 0 ms SOA trials for Experiment 1. Data are separated by first and second halves and between-subjects practice group. Error bars show 95 % Confidence

Intervals. BCEs are observed in both halves of the Practice-Both group, and in the second halves of Practice-T1 and Practice-None groups. No BCE was observed in the Practice-T2 group

modifying a significant main effect of response compatibility, $F(1, 20) = 4.917, p < .05$, and no main effect of experiment half, $F < 0.1$. Subsequent t tests to examine this interaction showed no effect of response compatibility in the first half of the experiment, $t(20) = 0.865, p = .172$, but a significant BCE in the second half, $t(20) = 3.466, p < .01$.

Interpreting BCE prior-practice effects

The observed pattern of BCE effects across prior-practice conditions is difficult to account for under the WM-mediated transient-link model or a traditional S-R learning direct-link model, but it is well-predicted by the episodic direct-link model. When participants had no prior experience with Task 2 prior to dual-task performance (Practice-T1 and Practice-None groups), previous Task 1 experience did not seem to matter for BCEs – we observed no BCE in the initial half of dual-task performance, but a clear BCE had developed by the second half in both of these groups. In the Practice-Both condition, with prior single task practice with Task 1 and Task 2, the BCE was well-developed within the first half of dual-task performance and persisted over halves.

Against this background of demonstrated BCEs, the Practice-T2 condition provides a critical dissociation of episodic direct-link versus other models. Given the elicitation of BCEs relative to practice as above, the WM-mediated model would strongly predict a BCE to be present, given prior Task 2 practice. If experience with Task 2 is primarily important, a BCE might be predicted in the first half of dual-task performance for Practice-T2, given the BCE observed in the Practice-Both condition; if not, then at least by the second half, where BCEs had developed with much less experience with Task 2 in Practice-T1 and Practice-None conditions. The simple S-R learning direct-link model makes a similar prediction for any prior Task 2 practice to produce BCEs, via prior strengthening of S-R links in long-term memory. However, the BCE was not observed in the Practice-T2 condition, even in the latter half of dual-task performance.

This absence of BCEs in Practice-T2 dual-task performance occurred despite prior Task 2 practice producing substantial overall improvement on Task 2 RT in these trials. Considering the episodic model, we contrast this absence of a BCE in Practice-T2 with the development of BCEs over trials in Practice-T1 and Practice-None conditions. With no prior contextually sensitive learning of Task 2 to compete or interfere, Task 2 S-R automaticity could develop relatively rapidly over trials to produce BCEs. In the Practice-T2 condition, prior Task 2 practice actually appears to prevent the development of a BCE over the same timecourse, while still showing expected improvements on overall Task 2 RT performance. The difference between conditions with prior Task 2 practice (Practice-T2 vs. Practice-Both) would appear to be

the contextual similarity of the other prior single-task practiced in alternating blocks with Task 2. We suggest that having blocks of Task 1 and Task 2 interleaved in single-task practice may have provided a better match to eventual dual-task context, compared to interleaving Task 2 with a different task. This pattern of selective context-sensitive disruption of BCEs in the presence of otherwise facilitative practice effects on Task 2 RT, and clear development of BCEs in the same dual-tasks in other conditions, provides specific evidence in favor of the episodic model of BCEs and cannot be well accounted for by the WM-mediated model.

Interresponse interval quartile analyses

Finally, we assessed our 0 ms SOA Task 1 RT data for evidence that BCEs may have been generated only in trials with very short interresponse intervals. Such a finding would suggest compatibility effects could be caused by deliberate consideration of both tasks prior to responding rather than any requirement for parallel activation of Task 2 response information during Task 1 performance. For datasets showing BCEs (all groups except for Practice-T2 data), we determined interresponse interval (IRI) quartiles for 0 ms SOA trials for each participant and compatibility condition and then calculated mean Task 1 reaction times for each quartile. These data are shown in Table 1. We assessed these data with 2 (response compatibility) \times 4 (IRI quartile) repeated measures ANOVAs for each of the separate practice groups. We observed no interactions of response compatibility effects with IRI quartile, all $F_s < 1$, suggesting that any potential response grouping was not the cause of our observed BCEs.

Error rate analyses

Given the strong theoretical and experimental focus on 0 ms SOA data, for brevity we limit our reporting of error rate data here to errors in 0 ms SOA trials for Task 1 and Task 2. Error rates in 0 ms SOA data were overall slightly higher than at longer SOAs, likely due to the high degree of task concurrency, and represent the general profile of error performance across SOA conditions. Mean Task 1 and Task 2 error rates for 0 ms SOA trials are summarized in Table 2. Trials with an error committed on Task 1, regardless of Task 2 accuracy, and trials with an error committed on Task 2 following accurate Task 1 performance were submitted to separate $2 \times 2 \times 2 \times 2$ repeated measures ANOVAs, with within-subject factors of experiment half (first, second), response compatibility (compatible, incompatible), and between-subjects factors of prior Task 1 practice (yes, no) and prior Task 2 practice (yes, no). There were a very small number of trials with errors committed on both Task 1 and Task 2, which we excluded from analysis here.

Table 1 Mean Task 1 reaction times (ms) and standard errors of the mean (*SE*) for backward compatibility effect data (compatible vs. incompatible trials) at 0 ms stimulus onset asynchrony, separated by interresponse interval (IRI) quartiles

Task 1 Response Compatibility	IRI Quartile							
	1		2		3		4	
	Mean	<i>SE</i>	Mean	<i>SE</i>	Mean	<i>SE</i>	Mean	<i>SE</i>
Exp. 1, Practice-Both group								
Compatible	713	30	701	34	682	23	668	23
Incompatible	787	29	767	23	749	34	702	29
Exp. 1, Practice-T1 group								
Compatible	820	39	790	38	789	41	755	28
Incompatible	823	29	806	30	807	39	792	33
Exp. 1, Practice-Both group								
Compatible	850	39	825	50	819	47	838	49
Incompatible	899	42	887	47	828	46	864	44
Exp. 2, Before phase, T2-Reversed group								
Compatible	765	26	759	28	778	30	794	25
Incompatible	847	26	831	28	817	32	790	28
Exp. 2, Before phase, T2-Conflict group								
Compatible	696	28	711	30	694	32	747	27
Incompatible	763	28	742	30	735	34	769	29
Exp. 2, Before phase, T2-Different group								
Compatible	716	27	745	30	754	31	763	27
Incompatible	835	28	813	30	781	34	777	29
Exp. 2, Interference Phase, T2-Different Group								
Compatible	694	29	710	30	697	44	740	35
Incompatible	725	24	810	41	741	33	792	33
Exp. 2, After phase, T2-Reversed group								
Compatible	777	33	746	35	756	28	810	30
Incompatible	846	31	829	34	799	37	806	35
Exp. 2, After phase, T2-Conflict group								
Compatible	723	36	714	38	703	30	760	32
Incompatible	805	33	748	36	790	39	779	37
Exp. 2, After phase, T2-Different group								
Compatible	700	35	699	37	665	30	689	32
Incompatible	788	33	783	35	752	39	761	36

For Task 1 error rate, we observed a main effect of response compatibility, $F(1, 92) = 32.78, p < .001$, indicating that compatible trials were more accurate than incompatible trials. The interaction of response compatibility and experiment half approached significance, $F(1, 92) = 3.38, p = .069$, suggesting a trend toward a larger compatibility effect in the second half of the experiment. There were no significant effects involving prior Task 1 or Task 2 practice, although the interaction of response compatibility and Task 1 practice was marginally significant, $F(1, 92) = 3.32, p = .072$, describing a trend toward larger compatibility effect when Task 1 had been practiced previously. Most importantly for our interpretation of backward compatibility effects in Task 1 RT data, error rate

data for Task 1 performance indicated no evidence of a speed-accuracy trade-off with respect to response compatibility effects.

For Task 2, we observed a marginal main effect of experiment half, $F(1, 92) = 3.09, p = .082$, suggesting a trend toward more accurate performance in the second half of trials. In addition, there was a strong reversed compatibility effect $F(1, 92) = 25.45, p < .001$, with more accurate performance on incompatible trials compared with compatible trials. There were no significant effects involving prior Task 1 or Task 2 practice, but the three-way interaction of experiment half, response compatibility, and Task 1 practice approached significance, $F(1, 92) = 2.94, p = .090$.

Table 2 Mean error rates (%Err) and standard errors of the mean (SE) for 0 ms stimulus onset asynchrony dual-task performance in Experiment 1

Experiment half	Task 1				Task 2			
	Compatible		Incompatible		Compatible		Incompatible	
	%Err	SE	%Err	SE	%Err	SE	%Err	SE
Practice-Both								
First	3.50	0.01	8.11	0.02	12.23	0.03	7.24	0.02
Second	1.46	0.01	11.63	0.03	9.20	0.01	3.93	0.01
Practice-T2								
First	4.04	0.01	6.47	0.02	11.94	0.02	6.03	0.02
Second	2.23	0.01	5.32	0.01	10.35	0.02	6.31	0.01
Practice-T1								
First	2.99	0.01	4.76	0.01	9.19	0.02	8.89	0.02
Second	1.51	0.01	6.18	0.02	11.01	0.02	7.84	0.02
Practice-None								
First	3.89	0.01	5.69	0.02	11.42	0.02	3.93	0.02
Second	2.37	0.01	6.02	0.02	6.62	0.02	3.94	0.02

We suggest that these reversed compatibility effects might represent some kind of partial feature match interference effect, with respect to the greater dual-task context (e.g., Hommel, 2004, 2007, 2009) – following a correct response to Task 1 with an index finger, a perfect match to this event file would be to make the same response to the same task and stimulus; making the same response for Task 2 in a response-compatible dual-task trial is a partial mismatch to this immediately prior performance, and elicits interference, where making a subsequent different response to a different task may elicit less conflict (Hommel, 2007, 2009). In contrast, Task 1 performance does not have an immediate prior event that interferes to this extent, and so we see the influence of compatibility of simultaneously activated response information from Task 2 stimuli on Task 1 RT.

Experiment 2

In Experiment 1, we assessed the effect of prior single-task practice with Task 1 and/or Task 2 of a dual-task paradigm on the development of BCEs in dual-task performance. Experiment 1 demonstrated that BCEs are sensitive to the context in which prior Task 2 learning has occurred and that these effects are dissociable from practice effects on overall dual-task RT performance. The presence and timecourse of development of BCEs elicited over different practice conditions in Experiment 1 is well predicted by an episodic direct-link account of the BCE, but it is much less well accounted for by the WM-mediated transient-link model.

Experiment 2 was designed to test direct-link model predictions of how BCEs should develop over trials. Experiment 2 does not make strong experimental dissociations between traditional S-R learning versus episodic versions of a direct-link account. Instead, we focus on a common set of detailed predictions for long-term memory direct-link representations of Task 2 S-R relationships and how these representations should develop over time to produce BCEs and differentially interfere with BCEs with subsequent different Task 2 performance.

Ellenbogen and Meiran (2008) assessed their own data for potential direct-link effects and found considerable evidence of increases in BCE over blocks of practice. Ellenbogen and Meiran suggested that this kind of learning over blocks for BCEs was suggestive of long-term memory S-R direct-link-mediated activation of Task 2 response information, independent of WM-mediated rules. They argued that this direct-link element developed over experience and likely contributed to BCE production in addition to the main WM-mediation mechanism they primarily argued for.

To more closely evaluate direct-link predictions about the timecourse of development of BCEs relative to influences of prior Task 2 learning, Experiment 2 first trained all participants in the same dual-task paradigm to observe the development of BCEs over practice, in an initial Before phase – participants performed a shape discrimination task as Task 1 and judged digits as higher or lower than 5 as Task 2. Here, a direct-link model predicts BCEs to develop over practice and establish strong Task 2 learning for the rest of the experiment.

Next, in the Interference phase, participants performed the dual-task paradigm with one of three different alternatives for Task 2: In the Reversed-T2 group, participants continued the original Task 2 (digit high/low task) but with reversed response mapping requirements; in the Conflict-T2 group, participants performed a new categorization task but using the same stimuli as the original Task 2 (classifying digits as odd or even); and in the Different-T2 group, participants performed an unrelated task with different stimuli (a color-discrimination task). The timecourse of BCE development in the Interference condition was assessed across groups. Here, the direct-link model predicts that the development of BCEs in the Reversed-T2 and the Conflict-T2 conditions should be adversely affected due to conflicting prior learning on Task 2 in the Before phase. If BCEs in these conditions differ, we predict that the Reversed-T2 group should be more adversely affected because of greater interference from prior Task 2 learning. In contrast, BCEs should develop relatively quickly for the Different-T2 group, where prior Task 2 learning from the Before phase has little relationship with the new Task 2.

Finally, in the After phase, all participants returned to the original dual-task, and we assessed the timecourse of reestablishment of BCEs with respect to Task 2 learning in the Interference phase. The direct-link model predicts that BCE development here should occur with a faster timecourse than for the Interference phase, as there is substantial prior Task 2 learning from the Before phase to support this. The Different-T2 group should show relatively faster reestablishment of BCEs here, given the lack of overlap of prior Task 2 learning from the Interference phase. Conversely, the Reversed-T2 group should show relatively slower reestablishment of BCEs here due to Task 2 learning of opposite response mappings for the same task in the preceding Interference phase. We predict that the Conflict-T2 group should show an intermediate effect between these two.

The Before and Interference phases of the Reversed-T2 experimental group in the present Experiment 2 represent a close replication of the design used in Hommel and Eglau's (2002) Experiment 4. They showed that backward compatibility effects observed in the first half of their experimental trials disappeared in the second half of trials when the S-R mapping in Task 2 was reversed. Each experiment half consisted of 100 trials, which is equivalent to the length of the Interference phase of the present experiment (96 trials). Hommel and Eglau argued that once S-R associations are acquired, this learning makes it difficult to associate the same codes in different ways. Even though it is likely that backward compatibility effects would emerge under the new reversed response mapping with a sufficient amount of practice, the results of Hommel and Eglau suggest that the amount of practice required is potentially greater than that needed to acquire the original associations. Experiment 2 extends Hommel and Eglau's approach, to compare the timecourse of BCE development over the Interference phase with the recovery of original Task 2 learning in the After phase and assesses all of these effects across multiple degrees of interference on Task 2 learning across experimental groups. We consider our results below with respect to the detailed and specific pattern of predicted BCEs via the general direct-link model.

Method

Participants

One hundred one participants (74 females, mean age = 18.8 years) were recruited from the McMaster University undergraduate population. They were all enrolled in psychology courses and received partial course credit for their participation. This study was approved by McMaster's Research Ethics Board, and all participants gave written informed consent prior to beginning the experiment. All participants had normal or corrected to normal visual acuity, and normal color vision.

Apparatus and stimuli

Apparatus was the same as that used for Experiment 1. Four basic tasks were used to create a number of different PRP paradigms. Two of these tasks were the same *shape* and *color* tasks used in Experiment 1. The *high/low* task asked participants to classify the value of single digit stimuli as higher or lower than 5, and the *odd/even* task asked participants to classify single digit stimuli as odd or even. *High/low* and *odd/even* tasks used the same set of eight single-digit stimuli (1 to 9, excluding 5), presented in white against a black background, with a height of approximately 1.25 degrees of visual angle.

Design and procedure

An outline of the experimental design and procedure is shown in Fig. 4. All participants completed three typical PRP dual-task phases: an initial Before phase, a subsequent Interference phase, and, finally, the After phase. In the Before and After phases, all participants performed the same PRP task, with the *shape* task as Task 1 and the *high/low* task as Task 2. For the Interference phase, all participants still performed the *shape* task as Task 1, but Task 2 was varied between three between-subjects groups. In the Reversed-T2 group, participants continued to perform the *high/low* task as Task 2, but with the response key mapping reversed. In the Conflict-T2 group, participants performed the *odd/even* task as Task 2, making different task judgements to the same set of digit stimuli previously seen in the *high/low* task. In the Different-T2 group, participants performed the *colour* task as Task 2, with minimal task or stimulus overlap with the previously performed *high/low* task.

Trial elements and timing for all three phases were identical to the dual-task phase of Experiment 1, except for the specific Task 2 stimuli presented. The Before phase consisted of 16 practice trials that were not included for analysis and 288 experimental trials made up of three iterations of the factorial combination of the four Task 1 shape stimuli, the eight Task 2 digit stimuli, and the three SOAs. The Interference and After phases consisted of 16 practice plus 96 experimental trials each, made up of factorial combinations of Task 1 and Task 2 stimuli and SOAs. In all phases, response mapping was counterbalanced across tasks and participants. Participants were given rest breaks every 32 trials throughout the experiment, along with on-screen feedback about overall accuracy and mean Task 1 RT, and self-initiated subsequent sets of 32 trials. The entire experiment was completed in a single 1-hour session.

Data analysis

The experiment was initially performed and analyzed with an original 55 participants; we subsequently doubled the size of

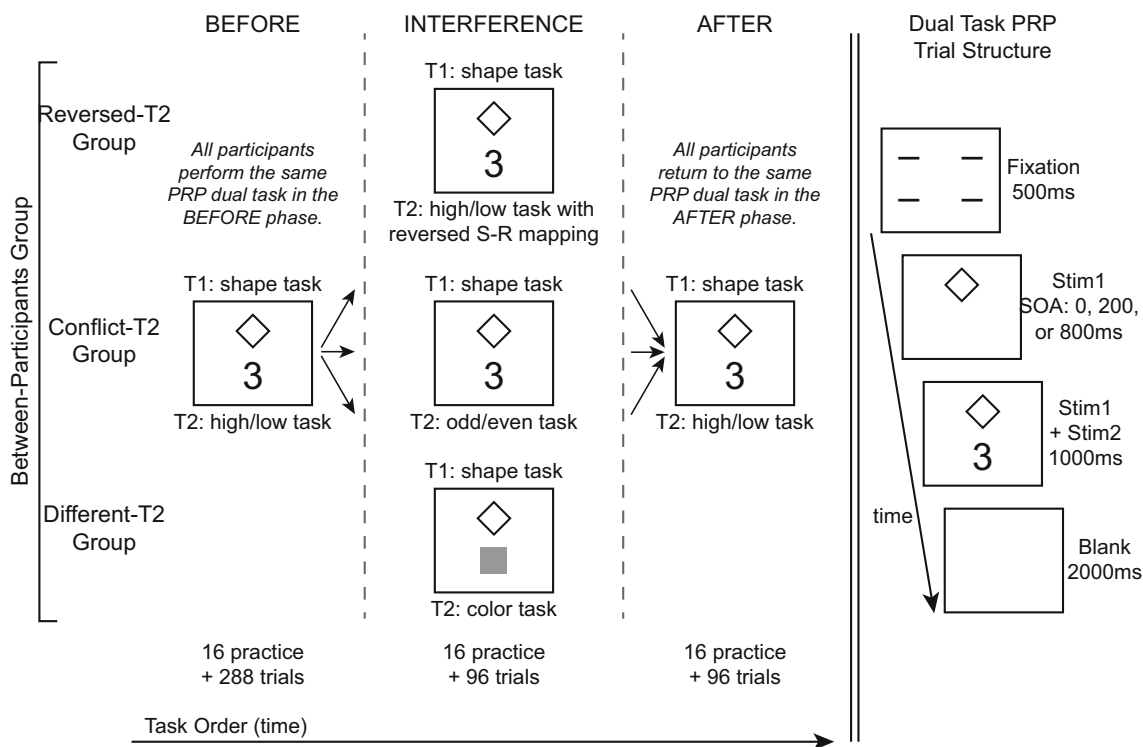


Fig. 4 Method and design for Experiment 2. Three separate groups of participants completed the same PRP dual task in the Before phase of the experiment, classifying shapes as Task 1 (T1), and classifying digits as higher or lower than 5 as Task 2 (T2). In the Interference phase, Task 2 was altered for each group; in the After phase, Task 2 returned to the same

design as in the Before phase. Backward compatibility effects were assessed with respect to potential conflict from overlapping Task 2 S-R mapping rules in the Interference phase compared to Before and After phases. (SOA = Stimulus Onset Asynchrony)

the design (collected another 46 participants) in order to increase power. To guard against possible inflation of Type I error rate (potential Type I error in our initial dataset may lead us to trust these outcomes and collect more data – hence the *p* values of the whole dataset may be exaggerated), we treated the first and second samples as separate replications of the same study and assessed these two datasets for similar effects of BCE over task and training manipulations and potential influences of sample on these effects. For the two main omnibus ANOVAs reported at length below (overall BCE effects across the design, and block-by-block analyses of BCE timecourse over training), we performed additional analyses with the added factor of dataset (first vs. second). We observed extremely similar effects in both datasets and in all of these analyses observed no interactions of dataset with any other factor, all *F*s < 1.2. Given the similarity of effects in both datasets and the lack of any interactions between datasets, we conclude that our second set of data closely replicates the first. As such, we present the subsequent analyses collapsed over both sets of data.

Mean reaction times for each condition were computed from trials where both Task 1 and Task 2 responses were correct. As in Experiment 1, participants’ data were excluded from analysis if Task 1 accuracy was less than 85 % across the experiment or if the overall accuracy measure was less than

70 %, representing a per-task error rate of approximately 85 %. This resulted in the elimination of nine participants’ data, leaving 92 in total, with 33, 29, and 30 participants in Reversed-T2, Conflict-T2, and Different-T2 groups, respectively. Trials with response latencies of less than 200 ms on either Task 1 or Task 2 or with greater than 2,000 ms for Task 1 or 2,500 ms for Task 2 were also excluded from analysis.

Data analysis in Experiment 2 focused even more directly on the presence versus absence and development over experience of backward compatibility effects, assessed as response compatibility effects on Task 1 RT at 0 ms SOA, within a dual-task PRP paradigm. The design of our experiment was to (a) induce BCEs with the same set of tasks in identical PRP paradigms in all three experimental groups in the initial Before phase, (b) give different Task 2 requirements to our three experimental groups and observe the effect of prior learning on development of BCEs in the subsequent Interference phase, and then (c) return all groups to the original PRP dual-task and look at how previously established BCEs recovered in the After phase. Initial analyses focused on response compatibility effects in Task 1 RT at 0 ms SOA across groups, separately for Before, Interference, and After phases. Subsequent analyses divided these phases more finely into 48-trial segments and assessed the timecourse of development and disruption of BCEs across these phases. We additionally

examined BCEs across interresponse interval quartiles, to determine if compatibility effects were caused by response grouping, and assessed error data to rule out speed-accuracy trade-off effects.

Results and discussion

Initial analyses of BCEs

Mean reaction-time data for Task 1 and Task 2 in Experiment 2 are shown in Fig. 5, separated by experiment phase (Before, Interference, and After) and experiment group.¹ BCEs are seen as the difference between response compatible versus incompatible trials in Task 1 RT, critically at the 0 ms SOA. We initially performed an omnibus ANOVA to assess BCEs across different phases of Experiment 2 and across our three experimental groups. We performed a $2 \times 3 \times 3$ repeated measures ANOVA on Task 1 RT data at 0 ms SOA, with within-subjects factors of response compatibility (compatible, incompatible), and experiment phase (Before, Interference, After), and a between-subjects factor of experiment group (Reversed-T2, Conflict-T2, Different-T2). We observed a strong main effect of response compatibility, $F(1, 89) = 69.029, p < .001$, with interactions of response compatibility with experiment group, $F(2, 89) = 5.270, p < .01$, and with experiment phase, $F(2, 178) = 7.931, p < 0.001$. Considering Fig. 5, these findings are consistent with the appearance of BCEs in all experiment groups in the Before and After phases, and for the Different-T2 group in the Interference phase but the absence of BCEs in the Interference phase for Reversed-T2 and Conflict-T2 groups. Additional effects of phase $F(2, 178) = 9.382, p < .001$, and phase by experiment group, $F(4, 178) = 6.215, p < .001$, support the observation of elevated Task 1 RT in the Interference phase for Reversed-T2 and Conflict-T2 groups, relative to other conditions.

Following up these omnibus tests, we conducted separate response compatibility by experiment group ANOVAs on 0 ms SOA Task 1 RT data, separately for each experiment phase, to better assess these effects and more directly test our hypotheses. The Before phase represents 288 trials of a dual-task PRP paradigm that was identical across all three experimental groups, designed to induce a comparable BCE in all three groups. ANOVA showed a strong effect of response compatibility, $F(1, 89) = 70.958, p < .001$. While the main effect of experiment group was also significant, $F(2, 89) = 3.504, p < .05$, suggesting some differences in overall RT

between groups, there was no interaction of this group effect with response compatibility, $F < 1$.

Having established comparable BCEs in each group, the Interference phase presented each group with 96 dual-task trials with a different Task 2, with varying relationships to the original Task 2 high/low digit-judgement task. Our prediction was that the high degree of interference involved from previous Task 2 learning on new performance in the Reversed-T2 and Conflict-T2 groups would interfere with development of BCEs arising from the new Task 2; in comparison, we predicted that the relatively low degree of interference from previous high/low digit-task learning on new performance in the Different-T2 group with the color-judgement task would allow relatively quicker development of a BCE arising from this new color Task 2. ANOVA showed an interaction of response compatibility and experiment group, $F(2, 89) = 3.292, p < .05$, modifying a main effect of group, $F(2, 89) = 7.113, p < .001$. There was no significant main effect of response compatibility, $F(1, 89) = 1.766, p = .187$. Subsequent *t* tests to assess the interaction revealed no response compatibility effects in Reversed-T2, $t(32) = 0.422, p = .676$, or Conflict-T2 groups, $t(28) = 0.113, p = .911$, but did show a significant BCE in the Different-T2 group, $t(29) = 2.71, p < .05$. We note that while high/low and odd/even tasks did not produce BCEs in our Interference phase, these tasks will produce typical BCEs of comparable sizes under more straightforward practice conditions, as seen in prior literature (e.g., Watter & Logan, 2006), and in the Before phase data in the present experiment.

The After phase of the experiment returned all groups to the same original Task 2 that all participants had initially performed in the Before phase, for a final 96 trials. We had predicted that BCEs were likely to return to all groups, given enough practice, but that participants in the Reversed-T2 and Conflict-T2 conditions might be delayed in redeveloping BCEs in the After phase relative to participants in the Different-T2 condition. Given that backward compatibility effects seem to develop over time, we might observe this as a presence versus absence of BCEs, or as a relative size effect on the BCE, depending on our sampling of this process. ANOVA revealed a strong main effect of response compatibility, $F(1, 89) = 46.967, p < .001$, with BCEs apparent in all three experiment groups. The interaction of response compatibility with group was not significant, $F(2, 89) = 1.29, p = .280$, nor was the effect of group, $F(2, 89) = 2.321, p = .104$. These results suggest that relative to the 96-trial interference phase, where BCEs did not develop for Reversed-T2 or Conflict-T2 groups following initial related but interfering Task 2 learning in the Before phase, the 96-trial After phase demonstrated relatively rapid recovery of original Task 2 learning, helping to overcome interfering Task 2 learning from the Interference phase and reestablish BCEs over this same shorter timeframe.

¹ For brevity, we focus only on Task 1 RT analyses of BCEs here in Experiment 2, as they speak most directly to the research questions at hand. We present the complete set of RT means for all conditions in Fig. 5 to show our early Task 1 backward compatibility effects in their complete dual-task context.

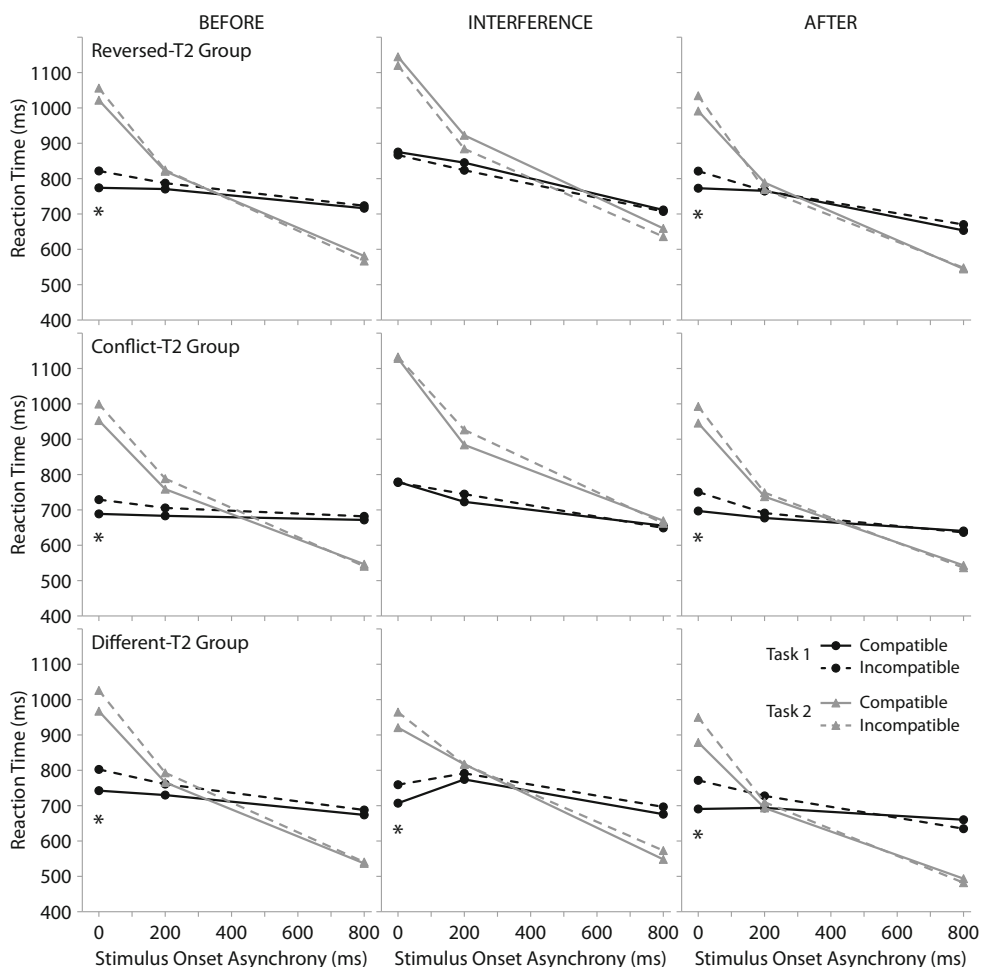


Fig. 5 Mean reaction time data for Experiment 2. Data are separated by experiment phase (Before, Interference, After) and between-subjects Task 2 Interference manipulation group. Each panel shows data for Task 1 and Task 2, divided by stimulus onset asynchrony and response compatibility.

Backward compatibility effects (BCE) are observed in all groups in Before and After phases, but only in the Different-T2 group in the Interference phase, and are indicated by an asterisk

Development of BCE over trials

As a further test of our hypotheses about the direct-link nature of BCEs, we conducted a more fine-grained analysis of the development of BCEs over trials within the task. We did this by dividing our Experiment 2 data into bins of 48 consecutive dual-task trials, and then assessing the 0 ms SOA Task 1 RT data within these bins for response compatibility effects. We chose to divide our data this way (using all PRP trials), and not by exact numbers of 0 SOA trials per bin, as all trials likely contribute to the development of expertise and automaticity on Task 2 performance that in turn is reflected as BCEs on Task 1 when tasks overlap sufficiently on short SOA trials. Divisions of 48 trails gave us six sequential bins of trials in the Before phase and two bins each in the Interference and After phases, with approximately eight compatible and eight incompatible 0 ms SOA trials per bin for each participant. Figure 6 shows mean RTs for these 0 ms SOA Task 1 data

for each experimental group, separated by bin and response compatibility.

In the Before phase, we assessed the development of BCEs over time with a $2 \times 6 \times 3$ repeated measures ANOVA, with within-subjects factors of response compatibility (compatible, incompatible) and bin (1 to 6), and a between-subjects factor of experiment group (Reversed-T2, Conflict-T2, Different-T2). For our subsequent analyses of BCE over time, we employed this same form of ANOVA, varying the number of bins included. We observed a significant interaction of response compatibility with bin, $F(5, 445) = 3.133, p < .01$, modifying the strong main effect of response compatibility, $F(1, 89) = 63.668, p < .001$. Within-subjects contrasts of the response compatibility by bin interaction showed significant linear, $F(1, 89) = 10.293, p < .01$, and quadratic trends, $F(1, 89) = 4.313, p < .05$, indicating the progressive development of the BCE over the Before phase in all three groups. The ANOVA showed no evidence of differential effects on BCEs

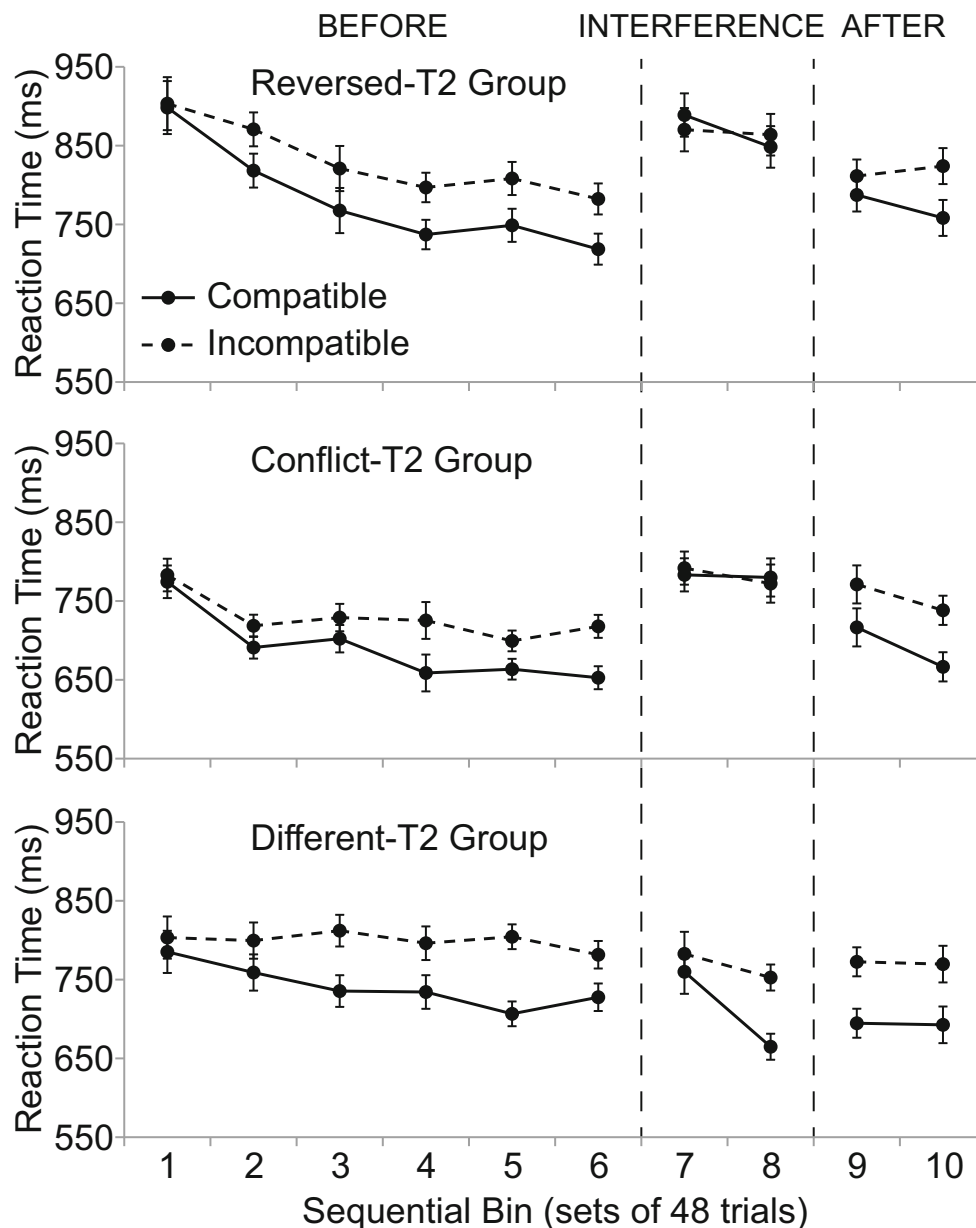


Fig. 6 Backward compatibility effect (BCE) data for Experiment 2. Mean Task 1 RT data at 0 ms SOA are shown for each sequential 48-trial bin over the experiment, separated by experiment phase and between-subjects Task 2 Interference manipulation group. Data are

divided by response compatibility, indicating the BCE in each bin in each group. Error bars represent the standard error of the difference between means for each response compatibility pair

between groups, with no interactions between experiment group and response compatibility, $F_s < 0.9$.

Group differences were apparent in overall RT, $F(2, 89) = 3.814, p < .05$, and in the interaction of experiment group with bin, $F(10, 445) = 2.360, p < .01$, highlighting the different rates of improvement in overall RT over bin between experimental groups. The main effect of bin was also significant, $F(5, 445) = 15.902, p < .001$, representing this general trend of improvement in RT over time. Repeating this ANOVA with only the second half of the Before phase (Bins 4 to 6) suggested that the experiment groups had equivalent and stable

BCEs by this time, with a strong response compatibility effect, $F(1, 89) = 79.056, p < .001$, and a remaining main effect of group, $F(1, 89) = 3.371, p < .05$, with no other effects or interactions, all $F_s < 1.5$.

The division of the Interference phase data into bins gives us the opportunity to compare the timecourse of BCE development in dual tasks with a new Task 2, in high- versus low-conflict situations. Given the differences observed between groups in the initial Interference BCE analyses, we assessed the two bins (Bins 7 and 8) of Interference data here separately, to test for presence versus absence of BCEs differentially

across groups at each time point. For Bin 7 data, there were no apparent BCEs in any experiment group. ANOVA revealed no effect of response compatibility, and no interaction with experiment group, all F s < 0.7. For Bin 8 (the second half of the Interference phase), a BCE was observed only in the Different-T2 group. ANOVA revealed a significant main effect of response compatibility, $F(1, 89) = 5.906, p < .05$, and a significant interaction with experiment group, $F(2, 89) = 4.710, p < .05$. T tests confirmed that a significant BCE in the Different-T2 group was driving this interaction, $t(29) = 5.141, p < .001$.

For the After data, our prior analyses have shown that BCEs were sufficiently reestablished in all experimental groups over the course of the After phase. Of particular interest is whether we might observe some difference in timecourse of the reestablishment of BCEs in the well-practiced original dual-task paradigm relative to the different degrees of interfering Task 2 experience acquired during the prior Interference phase. Given our initial analyses of After phase data, it is clear that BCEs were quickly reestablished following the Interference phase in all conditions. To observe a difference in timecourse, our most sensitive test is to compare response compatibility effects in the first bin of After data (Bin 9) between our two most extreme groups. We would expect the Different-T2 group to reestablish the original BCE with relatively fewer trials, given predicted minimal interference from prior T2 learning of the *color* task in the Interference phase; at the other extreme, we would expect the Reversed-T2 group to take relatively more trials in the After phase to develop a BCE, given learning of opposite response associations to the same high/low Task 2 in the Interference phase.

From Fig. 6, a BCE appeared well established in the Different-T2 group in the initial After phase data (Bin 9) but not yet well established in the Reversed-T2 group. A 2×2 ANOVA on Bin 9 data revealed a marginal interaction of response compatibility and experiment group, $F(1, 61) = 3.790, p = .056$, modifying a strong main effect of response compatibility, $F(1, 61) = 13.515, p < .001$. T tests revealed no effect of response compatibility in the Reversed-T2 group, $t(32) = 0.860, p = .397$, but a significant effect of response compatibility in the Different-T2 group, $t(29) = 4.159, p < .001$. While somewhat limited by the resolution with which we can measure the development of BCEs over trials in this study, the early After phase data is suggestive of a differential rate of development of BCEs under differing degrees of interference from prior S-R learning in Task 2.

Separate assessment of the Conflict-T2 group also showed a significant effect of response compatibility in Bin 9 data, $t(28) = 2.299, p < .05$. Finally, as suggested by initial After phase analyses, in the latter set of After trials (Bin 10), the response compatibility effect was well established across experiment groups, with a strong main effect of response

compatibility $F(1, 89) = 33.223, p < .001$, and no interaction with experiment group, $F < 0.1$.

Interresponse interval quartile analyses

Finally, we again assessed our 0 ms SOA Task 1 RT data for evidence of BCEs being generated only in trials with very short interresponse intervals, in order to ensure that our BCEs were not produced by response grouping. We determined IRI quartiles for 0 ms SOA trials for each participant and compatibility condition separately for Before, Interference, and After phases in which BCEs were observed (excluding Interference data for Reversed-T2 and Conflict-T2 groups). These data are shown in Table 1. We assessed data from Before and After phases with separate 2 (response compatibility) $\times 4$ (IRI quartile) $\times 3$ (T2 manipulation group) repeated measures ANOVAs, with a third ANOVA without the group variable to assess Different-T2 group data from the Interference phase. In the Before phase, there was a significant interaction of response compatibility and IRI quartile, $F(3, 267) = 12.21, p < .001$. However, follow up t tests showed that the BCE was present in all but the slowest quartile of IRIs, first: $t(91) = 8.32, p < .001$; second: $t(91) = 5.42, p < .001$; third: $t(91) = 3.50, p = .001$; fourth: $t(91) = 1.06, p = .293$. The three-way interaction was not significant, $F < 1.4$. After phase data showed a similar pattern, with a marginal interaction of response compatibility and IRI quartile, $F(3, 267) = 2.10, p = .10$. Follow-up t tests again showed significant BCEs in all but the slowest quartile of IRIs, first: $t(91) = 4.84, p < .001$; second: $t(91) = 4.36, p < .001$; third: $t(91) = 4.63, p < .001$; fourth: $t(91) = 1.60, p = .113$. The three-way interaction was not significant, $F < 1$. For Interference data from the Different-T2 group, there was no interaction of response

Table 3 Mean error rates (%Err) and standard errors of the mean (SE) for 0 ms stimulus onset asynchrony performance in Experiment 2

Phase	Task 1		Task 2		Task 1		Task 2	
	Compatible		Incompatible		Compatible		Incompatible	
	%Err	SE	%Err	SE	%Err	SE	%Err	SE
Reversed-T2								
Before	2.29	0.00	5.09	0.01	10.75	0.01	5.81	0.01
Interference	2.84	0.01	4.73	0.01	17.27	0.02	9.13	0.02
After	1.14	0.00	4.37	0.01	7.56	0.01	4.32	0.01
Conflict-T2								
Before	2.37	0.01	8.24	0.02	7.66	0.01	7.86	0.02
Interference	2.82	0.01	3.72	0.01	13.04	0.03	10.03	0.02
After	0.86	0.00	5.91	0.01	6.41	0.02	4.75	0.01
Different-T2								
Before	2.20	0.01	5.12	0.01	9.56	0.02	8.18	0.01
Interference	16.62	0.02	15.22	0.02	5.78	0.01	6.73	0.01
After	1.90	0.01	6.71	0.01	6.23	0.02	5.59	0.01

compatibility with IRI quartile, $F < 1$. These analyses indicate that any potential response grouping was not the cause of our observed BCEs.

Error rate analyses

Given our focus on 0 ms SOA data for BCE effects, for brevity we again limited our focus on error rate data to 0 ms SOA trials for Task 1 and Task 2. Error rates in 0 ms SOA data were again overall slightly higher than at longer SOAs, likely due to the high degree of task concurrency, and represent the general profile of error performance observed across SOA conditions. Mean Task 1 and Task 2 error rates for 0 ms SOA trials are summarized in Table 3. Trials with an error committed on Task 1, regardless of Task 2 accuracy, and trials with an error committed on Task 2 following accurate Task 1 performance, were submitted to separate 2×3 repeated measures ANOVAs for each phase, with response compatibility (compatible, incompatible) as a within-subjects factor and experiment group (Reversed-T2, Conflict-T2, Different-T2) as a between-subjects factor.

For Task 1 error rate, in the Before phase we observed a main effect of response compatibility, $F(1, 89) = 47.63$, $p < .001$, indicating that compatible trials were more accurate than incompatible trials. This effect was modified by an interaction with experiment group, $F(2, 89) = 3.15$, $p < .05$. Subsequent t tests revealed that the effect of response compatibility was significant for each group: Reversed-T2, $t(32) = 3.85$, $p < .01$; Conflict-T2, $t(28) = 4.31$, $p < .001$; Different-T2, $t(29) = 3.90$, $p < .01$. In the Interference phase, there was a main effect of group, $F(2, 89) = 54.25$, $p < 0.001$, where participants in the Different-T2 group committed substantially more errors (15.9 %) than those in the Reversed-T2 (3.8 %) or Conflict-T2 groups (3.3 %). There were no significant effects involving response compatibility, $F_s < 1$. In the After phase, we observed only a main effect of response compatibility, $F(1, 89) = 32.93$, $p < .001$, with compatible trials more accurate than incompatible trials, as in the Before phase.

For Task 2 error rate, in the Before phase we observed a reversed effect of response compatibility, $F(1, 89) = 5.94$, $p < .05$, where compatible trials were less accurate than incompatible trials, as in Task 2 errors for Experiment 1. This effect interacted with experiment group, $F(2, 89) = 3.40$, $p < .05$. Subsequent paired samples t tests revealed that the effect of compatibility was present only in the Reversed-T2 group, $t(32) = 4.04$, $p < .001$ (other $t_s < 1$). This pattern of errors was repeated in the Interference phase, with better accuracy for incompatible trials, $F(1, 89) = 8.26$, $p < .01$, and a significant interaction of response compatibility and group, $F(2, 89) = 5.11$, $p < .01$. Subsequent t tests showed that this effect of compatibility was again present only in the Reversed-T2 group, $t(32) = 3.39$, $p < .01$, Conflict-T2, $t(28) = 1.54$, $p = .134$; Different-T2, $t < 1$. In addition,

there was a main effect of group, $F(2, 89) = 5.17$, $p < .01$, where participants in the Different-T2 group committed fewer errors (6.3 %) than those in the Reversed-T2 (13.2 %) or Conflict-T2 (11.5 %) groups. Participants in the Different-T2 group were therefore less accurate in Task 1 and more accurate in Task 2 as compared with participants in the other two groups. Finally, in the After phase, we observed only a marginal main effect of compatibility, $F(1, 89) = 3.42$, $p = .068$, again with incompatible trials numerically more accurate than compatible trials.

There are two interesting patterns of error data here. The first is a replication of the reversed compatibility effect for Task 2 error data seen in Experiment 1. We again suggest that these reversed compatibility effects are consistent with previously described partial match effects, well explained by Hommel's (2007, 2009) event-file theory of episodic control of behavior. The second pattern of data relates to an apparent speed-accuracy trade-off in Task 1 data in the Interference phase. We suggest that the pattern of slowing observed across RT for Task 1 and Task 2, and the consistent Task 2 error effects with group interference manipulation, represents task-wide effects of enhanced cognitive control elicited by high-response conflict in Task 2 in our Reversed-T2 and Conflict-T2 conditions (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick & Cohen, 2004). We discuss this possibility and its implications for episodic versus WM-mediated BCE effects further in the General Discussion, below.

General discussion

The present paper sought to test whether the backward compatibility effect is due to episodically-mediated Task 2 response activation in parallel with attended Task 1 performance, or whether Task 2 response activation is mediated via Task 2 rules held in Working Memory during attended Task 1 performance. Experiment 1 directly compared predictions of WM-mediated transient-link, traditional S-R learning direct-link, and episodic direct-link models, under a range of task practice conditions. The WM-mediated model proposed by Ellenbogen and Meiran (2008) to account for BCEs predicted that prior practice of Task 2 should enhance dual-task BCEs along with general RT performance. In contrast, an episodic direct-link model predicted a dissociation of these effects, with the presence and timecourse of development of BCEs sensitive to the extent and context of prior Task 2 learning.

Evidence for an episodic direct-link model of BCE

Prior Task 2 practice was observed to produce rapidly-developing and sustained BCEs versus an absence of BCEs in the same dual-task paradigm, depending on

whether prior Task 2 single-task practice had been performed along with Task 1 of the subsequent dual-task (similar other-task context at practice, prominent BCEs) or with a different task (dissimilar other-task context at practice, no BCEs). Groups without any prior Task 2 practice produced substantial BCEs in the second half of dual-task performance in Experiment 1, suggesting that prior Task 2 practice in a dissimilar context was actively disruptive to the development of BCEs in this experiment. This pattern of data is at odds with strong predictions and assumptions of the WM-mediated model, and also does not fit with the simple Task 2 practice predictions of a traditional S-R learning direct-link model. We suggest that data from Experiment 1 make a strong case in support of the episodic direct-link nature of backward compatibility effects.

In Experiment 2, we more closely investigated the timecourse of development of BCEs over trials, with a number of Task 2 learning and interference manipulations with specific direct-link predictions. A detailed pattern of predicted direct-link BCE effects – progressive development of BCEs over initial practice (Before data), selective interference on BCEs under new Task 2 rules from initial Task 2 learning (Interference data), and differential recovery of initial BCEs depending on overlap with the interference task (After data) – were observed very faithfully in our data, and closely align with the findings in Experiment 1 to support a general direct-link account.

Taken together, our data argue that BCEs are the result of episodic elicitation of Task 2 response activation from related Task 2 stimuli, in parallel with attended Task 1 performance, observable as a response compatibility effect on Task 1 RT when the two tasks overlap sufficiently, typically at very short SOAs in a PRP or similar dual-task paradigm. Data supporting a direct-link account of BCEs is present in Hommel and Eglau's (2002) persistence of BCEs in the absence of relevant Task 2 performance demands (Experiments 3 and 4), in Ellenbogen and Meiran's (2008) finding of increasing BCEs over practice, and in our current detailed pattern of BCE development and interference effects in Experiment 2. Both a traditional S-R learning account and an episodic learning account predict these data well, and do so in preference to a WM-mediated transient-link account. Data from our current Experiment 1 help us further dissociate these models, favouring an episodic direct-link account. From this growing set of studies, we suggest that an episodic direct-link model of BCEs best explains the wide range of data observed.

Effects of working memory on BCE

The effect of WM load on BCEs is not directly addressed by our study, and has been taken as strong evidence in favour of a

WM-mediated transient link account of BCEs. While Ellenbogen and Meiran (2008) demonstrate a convincing methodological effect of WM load on BCEs, we suggest an alternative mechanism for this, considering the otherwise strong evidence for a direct-link model of the BCE. We agree that WM is important in the development of Task 2 automaticity and BCEs, but argue that this is only an indirect relationship, and that WM is not required to represent Task 2 rules along with Task 1 performance to produce the BCE.

Instead, we suggest that WM load influences attended Task 2 performance, where limited executive control and WM capacity under high task demand limits the extent and quality of stimulus- and category-response learning during attended Task 2 performance. This would be seen experimentally as an effect of high WM load disrupting BCEs, though via a very different mechanism than proposed by Ellenbogen and Meiran (2008) – this represents a disruption of direct-link acquisition of automaticity, rather than a disruption of a BCE-producing WM-related mechanism in a transient link model. This account comes very directly from basic predictions of episodic learning and automaticity, and can account for much of the data suggested in support of the WM-mediated transient-link model (e.g., Ellenbogen & Meiran, 2008).

These alternative accounts of how WM load may influence the BCE are directly testable. For example, dual-task training under high WM load followed by a switch to a lower WM load Task 1 should disrupt development of BCEs from Task 2 with a direct-link mechanism, but show a rapid presence of BCEs in a WM-mediated model once participants have available WM capacity. Similarly, lower WM demand training to develop BCEs in a direct-link model should show continued BCEs even with a switch to a more complex Task 1, where a WM-mediated model would predict a sudden absence of BCEs with a switch to a higher WM load.

Cognitive control demands under high interference

In Experiment 2, we observed slower RT and improved accuracy in Task 1 for Reversed-T2 and Conflict-T2 groups, compared to faster but more errorful performance in Different-T2 data. We agree that there are clear differential effects in performance across groups in the Interference phase; however, careful inspection of these data suggests a more interesting and episodically-relevant picture. We suggest that Task 2 in the Interference phase of Reversed-T2 and Conflict-T2 dual-tasks generates a high degree of response conflict in the Interference phase, eliciting a strong task-general cognitive control response (e.g., Botvinick, Braver, Barch, Carter & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004). We speculate that this enhanced task-wide control is the cause of general slowing of the entire dual task, where both Task 1 and Task 2 are subject to stronger inhibitory control and error

monitoring, with a relatively reduced influence of automaticity on overt behavior. In many PRP studies, we often worry when we see the influence of Task 2 difficulty on overall Task 1 RT – we are concerned that participants may be delaying Task 1 performance in order to covertly partially perform Task 2. Analysis of IRIs suggests this is not the case here.

Instead of a speed-accuracy trade-off generated within Task 1 performance in Interference data, our data suggest a general task-wide RT cost due to high Task 2 response conflict across trials in Reversed-T2 and Conflict-T2 groups. This cognitive control delay and enhanced attentional focus then nicely predicts all the error data. For Task 2, higher error rates are seen with higher conflict tasks versus the Different-T2 condition. For Task 1, the increased focus and processing time in Reversed-T2 and Conflict-T2 leads to very accurate performance on the familiar Task 1. In comparison, Different-T2 dual-task performance is not constrained by Task 2-elicited cognitive control influences, and Task 1 becomes more errorful with faster RT. Despite this speeded performance, Task 2 is still less errorful here, reinforcing the large amount of response conflict present in other groups. It is possible that this high degree of cognitive control may also play a role in the absence of BCEs, given the suppression of automaticity often observed in such conditions.

Theoretical issues for the WM-mediated model

Finally, we consider the theoretical implications of a WM-mediated BCE. We agree with descriptions by Ellenbogen and Meiran (2008), Hommel (2009), and others in describing a modern version of the “prepared reflex” – that WM represents task rules or sets that allow the rapid and automatic activation or retrieval of behavior in response to particular stimuli. Ellenbogen and Meiran suggested that holding Task 2 rules in WM during Task 1 performance (along with Task 1 rules) will give a preparation benefit for eventual Task 2 performance; the byproduct of this is the BCE, with automatic Task 2 response activation driven by the Task 2 stimuli via the Task 2 rules in WM. While this arrangement might lead to an eventual Task 2 preparation benefit, we suggest that it would lead to substantial performance costs on general dual-task behavior.

Logan and Gordon’s (2001) ECTVA model provides extensive model demonstrations of exactly these issues in dual-task performance with simultaneous versus sequential activation of task rules in WM, allowing strategically parallel or serial performance, respectively. Simultaneous rule activation for both tasks introduces a classic binding problem – information generated from each task causes interference via crosstalk onto the other, and responding becomes extremely slow and/or errorful. While ECTVA concentrated on stimulus feature overlap for categorization, the response code overlap in BCEs presents a similar issue here. Logan and Gordon (2001)

showed that the ECTVA simulated realistic BCE-like RT effects along with realistic Task 1 and Task 2 dual-task RTs when Task 2 rules were quite weakly represented – this was essentially the model’s strategic serial performance solution to the dual-task binding problem, with Task 1 and Task 2 rules alternately activated, to minimize cross-task interference and allow good speeded performance in each task.

We suggest that the WM-mediated model of BCEs cannot account for all these issues at once. If Task 2 rules are represented in WM to a substantial degree along with Task 1 rules (as suggested, to enhance Task 2 preparation), they should drive the activation of response information in keeping with our “prepared reflex” expectations. This would produce a substantial crosstalk/interference problem and would require an additional powerful cognitive control mechanism to solve this binding problem. In contrast, the data we and other authors typically observe for BCEs suggest a relatively small response compatibility effect, without overwhelming costs on reasonably fast and accurate performance. We suggest there is a substantial mismatch between the proposed degree of representation of Task 2 rules in WM during Task 1 performance, and the observed effects on performance that task rule mediated, prepared reflex-like behavior would predict given well-represented task rules.

Conclusion

In summary, we suggest that the dual-task BCE is the product of episodic automatic activation of Task 2 response information, in parallel with attended Task 1 performance. We suggest that our data present a convincing demonstration of a range of well-predicted direct-link effects on the BCE and provide a clear empirical dissociation of episodic versus WM-mediated BCE models, in favor of the episodic account.

Author Note We thank Karin R. Humphreys, and members of McMaster University’s Cognitive Science Laboratory (Esther Manoian, Mohammed Shariq Wani, Monika Polakowska, Claire Howarth, Rachel Den Boer, Stefania Cerisano, Sudhir Karthikeyan, and Tim Li) for useful discussion and assistance with data collection. This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Grant #327454 to S. W. and by an NSERC postgraduate scholarship (PGS-D) to S. J. T.

References

- 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.
- Crump, M. J., & Milliken, B. (2009). The flexibility of context-specific control: Evidence for context-driven generalization of item-specific control settings. *The Quarterly Journal of Experimental Psychology*, *62*, 1523–1532.

- Crump, M. J., Vaquero, J. M., & Milliken, B. (2008). Context-specific learning and control: The roles of awareness, task relevance, and relative salience. *Consciousness and Cognition*, *17*, 22–36.
- Ellenbogen, R., & Meiran, N. (2008). Working memory involvement in dual-task performance: Evidence from the backward compatibility effect. *Memory & Cognition*, *36*, 968–978.
- Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1368–1384.
- Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494–500.
- Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research*, *71*, 42–63.
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, *73*, 512–526.
- Hommel, B., & Eglau, B. (2002). Control of stimulus-response translation in dual-task performance. *Psychological Research*, *66*, 260–273.
- Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*, 344–638.
- Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 236–259.
- Lien, M.-C., & Proctor, R. W. (2000). Multiple spatial correspondence effects on dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1260–1280.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393–434.
- Logan, G. D., Miller, A. E., & Strayer, D. L. (2011). Electrophysiological evidence for parallel response selection in skilled typists. *Association for Psychological Science*, *22*, 54–56.
- 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 and Performance*, *32*, 149–165.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220–244.
- Pashler, H. E., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. E. Pashler (Ed.), *Attention* (pp. 155–189). Hove, UK: Psychology Press.
- Ruthruff, E., Johnston, J. C., & Remington, R. W. (2009). How strategic is the central bottleneck: Can it be overcome by trying harder? *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1368–1384.
- Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001a). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 3–21.
- 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. (2001b). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, *8*, 73–80.
- Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, *14*, 1–36.
- Thomson, S. J., & Watter, S. (2013). Information continuity across the response selection bottleneck: Early parallel Task 2 response activation contributes to overt Task 2 performance. *Attention, Perception, & Psychophysics*, *75*, 934–953.
- Thomson, S. J., Watter, S., & Finkelshtein, A. (2010). Parallel response selection in dual-task situations via automatic category-to-response translation. *Attention, Perception, & Psychophysics*, *72*, 1791–1802.
- Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology*, *39*, 212–222.
- Tombu, M., & Jolicoeur, P. (2004). Virtually no evidence for virtually perfect time-sharing. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 795–810.
- Van Selst, M. A., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the Psychological Refractory Period effect? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1268–1283.
- Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology*, *46*, 361–413.
- Watter, S., & Logan, G. D. (2006). Parallel response selection in dual-task situations. *Perception & Psychophysics*, *68*, 254–277.
- Welford, A. T. (1952). The ‘Psychological Refractory Period’ and the timing of high-speed performance – A review and a theory. *British Journal of Psychology: General Section*, *43*, 2–19.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.