

# The Dual Mechanisms of Cognitive Control (DMCC) project: Validation of an online behavioural task battery

Quarterly Journal of Experimental Psychology  
1–24  
© Experimental Psychology Society 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/17470218221114769  
qjep.sagepub.com



Rongxiang Tang<sup>1</sup> , Julie M Bugg<sup>1</sup> , Jean-Paul Snijder<sup>2</sup> ,  
Andrew RA Conway<sup>2</sup> and Todd S Braver<sup>1</sup> 

## Abstract

Cognitive control serves a crucial role in human higher mental functions. The Dual Mechanisms of Control theoretical framework provides a unifying account that decomposes cognitive control into two qualitatively distinct mechanisms—proactive control and reactive control. Here, we describe the Dual Mechanisms of Cognitive Control (DMCC) task battery, which was developed to probe cognitive control modes in a theoretically targeted manner, along with detailed descriptions of the experimental manipulations used to encourage shifts to proactive or reactive mode in each of four prototypical domains of cognition: selective attention, context processing, multitasking, and working memory. We present results from this task battery, conducted from a large ( $N > 100$ ), online sample that rigorously evaluates the group effects of these manipulations in primary indices of proactive and reactive control, establishing the validity of the battery in providing dissociable yet convergent measures of the two cognitive control modes. The DMCC battery may be a useful tool for the research community to examine cognitive control in a theoretically targeted manner across different individuals and groups.

## Keywords

Cognitive control; working memory; Stroop; task-switching; context processing

Received: 19 February 2022; revised: 19 June 2022; accepted: 1 July 2022

## Introduction

There is widespread agreement that the capacity for cognitive control is a central element of human adaptability, achievement, and flourishing. While this ability to flexibly regulate, update, and coordinate thoughts and actions in accordance with internally maintained goals is one of humans' most cherished higher mental functions, it is also quite vulnerable to impairment, and even in healthy individuals can vary substantially (Braver, 2012). Importantly, cognitive control is not a unitary process, as it encompasses a diverse range of mental functions spanning different domains of cognition (i.e., attention, working memory [WM], and decision-making) (Kane & Engle, 2002; Miller & Cohen, 2001; Miyake et al., 2000). Consequently, a major challenge has been to characterise and explain cognitive control through a unifying and coherent theoretical framework, which ideally should provide meaningful and operationalisable core constructs that can account for both contextual (i.e., state) and individual (i.e., trait) variations in these functions.

The Dual Mechanisms of Control theoretical framework decomposes cognitive control into two qualitatively distinct mechanisms—proactive control and reactive control (Braver, 2012; Braver et al., 2007). Proactive control refers to a sustained and anticipatory mode of control that is goal-directed, allowing individuals to actively and optimally configure processing resources prior to the onset of task demands. Reactive control, by contrast, involves a transient mode of control that is stimulus-driven and relies upon retrieval of task goals and the rapid mobilisation of

<sup>1</sup>Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA

<sup>2</sup>Division of Behavioral and Organizational Sciences, Claremont Graduate University, Claremont, CA, USA

### Corresponding author:

Todd S Braver, Department of Psychological and Brain Sciences, Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899, USA.  
Email: tbraver@wustl.edu

processing resources following the onset of a cognitively demanding event (Braver, 2012; Braver et al., 2007). In other words, proactive control is preparatory, while reactive control operates in a just-in-time manner. These two mechanisms proposed by the Dual Mechanisms framework have been dissociated based on both their temporal dynamics, computational mechanisms, and neural substrates in healthy and impaired populations (Braver et al., 2005, 2009; De Pisapia & Braver, 2006) and based on their behavioural signatures in young adults (Gonthier, Braver, & Bugg, 2016) and in cross-sectional studies examining age differences (Bugg, 2014a, 2014b; Paxton et al., 2008). Critically, extant empirical findings provide evidence that these two modes of control can be manipulated via distinct situational factors while also pointing to an important source of variation in control function at the individual and group level (i.e., age group or in clinical groups with differing levels of well-being and characteristics), in terms of the bias or preference to adopt one control mode over the other mode (Barch & Ceaser, 2012; Braver, 2012).

In addition to providing a unifying account for understanding intra-individual, inter-individual, and between-groups variability in cognitive control, the Dual Mechanisms framework describes a domain-general account of these two control mechanisms, postulating the presence of proactive and reactive control across multiple cognitive domains. However, there have been empirical and theoretical challenges in developing and optimising valid and reliable paradigms of proactive and reactive control in different task domains, and in establishing behavioural markers that provide robust indices for these two modes of control. The Dual Mechanisms of Cognitive Control (DMCC) project was initiated by our group to address these shortcomings (Braver et al., 2021). A major aim of the DMCC project has been to develop and systematically examine the validity and reliability of a battery of cognitive control tasks across four distinct cognitive domains: selective attention, context processing, multi-tasking, and WM (Braver et al., 2021). Much of our prior work in the DMCC project has been to successfully demonstrate the utility and validity of the DMCC task battery within the neuroimaging environment (Braver et al., 2021; Etzel et al., 2022; Freund et al., 2021; Singh et al., 2022; Tang et al., 2021). In parallel, we have initiated a systematic validation of the full task battery in terms of its behavioural characteristics. In the current article, we focus on group effects, testing for dissociations between behavioural markers of proactive and reactive control. Companion papers examine the psychometric properties of the task battery and its utility for individual difference analyses (Lin et al., 2022; Snijder et al., 2022).

Prior small-scale behavioural studies have focused on individual tasks and conditions within the DMCC battery, such as the AX version of the Continuous Performance Test (AX-CPT), in terms of their ability to isolate proactive and

reactive control (Bugg & Braver, 2016; Gonthier, Braver, & Bugg, 2016; Gonthier, Macnamara, et al., 2016). The goal of the current study is to provide a systematic and rigorous behavioural analysis of all optimised variants of all four task domains, with a large sample size ( $N > 100$ ) and utilizing a within-subject design. Indeed, in most studies of proactive and reactive control to date, whether they are designed to examine individual variation, temporal dynamics, or neural signatures of these two modes of cognitive control, the focus has been on a single task or a limited cognitive domain, and with fairly restricted participant samples. Although the focus on single tasks in measuring the two modes of control has been informative in contributing to our understanding of their mechanisms, it has precluded a rigorous test of the validity and the domain-generalizability of cognitive control modes across multiple cognitive domains.

The DMCC battery includes theoretically optimised adaptations of four well-established cognitive tasks (Stroop, AX-Continuous Performance Test (AX-CPT), Cued Task-Switching and Sternberg WM), one for each of the above-mentioned domains, respectively, that were theoretically optimised to capture variability in proactive and reactive control. Specifically, there were three variants of each task representing different experimental conditions: (1) a baseline condition that maximises within- and between-individual variability, which does not bias the adoption of proactive or reactive control; (2) a proactive condition that shifts individuals towards proactive control; and (3) a reactive condition that independently engages the reactive mode of control. As will be detailed for each task in the later sections, we contrasted theoretically specified behavioural performance patterns across the three variants, to determine whether proactive and reactive control variants did indeed produce the predicted shifts in control.

The present study implemented a multi-session within-subject design to systematically evaluate the validity of the full DMCC task battery. As the first large-sample ( $N > 100$ ) study of the DMCC task battery conducted in a purely behavioural context, it has several advantages and innovative features that had rarely been implemented in prior experimental studies of cognitive control. First, all tasks were computerised and made available through an online platform (Amazon Mechanical Turk [MTurk]) for data collection, which enables recruitment of a large sample size within a short time period while also allowing open accessibility and dissemination of the task battery for future investigations—both in laboratory settings and through online platforms. Second, the study was methodologically innovative in that it includes novel task variants within the battery that have never previously been reported, along with experimental manipulations that induce variability in the utilisation of proactive and reactive control modes affect behavioural performance profiles. Third, the validation of this task battery provides a firm foundation

for the more costly and time-consuming neuroimaging investigations of cognitive control, by identifying the most robust behavioural markers and metrics that can be linked to underlying neural mechanisms (Braver et al., 2021). Finally, this work also provides a foundation for future translational efforts, given that the battery provides (1) assessment tools by which to evaluate the domain-general (unity) and diversity of cognitive control function in different populations, including those with impaired cognitive control; and (2) potential targets for intervention efforts aimed at enhancing proactive and/or reactive control (Braver, 2012).

In the following section, we describe the experimental manipulations and rationale underlying the theoretically targeted variants of all four cognitive control tasks (Stroop, AX-CPT, Cued-TS, and Sternberg WM), specifically highlighting the innovative features of the task conditions included in the battery. This article focuses on the effectiveness of the experimental manipulations at the group level, in independently assessing proactive and reactive control modes. In particular, we compare task performance and primary outcome indices among the three conditions (baseline, proactive, reactive) to evaluate both divergent (discriminant) and convergent (cross-task) validity of the DMCC task battery in capturing variations in the two cognitive control modes. As indicated above, the primary goal of the article is to provide a comprehensive introduction to the DMCC battery and the associated dataset acquired with it, such that the scientific community can fully evaluate and make use of the dataset (which will be made available on public repositories at the time of publication).

## Dual Mechanisms of Cognitive Control task battery

### Stroop

The colour-word Stroop is widely recognised as a canonical task of cognitive control, in which top-down selective attention is required to focus processing on the task-relevant font colour of printed words while ignoring the irrelevant but otherwise dominant word name. A commonly used approach to manipulating cognitive control demands in the Stroop task is to vary *list-wide proportion congruence* (PC) (Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979). Under high list-wide PC conditions, congruent trials (word name matches font colour, for example, BLUE in blue font) are frequent and incongruent trials (word name indicates a different colour than the font colour, for example, RED in blue font) are rare within a block, such that control demands are on average low and intermittent. In contrast, under low list-wide PC conditions (rare congruent trials, frequent incongruent), the high probability that interference will occur within a block should lead to an upregulated cognitive control state.

In particular, we and others have hypothesised that under low list-wide PC conditions, the tendency to utilise *proactive control* will increase (Bugg, 2014a; Bugg & Chanani, 2011; Gonthier, Braver, & Bugg, 2016; Hutchison, 2011; Spinelli et al., 2019). In this case, proactive control is theoretically associated with sustained maintenance of the task goal to attend to the ink colour and ignore the word, which should be present in a consistent (i.e., global; present on all trials) and preparatory manner (i.e., engaged even prior to stimulus onset). Thus, the key prediction is that the Stroop effect (average slowing or increase in errors on incongruent relative to congruent trials) should be reduced on all trials, relative to a baseline, high list-wide PC condition, reflecting improved performance on incongruent trials and a reduction of facilitation on congruent trials (i.e., a congruency cost) (Gonthier, Braver, & Bugg, 2016).

In contrast, PC can also be manipulated in an item-specific, rather than list-wide fashion (Jacoby et al., 2003). In this case, specific colours will occur with low PC (e.g., items appearing in green font will frequently be incongruent), while others may occur with high PC (e.g., items appearing in red font will frequently be congruent), and these “items” are randomly intermixed such that participants cannot predict whether a low PC or high PC item will appear on a given trial. This type of item-specific PC manipulation is theoretically predicted to enhance the utilisation of *reactive control* for low PC items (Bugg & Dey, 2018; Bugg & Hutchison, 2013; Bugg et al., 2011). For these items, strong associations develop between a critical feature (a specific font colour, such as green) and increased control demands (i.e., high interference), leading to more effective goal retrieval and utilisation upon presentation of a stimulus that includes this feature (e.g., a word printed in a green font). The engagement of reactive control is expected to be transient, present only after stimulus onset, and only engaged by low PC incongruent items, particularly when these occur within the context of 50% congruent, or even higher, list-wide PC conditions.

The three Stroop task variants in the present battery varied as follows: the baseline condition had a high list-wide PC (67% congruent, 33% incongruent trials), whereas the proactive condition had a low list-wide PC (33% congruent, 67% incongruent trials). In contrast, the reactive condition approximated the high list-wide PC of the baseline condition (60% congruent, 40% incongruent) due to the inclusion of many high PC (100% congruent) filler items, but also featured specific items that were low PC (25% congruent, 75% incongruent). Another feature of the battery is the inclusion, in each condition, of a *set of unbiased, diagnostic items* (“PC-50,” 50% congruent, 50% incongruent) that did not share features (i.e., words or colours) with the other items in the condition. These PC-50 (diagnostic) items provide clearer behavioural markers from which to dissociate proactive and reactive control

(Braem et al., 2019). Similar versions of these Stroop conditions have been examined in prior work, using both picture-word (Gonthier, Braver, & Bugg, 2016) and colour-word variants (Dey & Bugg, 2021). Finally, it is worth noting that because of the large numbers of different font colours (8) included in each of the conditions, the task was implemented with vocal rather than manual responding, using built-in voice recognition software to extract response latencies.

### AX-CPT

The AX-CPT has become increasingly utilised as a task of context processing and cognitive control, given its simplicity, flexibility, and applicability in a wide range of populations (Barch et al., 2008; Chatham et al., 2009; Chun et al., 2018; Janowich & Cavanagh, 2018; Servan-Schreiber et al., 1996). In the paradigm, participants respond to letters presented one at a time, with each trial consisting of a cue–probe letter pair. When an A-cue is followed by an X-probe, a target response is required. Since the AX pairing occurs frequently, strong cue–probe associations develop. Cognitive control is postulated to be needed to maintain and utilise the information provided by contextual cues, particularly to minimise errors and response interference occurring on BX trials (where B refers to any letter except A), which occur when the X-probe is presented, but is not preceded by an A-cue. In prior work, shifts in the tendency to utilise *proactive* or *reactive control* have not only been observed when comparing different populations or groups, but have also been manipulated within-subjects (Braver et al., 2009).

The AX-CPT conditions included in the battery extend prior recent work using a task variant in which the A- and B-type contextual cues occur with equal frequency, thus eliminating confounds in earlier versions that could be due to the lower overall frequency of encountering B-cues (Gonthier, Macnamara, et al., 2016; Richmond et al., 2015). Furthermore, these conditions also include no-go trials, in which the probe is a digit rather than letter. Because of the increase in response uncertainty (i.e., three types of probe response are possible: target, non-target, no-go), the addition of no-go trials decreases the overall predictive utility of context information for responding, and as a consequence was found to reduce the overall proactive control bias typically observed in healthy young adults. As such, the no-go conditions result in a “low control” baseline, from which to more sensitively observe condition-related changes in control mode (Gonthier, Macnamara, et al., 2016). In all of the current AX-CPT versions tested in this battery, the task structure, trial types, and frequencies are identical, except for the specific manipulations described below for proactive and reactive conditions.

The *proactive* condition replicates prior work using context strategy training (Gonthier, Macnamara, et al.,

2016), as a means of increasing the predictive preparation of responses following contextual cue information. Specifically, participants are provided with explicit information regarding the frequencies of these cue–response associations and receive training and practice in utilising them to prepare the dominant responses. In addition, during inter-trial intervals, participants are provided with visual instructions to “remember to use the strategy.” The key prediction is that the increased utilisation of contextual cue information will lead to a bias to prepare a target response following an A-cue (analysed in terms of both AX and AY trials) and a non-target response following a B-cue, leading to reduced interference on BX trials. Yet a side effect of this preparatory bias is a predicted increase in errors and response interference on AY trials, which occur when the A-cue is *not* followed by an X-probe.

The *reactive* condition involved a new manipulation which has not previously been examined in prior work. Specifically, the reactive condition utilises context-specific probe cueing (similar to other context cueing manipulations in tasks, such as Stroop and flanker; for review, see Bugg and Crump, 2012), in that for high control demand trials (AY, BX, no-go) the probe item appears in a distinct spatial location, and with a distinct border colour surrounding it (presented briefly before the onset of the probe). Critically, because these featural associations are only present at the time of probe onset, they were not hypothesised to modulate the utilisation of proactive control strategies. Similarly, the probe features could not drive direct stimulus–response learning, since they do not directly indicate the appropriate response to be made. In other words, the probe feature cannot be used as a “stop signal,” since on high-control demand trials it signals the need for a go response as often as a no-go. Similarly, on low control demand trials, the probe feature predicts a target response (when it follows an A-cue) as often as it does a non-target response (when it follows a B-cue). In contrast, the probe features do serve as contextual cues signalling high control demand, and thus prompt more rapid and effective retrieval of contextual information to resolve the conflict. Because information about high-conflict probe features is not provided explicitly to participants (in contrast to the proactive condition), it has to be learned implicitly through experience. The key prediction is that utilisation of probe features should reduce the tendency to make BX errors but could increase BX reaction time interference (due to the tendency to utilise the probe to drive context retrieval).

### Cued-TS

Cued-TS has long been recognised as a critical paradigm to assess a core component of cognitive control—the ability to activate and update task representations in an online manner, to configure attention and action systems to process the

task-relevant features of a current target. The key aspect of the paradigm is that two or more tasks randomly alternate across trials, with target items typically being ambiguous, so that they can be processed according to multiple task rules. Consequently, the advance presentation of the task cue, prior to target onset, is what disambiguates the target and specifies the appropriate stimulus–response rules.

An important metric of cognitive control in task-switching paradigms is the task-rule congruency effect (TRCE), which refers to the increased interference (both errors and reaction time) when the target response required for the current task is incongruent with the response that would be required to the same target stimulus if the alternative task had been cued (Meiran & Kessler, 2008). Consider the letter–digit task-switching (also called consonant–vowel, odd–even [CVOE]) task comprising a letter task and a digit task. If in the letter task, a right button press is required for a consonant and a left button press for a vowel, while in the digit task, a right button press is required for odd and a left button press for even, the “D4” target stimulus would be incongruent (whereas the “A2” target stimulus would be congruent, since for either task, the left button press would be correct). Two additional important metrics are switch costs, which refer to the decrement to performance when the task to be performed on the current trial switches from that on the previous trial (relative to task-repeats, when the same task is performed on two consecutive trials) (Meiran, 1996; Rogers & Monsell, 1995), and mixing costs, which refer to the decrement to performance that occurs on task-repeat trials (relative to performance within a single-task block) (Braver et al., 2003; Los, 1996). These have also served as indices of cognitive control demands.

In prior work, including reward incentives on a subset of trials, with reward cues presented at the time of the task cue, led to a strong reduction in the mixing cost—and this was present even on the trials that were non-incentivised—but there was no effect on the TRCE (Bugg & Braver, 2016). This finding was interpreted as indicating that the mixing cost reductions reflected a list-wide (global) enhancement of *proactive control*, whereas the TRCE effect is primarily influenced by *reactive control*, and so less affected by advance reward incentive manipulations. The Cued-TS conditions included in the current battery build on this prior work by using variants of the CVOE (letter/digit) paradigm that aims to accentuate the robustness of the TRCE while also enabling clear utilisation of *proactive control* through the use of advance task cues with a long cue-to-target interval (CTI). A robust finding from prior work is that performance improves with longer preparation times (CTI), suggesting advanced preparation for relevant task rules and stimulus–response mappings for the upcoming target (Meiran, 1996).

In the baseline condition, target stimuli are list-wide mostly congruent (LW-MC; 67%), as prior work has found that mostly congruent conditions result in a large and robust

TRCE (Bugg & Braver, 2016). The *proactive* condition builds on Bugg and Braver’s (2016) study in keeping the same LW-MC structure as the baseline condition but adding reward incentives on a subset of trials. Specifically, on 33% of trials, reward cues are presented simultaneously with advance task cues (i.e., by presenting the task cue in green font), and indicate the opportunity to earn monetary bonuses if performance is accurate and fast (relative to baseline performance) on that trial. By only presenting reward cues on a subset of trials, the remaining subset of non-incentivised trials and target stimuli can be directly compared across the proactive and baseline conditions. A divergence from Bugg and Braver (2016) is that single-task conditions are not included as part of the battery (due to length constraints), which precludes direct calculation of mixing costs. Nevertheless, the key prediction is that enhanced *proactive control* will lead to a global improvement of performance (i.e., faster response times [RTs] without a loss in accuracy).

The *reactive* condition utilises a new manipulation which has not previously been examined in prior work. Specifically, the reactive condition includes punishment (rather than reward) incentives, again on the same 33% subset of trials that were incentivised in the proactive condition. However, in the reactive condition the incentive cue is presented at the time of the target stimulus, rather than with the task cue, which precludes the use of incentive motivation in a preparatory fashion. Participants are instructed that they will lose a component of their potential monetary bonus if they make an error on these incentivised trials. Critically, the incentivised trials occur preferentially (75%) with incongruent target stimuli. This manipulation is intended to associate punishment-related motivation with these high-conflict items, potentially leading to increased response monitoring and caution when incongruence is detected. As such, the key prediction is that enhanced reactive control should reduce the error TRCE, even on the non-incentivised trials, when compared with baseline and proactive conditions. Conversely, the RT TRCE should be increased, due to the tendency to utilise target features (detection of incongruency) to drive retrieval of task rules.

### Sternberg WM

The Sternberg item-recognition task has been one of the most popular experimental paradigms used to assess short-term/WM for over 50 years (Sternberg, 1966), but more recently has been adapted particularly for the study of cognitive control with the “recent probes” version (Jonides & Nee, 2006). Like standard versions of the paradigm, the recent probes version presents participants with a memory set of various load levels (number of items), to maintain over a short delay (retention period), after which a single item probe is presented, which requires a target response if

the probe was a part of the memory set. A classic finding in the literature is that as the memory set increases in size, WM load increases, and performance declines accordingly (higher error rates, longer RTs) (Shiffrin & Schneider, 1977; Sternberg, 1966). Under conditions in which the WM load is below capacity (3–4 items), active maintenance and rehearsal processes can be used to keep the memory set accessible, as an attentional template from which to prospectively match against the probe item (i.e., utilising *proactive control* strategies). In contrast, when the WM load is above capacity (~7 items), probe responses are likely to be driven by retrieval-focused processes, such as familiarity (i.e., *reactive control* strategies).

In recent probes versions, the key manipulation is that the probe item can also be a part of the memory set of the previous trial, but not the current trial, which is termed a “recent negative” (RN) probe. On these RN trials, the probe is associated with high familiarity, which can increase response interference and errors, unless cognitive control is utilised to successfully determine that the probe familiarity is a misleading cue regarding its status (target or non-target). The current versions of the Sternberg WM paradigm included in the battery are adapted from previous studies (Burgess & Braver, 2010; Speer et al., 2003), in using manipulations of WM load expectancy and RN frequency. Specifically, in all conditions, trials randomly vary in set size, with words used as stimuli, such that all items are novel on each trial, with the exception of RN probes. Under such conditions, Burgess and Braver (2010) found strong RN interference effects in both RT and errors. Similarly, following Speer et al. (2003), the set size in a given trial is revealed sequentially, leading to unpredictability and reliance on WM load expectancies to engage control strategies.

In the baseline condition, most trials have high WM load (6–8 items; 60%) and RN frequency is low (20% of non-target probes), which should reduce tendencies to engage either proactive or reactive control strategies. However, in the *proactive* condition, most trials have low WM load (2–4 items; 60%), leading to the expectancy that active maintenance-focused and proactive attentional strategies will be effective, while RN frequency remains low (matched at 20% non-target probes), such that the utility of reactive control should be unchanged. The critical prediction concerns the five-item set size which occurs equivalently in all conditions (40% of trials), and thus can be directly compared between them. The key hypothesis is that use of proactive control strategies will improve both RT and accuracy, primarily for the target probe items (termed novel positive, or NP, since they never overlap across trials).

In the *reactive* condition, WM loads are identical to the baseline condition, while the frequency of RN trials is increased (80% of non-target probes). Thus, in the reactive condition, it is familiarity-based interference expectancy that increases, rather than WM load expectancy. Based on the increased interference expectancy, the theoretical

hypothesis is that participants will not rely on familiarity as a cue for responding, and will rather evaluate the match of the probe to items stored in WM. Consequently, the key prediction is that performance on RN (or rather the RN effect, computed by subtracting performance on novel negative or NN trials) will be significantly improved relative to baseline.

## Methods

### Participants

Participants were recruited for the study via the MTurk online platform. The TurkPrime interface was used to post study descriptions, manage recruitment and payment, send out reminder emails, and handle all other communication with the participants. After reading a description of the study that indicated its multi-session nature and time commitment, interested participants accessed a link which allowed them to review and sign the consent form. After signing the consent, the web links for the first session of the study were made available over MTurk. Participants were instructed to use a computer or laptop for completing the sessions, as the tasks were not designed to work with mobile device or tablet. The study protocol was approved by the Institutional Review Board of Washington University, St. Louis. All data were collected across two separate testing waves held a few months apart; however, since the procedures for the tasks described below were identical across waves, the data from both waves are aggregated for reporting purposes below.

A total of 278 participants signed up for the study and 129 participants with complete data were included in the analyses. Participants were excluded for not completing sessions in the required period of time, for technical problems that precluded data analysis (particularly for the Stroop task, which involved vocal responses), or for data that indicated a failure to comply with task instructions. Participants were not restricted with regard to age; consequently, the final included sample of participants had a wide age range (22–64,  $M = 37.11$ ,  $SD = 9.90$ ; 82 females, 47 males). Data were analysed separately for each task, and only for participants that had complete useable data for that task; thus, tasks are not equivalent in terms of sample size (AX-CPT: 121, Stroop: 126, Cued-TS: 128, Sternberg: 128).

### Design and procedure

The study protocol consisted of 30 separate testing sessions that subjects completed in a sequential manner (15 for the test phase, and another 15 for retest). Participants were asked to complete the sessions at a rate of 5 per week, that is, 6 weeks to complete the full protocol. Each session lasted approximately 20–40 min in duration, with the exception of the first session, which was 1 hr in duration (and included a

Stroop practice block to validate operation of vocal response recording, plus a battery of demographic and self-report questionnaires). To both incentivise and prorate study completion, completion of the first session in each set of 5, for both test and retest phases, resulted in a US\$4 payment, and the others resulted in a US\$2 payment. Additional bonuses of US\$20 were paid for completion of the test phase and US\$30 for full study completion. Together, successful completion of the entire protocol resulted in a payment of US\$122, plus additional monetary bonuses associated with incentives in the Cued-TS sessions.

Each set of five sessions was posted at the beginning of the week through MTurk and sent through emails to the participants. Two reminder emails were also typically sent during the week to remind subjects of the completion deadline for the set (by the end of the week). If subjects failed to complete the weeks' sessions by the designated deadline, they were not invited back to participate in subsequent sessions. For each completed session, subjects would enter in a completion code and the experimenter would review each session results for completion and approve the payment within a week through TurkPrime. If subjects dropped from the study, they still received prorated payment for all sessions completed.

For each completed session, the experimenter checked for overall accuracy and completion of each task and questionnaire to make sure that subjects were complying with instructions and maintaining sufficient attention to the task. A criterion of 60% accuracy and response rate was used to determine whether the data would be included, and the subject invited to remain in the study. For each task or questionnaire that did not meet the criterion, the experimenter attempted to communicate with the subject first to determine whether they had trouble understanding the instructions or had technical difficulties. If so, the subject was given a second chance to complete the task before a designated deadline.

Within each of the test and retest phases, sessions were conducted in a fixed order for all participants, with the baseline conditions of all tasks performed first, followed by reactive conditions of all tasks and then finally proactive conditions. The AX-CPT, Cued-TS, and Sternberg were programmed with in-house JavaScript code (available upon request at <https://sites.wustl.edu/dualmechanisms/request-form/>), while the Stroop task was programmed and delivered using Inquisit software, as it included capabilities for online vocal response recording (script also available at link above).

## Tasks

In the following sections, we describe each task and its variants. Figure 1 is a general illustration of the four tasks and their associated experimental manipulations in each variant.

**Stroop.** In this vocal Stroop task, colour words are presented in coloured font and participants name the font colour out loud. For each trial, vocal response latencies were recorded, and the spoken word was detected using the computer's built-in voice recognition software. Accuracy was then automatically coded through the Inquisit software. Participants were given standard instructions to respond as quickly as possible (in a normal voice) while retaining accuracy. Adequacy of the automated voice recognition was validated in previous pilot testing, and individually for each subject based on their first testing session, which contained a practice block of 25 standard Stroop trials. If responses could not be detected for most of the trials, the subject was not asked to continue with further testing.

The current versions of the Stroop were based on the design of our previously reported work (Gonthier, Braver, & Bugg, 2016), and constructed using two different sets of four colours, in which the relative proportion of congruent and incongruent trials was manipulated in different ways. One set of four colours (red, blue, purple, white) was *biased* in the proportion of congruent and incongruent trials, either mostly congruent or mostly incongruent (MI), varied across conditions. The other set (black, green, pink, yellow) was an unbiased/diagnostic set in that the proportion of congruent to incongruent stimuli was 50:50 (hereafter, this set is termed *PC-50* items). The two sets of stimuli were non-overlapping, such that on incongruent trials, the word name was one of the three remaining colours from that set (e.g., green font with "black," "pink," or "yellow"; red font with "blue," "purple," or "white"). All trials consisted of the following stimulus parameters: items were presented centrally on a grey screen for 5,000-ms duration or until a response was detected, followed by a 250-ms inter-trial interval during which a blank screen was presented.

**Baseline session.** In the baseline session, the trials were manipulated in an LW-MC manner. Participants completed a total of 288 trials during the baseline session, in which there were 96 PC-50 trials (48 congruent, 48 incongruent) and 192 biased trials. The biased set had 75% congruent (144 trials) and 25% incongruent (48 trials) trials. Consequently, the list-wide proportion congruency for the baseline session was 67%. The session was divided into two blocks of 144 trials each, between which participants were instructed to rest for 1 min. Participants practised a slightly simpler version of the baseline condition in a practice block on the first session, to validate that response latencies could be accurately captured.

**Proactive session.** In the proactive session, the trials were manipulated in a list-wide, mostly incongruent (LW-MI) manner. Participants completed a total of 288 trials during the proactive session, in which there were 96 PC-50 trials

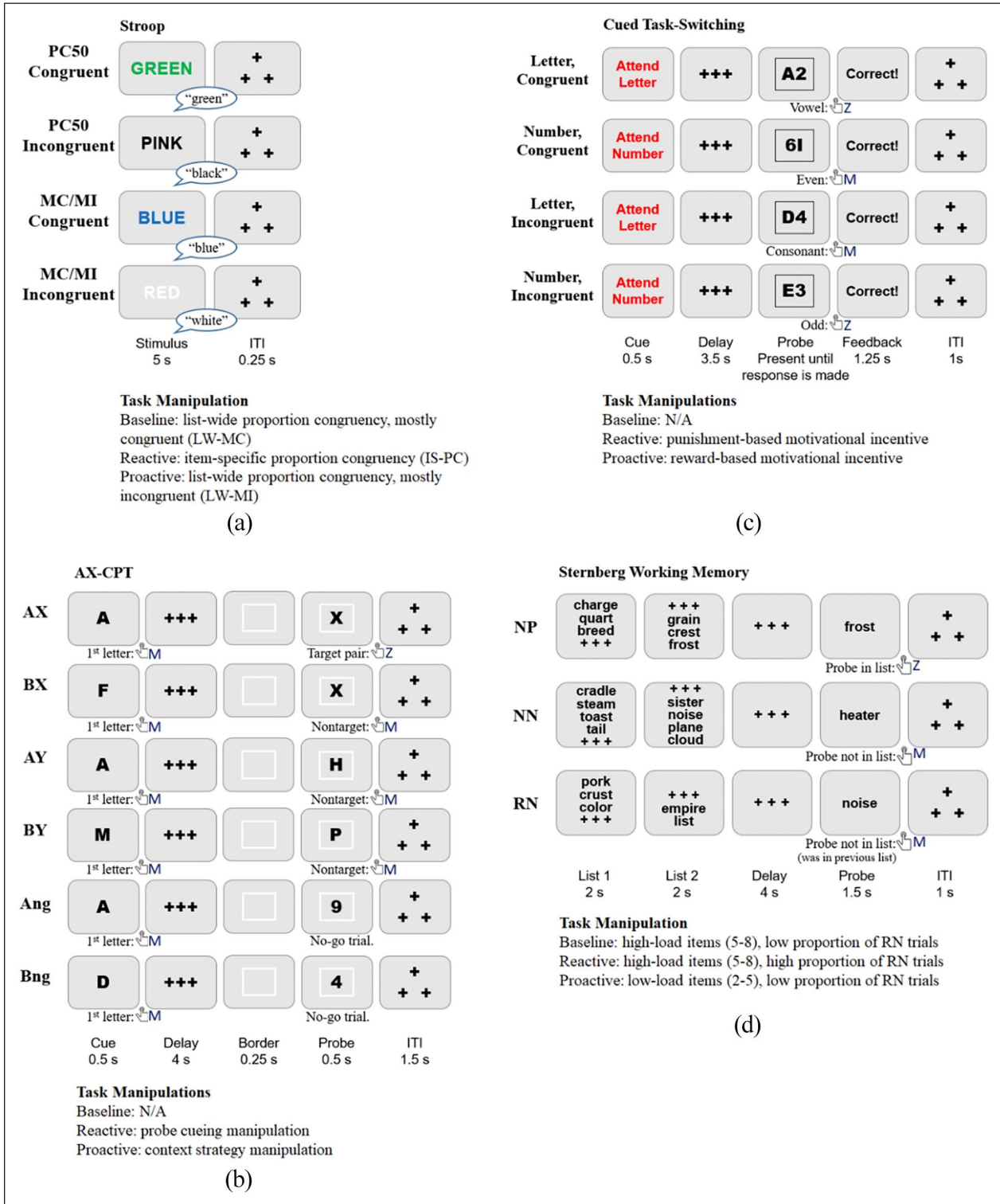


Figure 1. The DMCC battery: (a) Stroop, (b) AX-CPT, (c) Cued Task-Switching, and (d) Sternberg working memory.

(48 congruent, 48 incongruent), and 192 biased trials. The biased set had 25% congruent (48 trials) and 75% incongruent (144 trials) trials. Consequently, the list-wide proportion

congruency for the proactive session was 33%. The session was divided into two blocks of 144 trials each, between which participants were instructed to rest for 1 min.



**Reactive session.** In the reactive session, the proportion congruency manipulation was at the item level, i.e., item-specific proportion congruency (IS-PC). Specifically, blue and red colour font items were manipulated to be biased trials with PC-100 (i.e., these font-colour words were only presented as congruent trials; 192 trials). Purple and white colour font items served as biased trials, with PC-25 (i.e., 25% congruent, 48 trials; 75% incongruent, 144 trials). Finally, as in the baseline and proactive conditions, the remaining 96 trials were unbiased trials with PC-50 (i.e., equal amount of congruent and incongruent trials). Thus, subjects completed a total of 480 trials during the reactive session. The session was divided into three blocks of 160 trials each, between which subjects were instructed to rest for 1 min.

**Cognitive control measures.** Average response times (RTs) on correct trials and error rates were calculated for both congruent and incongruent trials for each subject in each session. The key measure of cognitive control was the Stroop interference effect (incongruent–congruent). We focus primarily on the PC-50 items (since these were matched across conditions) and RT (as is standard in the literature), though we also examined biased items and error rates for both types of items. To directly compare proactive and reactive conditions, two additional derived indices were also calculated: the transfer cost and congruency cost (see Gonthier, Macnamara, et al., 2016, for further descriptions). The transfer cost was computed as the difference in Stroop RT interference on PC-50 items relative to biased items; the congruency cost was computed by subtracting the baseline congruent trial RT from the congruent trial RT in proactive and reactive conditions (again focusing on PC-50 items, but also computed for biased items).

**AX-CPT.** In this version of the AX-CPT, participants make button press responses to visually presented cue–probe pairs. A target key press (“Z”) is made to the probe on AX trials; a non-target key press (“M”) is made to the probe on the other non-target (AY, BX, BY) trials, as well as to the cue on all trials. In addition to the four primary trial types, the task also includes no-go trials, which require withholding response to the probe; no-go trials are indicated by a digit (1–9) rather than letter probe. The task comprised 216 trials in total and included 72 AX trials, 72 BY trials, 18 AY trials, 18 BX trials, and 36 no-go trials (18 following an A-cue, 18 following a B-cue). All trial types and no-go trials were presented in random order. The task was performed in three 72 trial blocks, between which subjects were instructed to take a minimum of 1 min rest break. All trials consisted of the following parameters. The cue was presented centrally on a white screen for 500 ms duration. After a 4,000-ms blank cue–probe interval, the target (in same size font) was presented for 500 ms but immediately preceded by a 250-ms period during which a bounding box

was presented. A 1,500-ms inter-trial interval ended the trial (indicated by a central triangle of fixation crosses).

**Baseline session.** The baseline session identically followed the description above. After receiving task instructions, subjects performed a 12-trial practice block before beginning the actual task.

**Proactive session.** In the proactive condition, participants received strategy training before completing the AX-CPT. The strategy training occurred during a practice block consisting of two phases. In the first phase of six trials, an audio clip was played, which instructed subjects which button to prepare following the cue. In the second phase of six trials, after the cue was presented, they were asked to type which button they were preparing to press in response to the second item. Participants typed out “left” or “right” and the program told subjects if they were correct or not. If they were not correct, they were reminded what letter the first item was and asked to try again. This procedure was implemented to accommodate the online testing format, and deviated slightly from in-person versions, in which subjects responded verbally regarding the button they were preparing to press. In addition, during the test phase, in the inter-trial interval periods, subjects were given the visual message to “Use the strategy!” Otherwise, task structure was identical to the baseline session.

**Reactive session.** The occurrence of high-conflict trials (AY, BX, no-go) was implicitly signalled by presenting the probe in a distinct spatial location and preceded by a distinct border colour. Specifically, while cues were always presented centrally (as in the baseline and proactive conditions), the probe stimuli were presented either in the upper half (AX, BY) or in the lower half (AY, BX, no-go) of the visual display. Furthermore, probe stimuli were immediately preceded (250 ms before probe onset) by either a white border (AX, BY) or red border (AY, BX, no-go). Otherwise, the task structure and trial proportions were identical to baseline and proactive sessions.

**Cognitive control measures.** Average RTs on correct trials and error rates were calculated for each of the four primary trial types (AX, AY, BX, BY) for each subject in each session. Average error rates for no-go trials were calculated as well. The key measure of cognitive control was BX probe interference, which is calculated as the difference score on B-cue trials (BX – BY). This index allows for examination of the interference that occurs when an “X” probe follows a non-target cue “A” and a target trial response must be inhibited. We focused on BX probe interference in both errors and RT. To directly compare proactive and reactive conditions, we also computed an additional derived index, the A-cue bias. The A-cue bias measure reflects the bias to make a target response following an A-cue, and is

calculated by computing a  $c$  criterion from hits on AX trials and false alarms on AY trials as  $1/2 \times (Z[H] + Z[F])$ , with H representing hits on AX trials and F representing false alarms on AY trials (Richmond et al., 2015). Because BX probe interference and A-cue bias involve different trial types, they can be examined fully independently, which is useful when directly comparing proactive and reactive control conditions.

Although not a primary focus of interest in the current report, additional derived indices were computed and reported to maintain continuity with prior work:  $d'$ -context and the Proactive Behavioral Index (PBI; Gonthier, Macnamara, et al., 2016). The  $d'$ -context index was calculated by computing a  $d'$  index from hits on AX trials and false alarms on BX trials as  $Z(H) - Z(F)$ , with H representing hits on AX trials, F representing false alarms on BX trials, and Z representing the z-transform of a value (Servan-Schreiber et al., 1996). The PBI is calculated as  $(AY - BX)/(AY + BX)$  (Braver et al., 2009). This index reflects the relative balance of interference between AY and BX trials; a positive PBI reflects higher interference on AY trials, indicating proactive control, whereas a negative PBI reflects higher interference on BX trials, indicating reactive control. The PBI was computed separately for error rates (based on average error rates on AY and BX trials) and for RTs (based on average RTs on AY and BX trials). To correct for error rates that were equal to 1.00, a log-linear correction was applied to all error rate data prior to computing the  $d'$ -context, the A-cue bias, PBI, and BX interference (Braver et al., 2009; Hautus, 1995). This correction was applied as  $\text{error rate} = (\text{number of errors} + 0.5) / (\text{number of trials} + 1)$ .

**Cued-TS.** In the current Cued-TS paradigm, we used the letter–digit task, which involves bivalent target stimuli consisting of a letter and a digit (e.g., E3). On each trial, the subject is cued to perform either a letter task—consonant/vowel discrimination—or a digit task—odd/even discrimination (Minear & Shah, 2008; Rogers & Monsell, 1995). For the letter task, consonants required a right key press (“M”) and vowels required a left key press (“Z”). For the digit task, even numbers required a right (“M”) key press and odd numbers required a left (“Z”) key press. At the start of every trial, the task is cued by an on-screen message that indicates either “ATTEND LETTER” or “ATTEND NUMBER,” indicating whether attention and responding should be based on the letter or digit, respectively. Critically, because of the response mappings, certain stimuli are congruent, in that they require the same key press irrespective of the relevant task rule (e.g., H6, E3), while other stimuli are incongruent, in that the two tasks were associated with different required responses to the same target (e.g., I6, D4).

The target stimuli were constructed in terms of two distinct stimulus sets. One set of stimuli (A1, A2, B1, B2, 1A,

2A, 1B, 2B) was mostly congruent (80% congruent; 20% incongruent). The second set of stimuli (D4, E3, H5, I6, 4D, 3E, 5H, I6) was unbiased (50% congruent, 50% incongruent). Trials randomly alternated between an equal number of “ATTEND LETTER” and “ATTEND NUMBER” trials. Due to the random presentation order of the cues, switch and repeat trials were on average equivalent, but deviated slightly in number across conditions and subjects. Each session consisted of 192 total trials, 96 mostly congruent (80 congruent, 16 incongruent) and 96 unbiased (48 congruent, 48 incongruent) and also equally split between the two tasks (i.e., 96 letter, 96 digit). Trials were separated into three 64 trial blocks, between which subjects were required to take a minimum of 1 min rest break. Prior to starting each session subjects learned (or refreshed their memory) of these response mappings through a set of 16 practice trials. All trials consisted of the following stimulus parameters: trial initiation with a 300-ms alerting cue (flashing cross), followed by the task cue presented on a grey screen for 500-ms duration. After a 3,500-ms blank CTI, the target was presented until a response was made. The response was followed by a 1,250-ms feedback period and then a 1,000-ms inter-trial interval (indicated by a central triangle of fixation crosses).

**Baseline session.** In this condition, no manipulations were made to the unbiased stimuli. All task cues appear in red font and task stimuli appear in black font.

**Proactive session.** The proactive version of Cued-TS was identical to the baseline version except for the addition of a reward-based motivational incentive. This motivational incentive involved presenting subjects with a reward cue (green font) indicated during presentation of the task cue. When subjects responded to incentive trials faster than the baseline session’s median RT, while maintaining accuracy (this information was stored in a look-up table database, and accessed at the beginning of each session), they received a monetary bonus for that trial added to their compensation amount. Before the start of the proactive sessions, participants were informed by the instructions that they can obtain more payment on top of regular compensation by responding faster than before and maintaining accuracy on incentive trials, which are preceded by a green cue. Non-incentive trials were indicated by the task cue appearing in red font. Only the unbiased set of stimuli were incentivised (66% of unbiased, 33% of total, 64 trials) and presentation order was random with respect to the task cue and target stimuli pre-determined pairs. Subjects received feedback on all trials. The word “Reward!” appeared on the screen for 1,250 ms if the subject earned the reward. If subjects were too slow or made an incorrect response, the words “Too Slow!” or “Incorrect!,” respectively, appeared on the screen. The non-incentive trials also included feedback, showing “Correct” or “Incorrect” after each trial.

**Reactive session.** The reactive version of Cued-TS was identical to the baseline version except for the addition of a punishment-based motivational incentive. This motivational incentive involved presenting subjects with a punishment cue, that was indicated during presentation of the target stimulus (green font). When subjects made errors on incentive trials, they received a monetary penalty for that trial that was subtracted from their compensation amount. Before the start of the reactive sessions, subjects were informed by the instructions that if they make an error on incentive trials, they would be penalised strongly, with 25 cents of their potential bonus taken away for each error. Non-incentive trials were indicated by the target stimulus appearing in black font, while incentive (i.e., punishment) trials were indicated by the target stimulus appearing in green font.<sup>1</sup> Only the unbiased set of stimuli were incentivised, and these were applied in an item-specific manner such that all of the incongruent stimuli (H5, 6I, 5H, 6I; 48 trials) were incentivised, while only 33% of the congruent stimuli were associated with incentives (D4, E3, 3E, 4D; 16 trials). The sentence “Loss of 25 cents!” appeared on the screen for 1,250 ms if the subject made an incorrect response. If subjects were correct, the word “Correct” would appear on the screen. The non-incentive trials also included feedback, showing “Correct” or “Incorrect” after each trial.

**Cognitive control measures.** Average RTs on correct trials and error rates were calculated separately for congruent/incongruent trials, and for both the non-incentivised (biased) items and the incentivised (unbiased) items, for each subject in each session. A key measure of cognitive control is the TRCE (Meiran & Kessler, 2008), which is the difference between incongruent and congruent trials. We focus primarily on the non-incentivised items since these can be most straightforwardly compared across proactive and reactive conditions, although we also examined and report effects on incentivised items. In addition, although not a primary focus of interest in the current report, we also report the switch cost as another index of cognitive control. The switch cost is calculated by subtracting task-repeat trials from task-switch trials.

**Sternberg.** In the current Sternberg item-recognition task, participants are presented with a new, short list of words on each trial that served as a memory set (e.g., “WINE,” “SPLIT,” “GRILL,” “INTENT”). After encoding and a retention interval delay period, a probe item is presented, which requires a judgement as to whether it was part of the current trial’s memory set (i.e., a positive probe) or not (i.e., a negative probe). If the word was in the most recent list, a left key press (“Z”) is required. If the word was not in the most recent list, a right key press (“M”) is required. The current versions of the Sternberg were based on the design of Speer et al. (2003) and constructed using two

distinct sets of memory set items: critical items, had a constant memory set of five words; the other, variable-load set consisted of either low-load items (memory sets of 2–4 words) or high-load items (memory sets of 6–8 words). In addition, the probe consisted of three trial types: (1) NP, (2) NN, and (3) RN.

Each session consisted of 120 total trials, broken down into 48 critical items and 72 variable-load items. Trials were separated into three 40 trial blocks, between which subjects were required to take a minimum of 1 min rest break. Prior to starting each session, subjects learned (or refreshed their memory) of the task through a set of 10 practice trials. All trials consisted of the following stimulus parameters: visual presentation of the memory set across two encoding screens each of 2,000 ms duration; in the first screen, were presented above a central fixation cross, and in the second screen, below the cross. Following memory set presentation, a retention interval of 4,000 ms was presented (during which the fixation cross remained on screen), followed by 1,500 ms presentation of the probe item, and then a 1,000-ms inter-trial interval.

**Baseline session.** The baseline session involved high-load variable items and a low proportion of RN trials (20% of negative probes; 10% of total trials). Specifically, the variable-load set consisted of a mixture of high-load memory sets (12 six-item, 24 seven-item, 36 eight-item) and very few RN trials (4 RN, 32 NN, 36 NP). For the critical five-item set, the proportion was slightly adjusted, to increase the number of RN trials for analysis (8 RN, 16 NN, 24 NP).

**Proactive session.** In the proactive session, the variable-load items were instead a mixture of low-load memory sets (36 two-item, 24 three-item, 12 four-item). The proportion of RN, NN, and NP trials was identical to the baseline session for both variable-load (4 RN, 32 NN, 36 NP) and critical item sets (8 RN, 16 NN, 24 NP).

**Reactive session.** In the reactive session, the variable-load set used the identical mixture of high-load memory set items as the baseline session (12 6-item, 24 7-item, 36 8-item). However, the relative proportion of RN to NN trials was increased in both the variable-load (32 RN, 4 NN, 36 NP) and critical items (16 RN, 8 NN, 24 NP).

**Cognitive control measures.** Average RTs on correct trials and error rates were calculated per trial type (i.e., NN, NP, RN trials), and separately for critical items (five-item lists) and non-critical items (collapsed across the remaining list lengths). The key measure of cognitive control was the recency effect, which is calculated as the difference score on negative trials (RN–NN; Jonides & Nee, 2006). We focused on critical item performance, both in errors and RT and in terms of the recency effect, since these are most

easily compared across proactive and reactive conditions, though we also report findings on non-critical (high- or low-load) items as well.

**Data pre-processing and analysis.** The data were pre-processed in two steps: (1) removal of extreme outliers and (2) winsorisation of remaining outliers. In Step 1, all 128 subjects were screened for data abnormalities such as extremely slow RTs or high error rates. RT plots were examined and cut-off decisions were made for each task separately. Trials with RTs slower than the cut-off threshold were discarded. The threshold for Stroop was 4,000 ms; no RTs on correct trials surpassed the threshold. The threshold for AX-CPT was 2,000 ms; no RTs on correct trials surpassed the threshold. The threshold for Cued-TS was 5,000 ms and resulted in 0.3% of the task's data discarded. The threshold for Sternberg was 3,000 ms; no RTs on correct trials surpassed the threshold. After discarding trials with these RT outliers, the number of trials per condition remained sufficient for analyses.

In Step 2, a winsorisation procedure was conducted on RT data at the trial level (i.e., data split by phase, session, trial type, and subject). The winsorisation parameters for RTs were as follows: RTs lower than 200 ms were replaced by RTs of 200 ms and RTs above the mean plus 3 standard deviations were replaced by RTs of the mean plus 3 standard deviations. Across the four tasks, 1.9% of RT observations were adjusted by the procedure. The adjustments did not vary considerably across tasks, sessions, or trial types. For error rate, the winsorisation procedure was conducted at the level of trial type (data split by phase, session, and trial type), instead of at the subject level, which was examined in the first step of pre-processing. Following the cut-off used by Gonthier, Macnamara, et al. (2016), error rates above 40% were replaced with error rates of 40%. This resulted in nearly 5% of error rates being adjusted for the AX-CPT and Sternberg tasks (i.e., 4.78%, 4.69%, respectively). The Stroop and Cued-TS adjustments were much lower at .07% and 1.69%, respectively. Examining this more carefully revealed repeated subpar performance for some subjects (e.g., greater than 80% error rate in some conditions, large proportion of observations without responses) which inflated the winsorisation adjustment rates. Those subjects were excluded from the final sample. It should be noted that for all tasks (with the exception of no-go trials in the AX-CPT), trials in which no response was recorded were treated as incorrect trials. Finally, we retained 126 subjects for Stroop, 121 for AX-CPT, 128 for Cued-TS, and 126 for Sternberg.

After these pre-processing steps, statistical inference was conducted both within and across conditions using paired *t*-tests. Both classical frequentist and also Bayesian analyses were conducted and both sets of results are reported, in terms of both effect sizes (Cohen's *d*) and Bayes Factors (BFs). The reported *t*-values are always

positive when the pattern followed the predicted pattern; thus, a negative *t*-value refers to a pattern than went opposite to that predicted. We refer to an effect as having strong evidence in favour of the hypothesis when  $BF > 10$ , and strong evidence for a null effect with  $BF < 0.1$ . In cases where an effect yielded statistical significance via classical frequentist conventions (i.e.,  $p < .05$ ), but with  $BF < 10$ , we refer it as significant but lacking strong evidence. Data and analysis code are publicly available at <https://osf.io/pqvg/>.

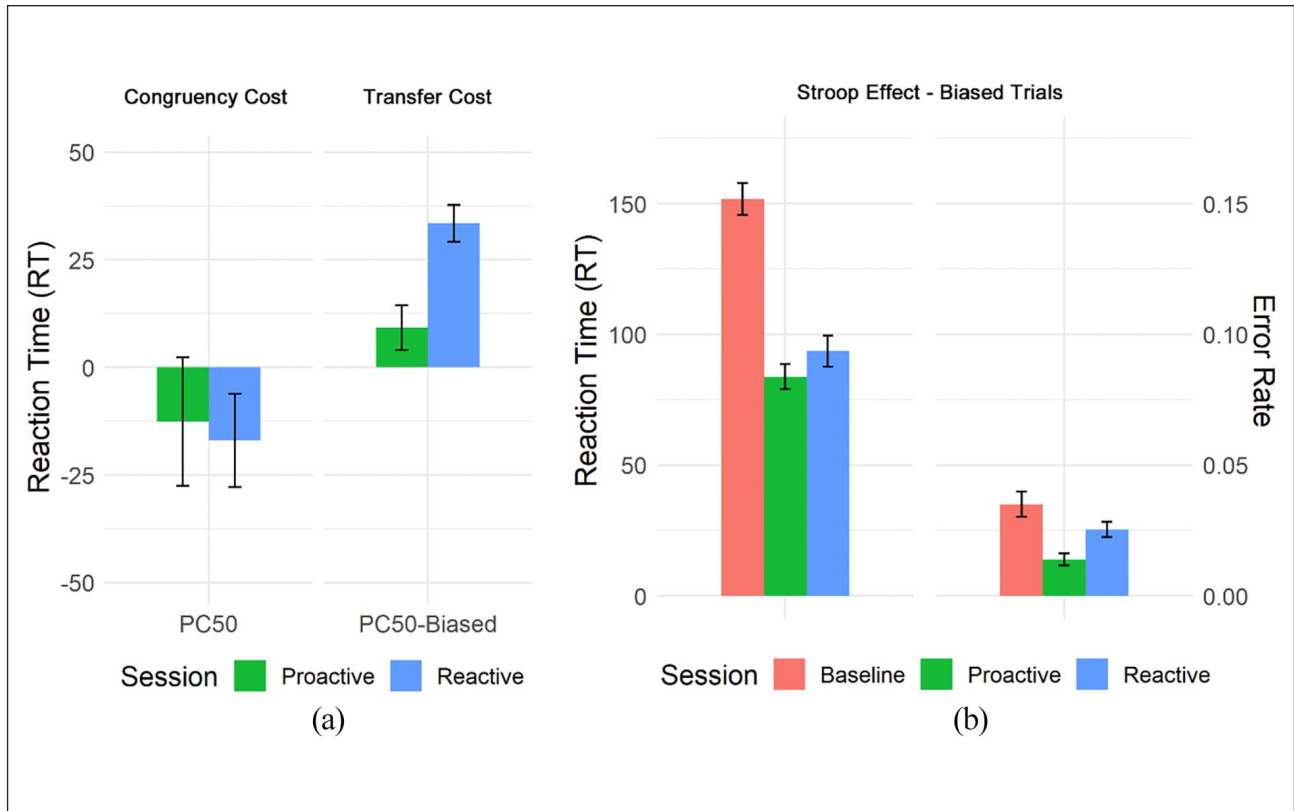
## Results

Primary and secondary predictions for each task are summarised in Figures 2 to 5 and Tables 1 to 4, respectively; the tables also indicate which predictions were confirmed. For all figures, primary results are depicted in Panel (a) and secondary results are depicted in Panel (b). Detailed descriptive data are presented in Supplementary Tables 1 to 4. Below, we present the key results and test statistics for each task measure, separately for each condition.

### Stroop

**Baseline effects.** We first verified the presence of a standard Stroop interference effect in terms of increased RT on incongruent (IC) relative to congruent (C) trials. We examined the biased (PC-75) and PC-50 items separately. In both cases, highly robust effects ( $>100$  ms) were observed—biased items, IC:  $M=920.27$ ,  $SD=381.32$ , C:  $M=768.51$ ,  $SD=372.96$ ,  $t(125)=24.49$ ,  $p < .001$ , Cohen's  $d=2.18$ ,  $BF_{10} > 100$ ; PC-50 items, IC:  $M=910.60$ ,  $SD=376.78$ , C:  $M=792.41$ ,  $SD=380.73$ ,  $t(125)=17.86$ ,  $p < .001$ , Cohen's  $d=2.18$ ,  $BF_{10} > 100$ . In the Stroop task, error rates tend to be very low overall, but also typically show Stroop interference effects as well. This pattern held in the current dataset—biased items, IC:  $M=0.07$ ,  $SD=0.08$ , C:  $M=0.03$ ,  $SD=0.06$ ,  $t(125)=7.314$ ,  $p < .001$ , Cohen's  $d=1.59$ ,  $BF_{10} > 100$ ; PC-50 items, IC:  $M=0.05$ ,  $SD=0.06$ , C:  $M=0.03$ ,  $SD=0.06$ ,  $t(125)=5.23$ ,  $p < .001$ , Cohen's  $d=0.47$ ,  $BF_{10} > 100$ .

**Proactive condition.** In the proactive condition, the list-wide PC manipulation was predicted to lead to a reduction in the Stroop RT interference effect. Critically, because of the list-wide nature of the manipulation, this reduction was predicted to affect PC-50 items as well as biased items (PC-25). This prediction was confirmed: both of these interference effects yielded strong evidence of reduction relative to the baseline condition—biased: baseline  $M=151.76$ ,  $SD=69.57$ , proactive  $M=83.74$ ,  $SD=53.43$ ,  $t(125)=12.10$ ,  $p < .001$ , Cohen's  $d=1.08$ ,  $BF_{10} > 100$ ; PC-50: baseline  $M=118.19$ ,  $SD=74.30$ , proactive  $M=92.96$ ,  $SD=68.66$ ,  $t(125)=3.76$ ,  $p < .001$ , Cohen's  $d=0.34$ ,  $BF_{10}=69.83$  (Figure 2b, left side). Similar numerical patterns were present in the error rate data, but weaker,



**Figure 2.** Stroop RT and error indices: (a) primary effects and (b) secondary effects.

**Table 1.** Stroop.

Primary results			
Measures	Formula	Predictions	Confirmation
Reaction time			
Stroop effect	Incongruent – congruent	P < B (biased, PC-50) R < B (biased) R = B (PC-50)	+* +* +
Congruency cost	Proactive/reactive congruent – baseline congruent	R < P (biased, PC-50)	+
Transfer cost	PC-50 incongruent – biased incongruent	P < R	+*
Secondary validation results			
Error			
Stroop effect	Incongruent – congruent	P < B (biased, PC-50) R < B (biased) R = B (PC-50)	+* + +

B: baseline; P: proactive; R: reactive; BF: Bayes Factor.

+Prediction confirmed.

\*Strong evidence (BF > 10 or BF < 0.1).

in that the effects were significant, but some lacked strong evidence—biased: baseline  $M=0.03$ ,  $SD=0.05$ , proactive  $M=0.01$ ,  $SD=0.0$ ,  $t(125)=4.64$ ,  $p < .001$ , Cohen's

$d=0.41$ ,  $BF_{10} > 100$ ; PC-50: baseline  $M=0.02$ ,  $SD=0.04$ , proactive  $M=0.01$ ,  $SD=0.03$ ,  $t(125)=2.07$ ,  $p = .041$ , Cohen's  $d=0.18$ ,  $BF_{10}=1.29$ ) (Figure 2b, right side).

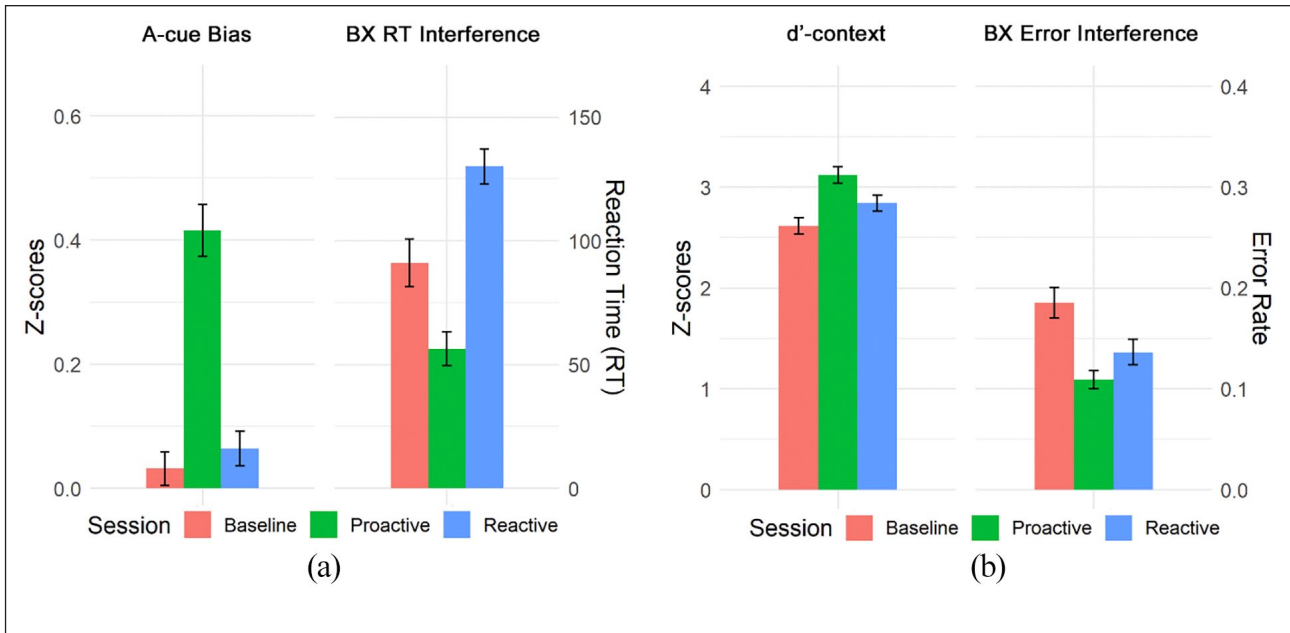


Figure 3. AX-CPT RT and error indices: (a) primary effects and (b) secondary effects.

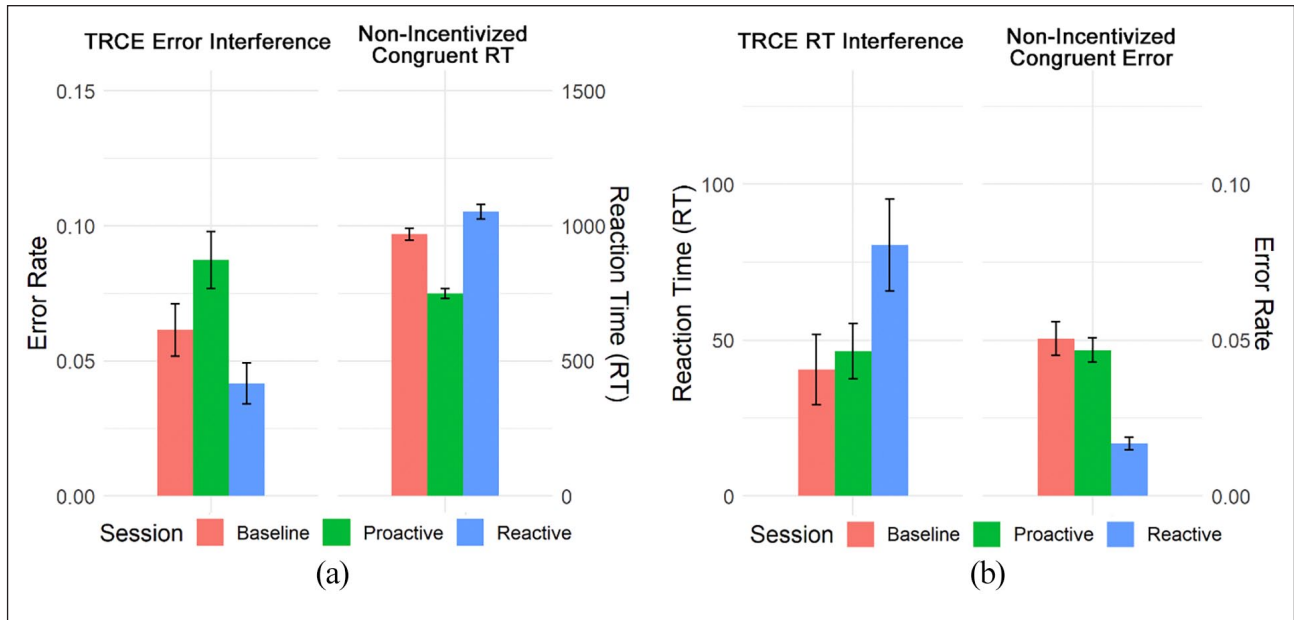
Table 2. AX-CPT.

Primary results			
Measures	Formula	Predictions	Confirmation
Reaction time			
BX interference effect	$BX - BY$	$P < B$ $R > B$ $R > P$	+* +* +*
Error			
BX interference effect	$BX - BY$	$P < B$ $R < B$	+* +*
A-cue bias	$0.5 \times (Z[\text{Hits on AX}] + Z[\text{False alarms on AY}])$	$P > 0$ $P > B$ $P > R$	+* +* +*
Secondary validation results			
Reaction time			
PBI	$(AY - BX)/(AY + BX)$	$P > 0$ $R < P$	+* +*
Error			
d'-context	$Z(\text{Hits on AX}) - Z(\text{False alarms on BX})$	$P > B$ $P > R$	+* +*
PBI	$(AY - BX)/(AY + BX)$	$P > 0$ $R < 0$	+* +*

B: baseline; P: proactive; R: reactive; BF: Bayes Factor.  
 +Prediction confirmed.  
 \*Strong evidence (BF > 10).

**Reactive condition.** In the reactive condition, the IS-PC manipulation was predicted to lead to a reduction in the Stroop RT interference effect on the biased (MI) items (PC-25). Conversely, because of the item-specific nature

of the manipulation the reduction in Stroop interference was predicted to not transfer to PC-50 items, with no change from baseline. This prediction was confirmed: biased interference effect ( $M=93.53$ ,  $SD=66.24$ ) and



**Figure 4.** Cued Task-Switching RT and error indices: (a) primary effects and (b) secondary effects.

**Table 3.** Cued Task-Switching.

Primary results			
Measures	Formula	Predictions	Confirmation
Reaction time			
Non-incentivised (biased) trials	—	P < B (biased, low-conflict biased) R > B (biased) P < R (low-conflict biased)	+* +* +*
Error			
TRCE interference (non-incentivised)	Incongruent – congruent	R < B R < P	+ +*
Secondary validation results			
Reaction time			
TRCE interference (non-incentivised)	Incongruent – congruent	R > B R > P	+* +
Error			
Non-incentivised (biased) trials	—	P = B (low-conflict biased) R < B	+ +*

B: baseline; P: proactive; R: reactive; BF: Bayes Factor.

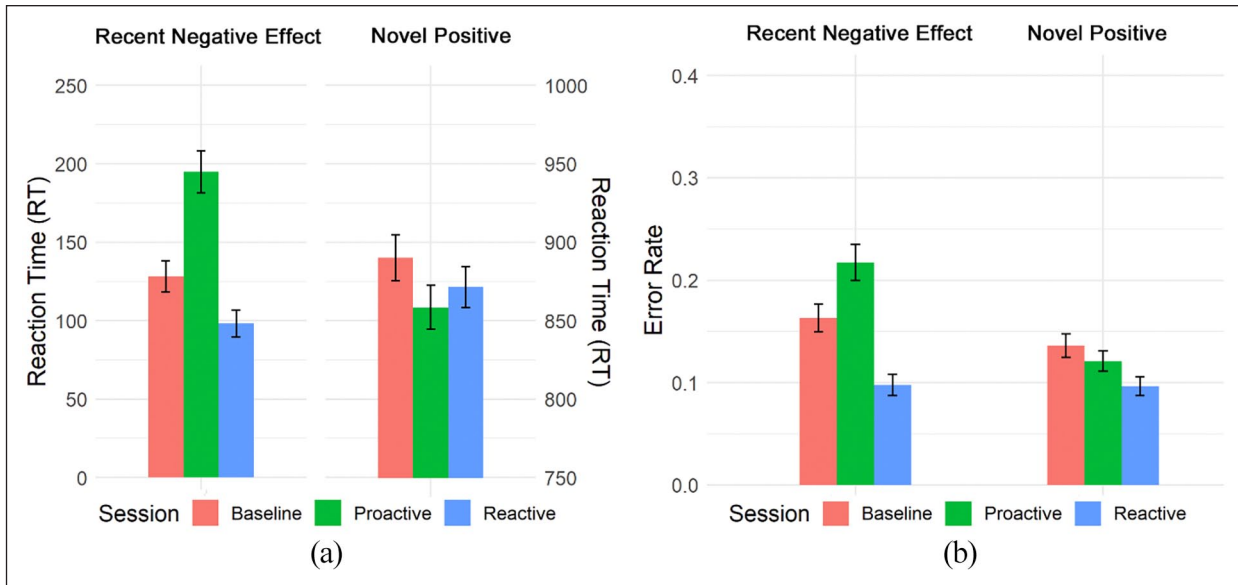
+ Prediction confirmed.

\*Strong evidence (BF > 10 or BF < 0.1).

PC-50 interference effect ( $M=127.01$ ,  $SD=73.96$ ). Only the biased items were significantly reduced relative to the baseline condition with strong evidence; in contrast, the PC-50 items showed some evidence for a null effect—biased:  $t(125)=9.00$ ,  $p<.001$ , Cohen's  $d=0.80$ ,  $BF_{10}>100$ ; PC-50:  $t(125)=-1.50$ ,  $p=.136$ , Cohen's  $d=0.13$ ,  $BF_{01}=0.29$ . Again, similar numerical patterns were present in the error rate data—biased:  $M=0.03$ ,

$SD=0.03$ ,  $t(125)=1.96$ ,  $p=.053$ , Cohen's  $d=0.21$ ,  $BF_{01}=0.63$ ; PC-50:  $M=0.02$ ,  $SD=0.03$ ,  $t(125)=-0.90$ ,  $p=.370$ , Cohen's  $d=0.09$ ,  $BF_{01}=0.15$ .

**Proactive vs. reactive.** Based on prior work, we predicted that the Stroop interference effect (in RT) would be reduced for PC-50 items in proactive, due to the differential transfer effect. This prediction was confirmed, with strong evidence:



**Figure 5.** Sternberg RT and error indices: (a) primary effects and (b) secondary effects.

**Table 4.** Sternberg.

Measures	Formula	Predictions	Confirmation
Reaction time			
Critical load performance	NP	P < B P < R	+* +
Recent negative interference effect	RN – NN	R < B R < P	+* +*
Error			
Critical load performance	NP	P < B P < R	+ –
Recent negative interference effect	RN – NN	R < B R < P	+* +*

B: baseline; P: proactive; R: reactive; BF: Bayes Factor.

+Prediction confirmed.

–Opposite of prediction.

\*Strong evidence (BF > 10).

PC-50,  $t(125)=6.43, p<.001$ , Cohen's  $d=0.57$ ,  $BF_{10} > 100$ . Conversely, we predicted no difference between conditions for biased items; there was some evidence in favour of the null effect,  $t(125)=-1.97, p=.051$ , Cohen's  $d=0.18$ ,  $BF_{01}=0.64$ . To more directly quantify these contrasting effects, we computed the “transfer cost,” which is the difference in Stroop interference across PC-50 and biased items. As predicted, the transfer cost was significantly greater in reactive ( $M=33.48, SD=47.91$ ) than proactive, with strong evidence ( $M=9.22, SD=58.63$ ),  $t(125)=3.97, p<.001$ , Cohen's  $d=0.35$ ,  $BF_{10} > 100$ ) (Figure 2a, right side).

Another prediction from prior work was that the proactive condition would reduce facilitation on congruent trials to a greater degree than reactive. The congruency cost was computed by subtracting out the baseline congruency from

both proactive and reactive on PC-50 items (and on biased items as well). Although the congruency cost effects were numerically in the predicted direction (i.e., greater in proactive), these were not close to statistically reliable—PC-50: proactive= $-12.56$ , reactive= $-17.03$ ,  $t(125)=0.36, p=.717$ , Cohen's  $d=0.03$ ,  $BF_{01}=0.11$ ; biased: proactive= $4.07$ , reactive= $-7.06$ ,  $t(125)=0.78, p=.436$ , Cohen's  $d=0.07$ ,  $BF_{01}=0.13$  (Figure 2a, left side). Thus, the congruency cost prediction was not supported in this dataset. Key predicted effects for the Stroop are summarised in Table 1.

### AX-CPT

**Baseline effects.** We verified the presence of standard AX-CPT interference effects, which include higher RTs and



error rates on high-conflict AY (RT:  $M=548.77$ ,  $SD=73.45$ ; errors:  $M=0.06$ ,  $SD=0.07$ ) and BX (RT:  $M=553.75$ ,  $SD=148.53$ ; errors:  $M=0.19$ ,  $SD=0.18$ ) non-target trials, relative to low-conflict BY non-target trials (RT:  $M=462.65$ ,  $SD=66.51$ ; errors:  $M=0.01$ ,  $SD=0.03$ ). For both trial types, highly robust effects with strong evidence were observed—AY, RT:  $t(120)=21.25$ ,  $p<.001$ , Cohen's  $d=1.93$ ,  $BF_{10}>100$ ; errors:  $t(120)=6.67$ ,  $p<.001$ , Cohen's  $d=0.61$ ,  $BF_{10}>100$ ; BX, RT:  $t(120)=9.38$ ,  $p<.001$ , Cohen's  $d=0.85$ ,  $BF_{10}>100$ ; errors:  $t(120)=10.99$ ,  $p<.001$ , Cohen's  $d=1.00$ ,  $BF_{10}>100$ . Moreover, as predicted from the inclusion of no-go trials, participants showed poorer performance on BX than AY trials in terms of error rates,  $t(120)=7.44$ ,  $p<.001$ , Cohen's  $d=0.68$ ,  $BF_{10}>100$ , and were even numerically, though not reliably slower in RT,  $t(120)=0.48$ ,  $p=.633$ , Cohen's  $d=0.04$ ,  $BF_{01}=0.11$ .

**Proactive condition.** In the proactive condition, the instructed strategy manipulation was predicted to lead to an increased utilisation of contextual cue information, as indexed by a significantly positive A-cue bias (the tendency to make a target response following an A-cue). This prediction was confirmed with strong evidence,  $M=0.42$ ,  $SD=0.46$ ,  $t(120)=10.02$ ,  $p<.001$ , Cohen's  $d=0.91$ ,  $BF_{10}>100$ . Moreover, the A-cue bias also exhibited strong evidence of increase relative to baseline ( $M=0.03$ ,  $SD=0.30$ ),  $t(120)=9.06$ ,  $p<.001$ , Cohen's  $d=0.82$ ,  $BF_{10}>100$  (Figure 3a, left side). In addition, the utilisation of context was also predicted to reduce BX interference effects in the proactive condition (errors:  $M=0.11$ ,  $SD=0.10$ ; RT:  $M=56.45$ ,  $SD=74.55$ ), in both errors and RT, relative to baseline (errors:  $M=0.19$ ,  $SD=0.17$ ; RT:  $M=91.11$ ,  $SD=106.84$ ) (Figure 3a, right side). This prediction was also confirmed with strong evidence—error:  $t(120)=5.07$ ,  $p<.001$ , Cohen's  $d=0.46$ ,  $BF_{10}>100$ ; RT:  $t(120)=3.74$ ,  $p<.001$ , Cohen's  $d=0.34$ ,  $BF_{10}=65.42$ . Although we now prefer the A-cue bias measure, because it more selectively indexes proactive control, for continuity with prior literature we further examined  $d'$ -context, which was also predicted to be improved in the proactive condition ( $M=3.12$ ,  $SD=0.88$ ), relative to baseline ( $M=2.61$ ,  $SD=0.93$ ) (Figure 3b, left side). This prediction was confirmed, as the proactive condition  $d'$ -context was significantly greater, with strong evidence,  $t(120)=6.13$ ,  $p<.001$ , Cohen's  $d=0.56$ ,  $BF_{10}>100$ .

**Reactive condition.** In the reactive condition, the probe cueing manipulation was predicted to lead to a reduction in BX error interference, but at a cost of increased BX RT interference (due to probe-triggered context retrieval). These predictions were also both confirmed, BX error interference in reactive ( $M=0.14$ ,  $SD=0.14$ ) showed strong evidence of reduction, relative to baseline ( $M=0.19$ ,  $SD=0.17$ ),  $t(120)=3.27$ ,  $p=.001$ , Cohen's  $d=0.30$ ,

$BF_{10}=15.15$  (Figure 3b, right side), whereas BX RT interference in reactive ( $M=130.07$ ,  $SD=77.40$ ) was significantly increased relative to baseline, also with strong evidence ( $M=91.11$ ,  $SD=106.84$ ),  $t(120)=3.99$ ,  $p<.001$ , Cohen's  $d=0.36$ ,  $BF_{10}>100$  (Figure 3a, right side). Although we now prefer the BX RT interference effect as a selective index of reactive control, for continuity with prior literature we further examined  $d'$ -context, which was also predicted to be improved in the reactive condition ( $M=2.84$ ,  $SD=0.85$ ), relative to baseline ( $M=2.61$ ,  $SD=0.93$ ; Figure 3b, left side). This prediction was confirmed, as the reactive condition  $d'$ -context was significantly greater than the baseline condition, but was lacking strong evidence,  $t(120)=2.76$ ,  $p=.007$ , Cohen's  $d=0.26$ ,  $BF_{10}=3.74$ .

**Proactive vs. reactive.** We predicted that the A-cue bias would be greater in proactive than reactive, whereas BX RT interference would be greater in reactive compared with proactive. Both effects were confirmed with strong evidence—A-cue bias:  $t(120)=7.97$ ,  $p<.001$ , Cohen's  $d=0.72$ ,  $BF_{10}>100$ ; BX RT interference:  $t(120)=10.13$ ,  $p<.001$ , Cohen's  $d=0.92$ ,  $BF_{10}>100$  (Figure 3a). Although we prefer these two measures as they are doubly dissociable, for completeness and comparison with prior studies we also examined the PBI and  $d'$ -context measures. In the proactive condition, the PBI had strong evidence of being positive in both error,  $M=0.18$ ,  $SD=0.53$ ,  $t(120)=3.68$ ,  $p<.001$ , Cohen's  $d=0.33$ ,  $BF_{10}=53.79$ , and RT indices,  $M=0.08$ ,  $SD=0.10$ ,  $t(120)=9.39$ ,  $p<.001$ , Cohen's  $d=0.85$ ,  $BF_{10}>100$ , whereas in reactive the PBI had strong evidence of being negative in errors,  $M=-0.19$ ,  $SD=0.52$ ,  $t(120)=4.12$ ,  $p<.001$ , Cohen's  $d=0.37$ ,  $BF_{10}>100$ . For RT, the PBI in reactive had strong evidence of being lower than proactive,  $t(120)=8.74$ ,  $p<.001$ , Cohen's  $d=0.79$ ,  $BF_{10}>100$ , consistent with predictions. Similarly, there was strong evidence for the  $d'$ -context measure being greater in proactive, relative to reactive ( $M=2.84$ ,  $SD=0.85$ ),  $t(120)=3.73$ ,  $p<.001$ , Cohen's  $d=0.34$ ,  $BF_{10}=63.31$  (Figure 3b, left side). Key predicted effects for the AX-CPT are summarised in Table 2.

### Cued-TS

**Baseline effects.** We verified the presence of standard Cued-TS effects, which include both the TRCE and the (residual) switch costs. For RT, there was strong evidence for both effects, in both the biased (mostly congruent) and unbiased items—TRCE, biased:  $M=40.51$ ,  $SD=126.90$ ,  $t(127)=3.61$ ,  $p<.001$ , Cohen's  $d=0.32$ ,  $BF_{10}=42.78$ ; unbiased:  $M=46.30$ ,  $SD=132.36$ ,  $t(127)=3.96$ ,  $p<.001$ , Cohen's  $d=0.35$ ,  $BF_{10}>100$ ; switch cost, biased:  $M=38.63$ ,  $SD=76.68$ ,  $t(127)=5.70$ ,  $p<.001$ , Cohen's  $d=0.50$ ,  $BF_{10}>100$ ; unbiased:  $M=39.96$ ,  $SD=115.66$ ,  $t(127)=3.91$ ,  $p<.001$ , Cohen's  $d=0.35$ ,  $BF_{10}>100$ .

These effects also tend to be present in error rate, and there was strong evidence confirming this pattern in the data as well, for both the TRCE, biased:  $M=0.06$ ,  $SD=0.11$ ,  $t(127)=6.33$ ,  $p<.001$ , Cohen's  $d=0.56$ ,  $BF_{10}>100$ ; unbiased:  $M=0.06$ ,  $SD=0.10$ ,  $t(127)=6.94$ ,  $p<.0001$ , Cohen's  $d=0.61$ ,  $BF_{10}>100$ , and switch cost, biased:  $M=0.02$ ,  $SD=0.04$ ,  $t(127)=4.91$ ,  $p<.001$ , Cohen's  $d=0.43$ ,  $BF_{10}>100$ ; unbiased:  $M=0.03$ ,  $SD=0.08$ ,  $t(127)=3.87$ ,  $p<.001$ , Cohen's  $d=0.34$ ,  $BF_{10}=97.60$ .

**Proactive condition.** We first examined the effects of the reward incentive manipulation. This manipulation was predicted to speed RTs, but also increase errors, which was confirmed with strong evidence, when comparing incentivised trials (RT:  $M=709.11$ ,  $SD=211.00$ ; errors:  $M=0.16$ ,  $SD=0.10$ ) to non-incentivised trials, RT:  $M=770.22$ ,  $SD=208.36$ ,  $t(127)=7.25$ ,  $p<.001$ , Cohen's  $d=0.64$ ,  $BF_{10}>100$ ; errors:  $M=0.09$ ,  $SD=0.07$ ,  $t(127)=8.89$ ,  $p<.001$ , Cohen's  $d=0.79$ ,  $BF_{10}>100$ . This shift in control strategy was predicted to affect even the non-incentive (biased) trials, which could be compared directly with baseline. Confirming this prediction, there was strong evidence for faster RTs on these non-incentivised trials in the proactive condition ( $M=770.22$ ,  $SD=208.36$ ), relative to baseline ( $M=988.41$ ,  $SD=254.83$ ),  $t(127)=16.66$ ,  $p<.001$ , Cohen's  $d=1.47$ ,  $BF_{10}>100$ . Moreover, even when restricting the focus to just low-conflict C trials, there was still strong evidence for this effect, proactive:  $M=746.98$ ,  $SD=197.89$ ; baseline:  $M=968.16$ ,  $SD=257.56$ ,  $t(127)=18.71$ ,  $p<.001$ , Cohen's  $d=1.65$ ,  $BF_{10}>100$  (Figure 4a, right side). In addition, the RT speeding on these trials occurred in the absence of a change in error rate, relative to baseline, with some evidence for the null, proactive:  $M=0.05$ ,  $SD=0.04$ ; baseline:  $M=0.05$ ,  $SD=0.06$ ,  $t(127)=0.79$ ,  $p=.431$ , Cohen's  $d=0.07$ ,  $BF_{01}=0.13$  (Figure 4b, right side), suggesting more than just a speed-accuracy shift.

**Reactive condition.** We first examined the effects of the punishment incentive manipulation. This manipulation was predicted to slow RTs, but also decrease errors. The RT prediction was confirmed with strong evidence when comparing incentivised trials (RT:  $M=1,202.03$ ,  $SD=338.29$ ) to non-incentivised trials (RT:  $M=1,094.76$ ,  $SD=303.19$ ),  $t(127)=9.55$ ,  $p<.001$ , Cohen's  $d=0.84$ ,  $BF_{10}>100$ . However, the effect of reduced errors was not detected, incentivised:  $M=0.04$ ,  $SD=0.05$ ; non-incentivised:  $M=0.04$ ,  $SD=0.05$ ,  $t(127)=0.83$ ,  $p=.407$ , Cohen's  $d=0.07$ ,  $BF_{01}=0.14$ . This shift in control strategy was predicted to affect even non-incentivised (biased) trials, which could be directly compared with the baseline condition. Confirming this prediction, there was strong evidence for RTs being slower in reactive ( $M=1,094.76$ ,  $SD=303.19$ ) relative to baseline,  $M=988.41$ ,  $SD=254.83$ ,  $t(127)=5.48$ ,

$p<.001$ , Cohen's  $d=0.48$ ,  $BF_{10}>100$ , and in this comparison there was also strong evidence for errors being lower as well, reactive:  $M=0.04$ ,  $SD=0.05$ , baseline:  $M=0.08$ ,  $SD=0.09$ ,  $t(127)=-6.50$ ,  $p<.001$ , Cohen's  $d=0.55$ ,  $BF_{10}>100$ . A stronger prediction was that this effect might be related to TRCE interference, which was also predicted to be reduced for error interference in reactive, but not TRCE RT interference. This prediction was partially confirmed, in that TRCE error interference in non-incentivised trials was numerically lower and close to statistically significant ( $M=0.04$ ,  $SD=0.09$ ), relative to baseline— $M=0.06$ ,  $SD=0.11$ ,  $t(127)=1.97$ ,  $p=.051$ , Cohen's  $d=0.17$ ,  $BF_{10}=1.57$  (Figure 4a, left side), but the effect lacked strong evidence. Conversely, TRCE RT interference in reactive ( $M=80.45$ ,  $SD=166.06$ ) was statistically greater when compared with baseline— $M=40.51$ ,  $SD=126.90$ ,  $t(127)=2.62$ ,  $p=.010$ , Cohen's  $d=0.23$ ,  $BF_{10}=2.59$  (Figure 4b, left side), but again this effect lacked strong evidence.

**Proactive vs. reactive.** We predicted that on non-incentivised trials, the TRCE error effect would be reduced in reactive relative to proactive. This prediction was confirmed, in that there was strong evidence that the TRCE error effect was lower in reactive ( $M=0.04$ ,  $SD=0.09$ ) compared with proactive— $M=0.09$ ,  $SD=0.12$ ,  $t(127)=4.56$ ,  $p<.001$ , Cohen's  $d=0.40$ ,  $BF_{10}>100$  (Figure 4a, left side). In addition, we predicted that in the proactive condition, there would be general response speeding relative to reactive, even on non-incentivised and low-conflict congruent trials. This effect was confirmed and also with strong evidence—proactive congruent RT:  $M=746.98$ ,  $SD=197.89$ , reactive congruent RT:  $M=1,054.53$ ,  $SD=308.78$ ,  $t(127)=-16.93$ ,  $p<.001$ , Cohen's  $d=1.50$ ,  $BF_{10}>100$  (Figure 4a, right side). Key predicted effects for the Cued-TS are summarised in Table 3.

## Sternberg

**Baseline effects.** We first verified the presence of standard WM load effects, by comparing the critical to higher load items. We expected higher error rates and longer RTs on the high-load items; there was strong evidence for these effects—errors: critical=0.13 (0.09), high load=0.21 (0.09),  $t(127)=11.31$ ,  $p<.001$ , Cohen's  $d=1.00$ ,  $BF_{10}>100$ ; RT: critical=897.02 (160.91), high load=943.34 (169.91),  $t(125)=6.36$ ,  $p<.001$ , Cohen's  $d=0.57$ ,  $BF_{10}>100$ . We also tested for the RN effect (RN-NN), which had strong evidence for both errors,  $M=0.16$ ,  $SD=0.15$ ,  $t(127)=12.11$ ,  $p<.001$ , Cohen's  $d=1.07$ ,  $BF_{10}>100$ , and RT,  $M=128.05$ ,  $SD=110.77$ ,  $t(127)=13.08$ ,  $p<.001$ , Cohen's  $d=1.16$ ,  $BF_{10}>100$ . Similarly, strong evidence was similarly obtained when examining non-critical (high-load) items for both errors,

$M=0.32$ ,  $SD=0.18$ ,  $t(127)=20.39$ ,  $p<.001$ , Cohen's  $d=1.80$ ,  $BF_{10}>100$ , and RT,  $M=174.25$ ,  $SD=190.14$ ,  $t(127)=10.29$ ,  $p<.001$ , Cohen's  $d=0.92$ ,  $BF_{10}>100$ .

**Proactive condition.** We first tested that the load manipulation was successful, comparing critical to low-load items, predicting lower error rates and faster RTs with lower load. These effects were confirmed with strong evidence—errors: critical=0.15 (0.11), low load=0.05 (0.08),  $t(127)=11.35$ ,  $p<.001$ , Cohen's  $d=1.00$ ,  $BF_{10}>100$ ; RT: critical=912.65 (173.84), low load=828.87 (163.50),  $t(126)=11.28$ ,  $p<.001$ , Cohen's  $d=1.00$ ,  $BF_{10}>100$ . The key prediction involved the critical items, which could be directly compared with baseline, which was predicted to show better performance on NP trials. This prediction was only partially confirmed, in that the effects were statistically significant for RT but lacking strong evidence, proactive:  $M=858.48$ ,  $SD=157.29$ , baseline:  $M=890.12$ ,  $SD=166.92$ ,  $t(127)=2.60$ ,  $p=.01$ , Cohen's  $d=0.23$ ,  $BF_{10}=2.47$  (Figure 5a, right side); the effects were in the correct numerical direction, but not significant for errors, proactive:  $M=0.12$ ,  $SD=0.11$ , baseline:  $M=0.14$ ,  $SD=0.13$ ,  $t(127)=1.42$ ,  $p=.159$ , Cohen's  $d=0.13$ ,  $BF_{10}=3.82$  (Figure 5b, right side).

**Reactive condition.** We first tested that the load manipulation was successful, as in the baseline condition, comparing critical to high-load items, predicting higher error rates and slower RTs with higher load. These effects were confirmed with strong evidence, errors: critical=0.09 (0.08), high load=0.18 (0.09),  $t(127)=14.52$ ,  $p<.001$ , Cohen's  $d=1.28$ ,  $BF_{10}>100$ ; RT: critical=890.79 (143.43), high load=948.47 (149.41),  $t(127)=11.62$ ,  $p<.001$ , Cohen's  $d=1.03$ ,  $BF_{10}>100$ . The key prediction was on critical items, which could be directly compared with baseline; we predicted a reduced RN effect in reactive. This prediction was confirmed. For errors, the RN interference effect was reliably reduced in the reactive condition, with strong evidence ( $M=0.10$ ,  $SD=0.12$ ) relative to baseline,  $M=0.16$ ,  $SD=0.15$ ,  $t(127)=4.37$ ,  $p<.001$ , Cohen's  $d=0.39$ ,  $BF_{10}>100$  (Figure 5b, left side). For RT, the effect was statistically significant, but lacked strong evidence, reactive:  $M=98.10$ ,  $SD=96.40$ ; baseline:  $M=128.05$ ,  $SD=110.77$ ,  $t(127)=2.48$ ,  $p=.014$ , Cohen's  $d=0.22$ ,  $BF_{10}=1.86$  (Figure 5a, left side).

**Proactive vs. reactive.** We predicted that for critical items, performance would be better in proactive on NP trials, but that the RN interference effect would be reduced in the reactive condition. This prediction was only partially confirmed. Although RTs were numerically faster on NP trials in proactive, relative to reactive, this effect was not statistically significant, proactive:  $M=858.48$ ,  $SD=157.29$ , reactive:  $M=871.40$ ,  $SD=149.05$ ,  $t(127)=1.26$ ,  $p=.210$ , Cohen's  $d=0.11$ ,  $BF_{10}=4.70$  (Figure 5a, right side). Furthermore,

contrary to our prediction, NP error rates were actually significantly higher in proactive relative to reactive, though this lacked strong evidence; proactive:  $M=0.12$ ,  $SD=0.11$ , reactive:  $M=0.10$ ,  $SD=0.10$ ,  $t(127)=-2.40$ ,  $p=.018$ , Cohen's  $d=0.21$ ,  $BF_{10}=1.55$  (Figure 5b, right side). Conversely, the RN effect provided strong evidence in support of the prediction, both in terms of errors—proactive:  $M=0.22$ ,  $SD=0.20$ , reactive:  $M=0.10$ ,  $SD=0.12$ ,  $t(127)=6.72$ ,  $p<.001$ , Cohen's  $d=0.59$ ,  $BF_{10}>100$ —and RT, proactive:  $M=194.77$ ,  $SD=148.99$ , reactive:  $M=98.17$ ,  $SD=96.78$ ,  $t(126)=7.24$ ,  $p<.001$ , Cohen's  $d=0.64$ ,  $BF_{10}>100$  (Figure 5a and b, left side). Key predicted effects for the Sternberg are summarised in Table 4.

## Discussion

The primary goal of this report was to comprehensively describe the newly developed DMCC task battery and to rigorously evaluate the degree to which experimental manipulations produce group-level shifts in proactive control and reactive control, as predicted by the Dual Mechanisms framework. In each of the four tasks, we compared task performance and primary outcome indices among the three conditions (baseline, proactive, reactive) to evaluate both convergent (cross-task) and divergent (discriminant) validity of the DMCC task battery in capturing variations in the two cognitive control modes.

In the Stroop task, the list-wide and IS-PC manipulations were generally successful in producing the predicted shifts towards proactive control and reactive control. In particular, three of the five key predictions were confirmed with strong evidence, except for the Stroop effect in reactive relative to baseline for the PC-50 items (reactive=baseline) and congruency cost (proactive < reactive), although both of these effects were in the correct numerical direction. For AX-CPT, findings indicated that the context strategy manipulation and the probe cueing manipulation were successful in dissociating the two modes of control. All eight of the key predictions were confirmed with strong evidence. Similarly, many additional measures of historical interest (e.g.,  $d'$ -context, PBI) also exhibited consistent patterns. For the Cued-TS, the reward and punishment incentive manipulations successfully produced differential effects on RT and error rates, supporting the dissociable nature of proactive control and reactive control. Specifically, four of the five key predictions were confirmed with strong evidence; for the TRCE interference effect on error rate, the difference between reactive and baseline conditions was in the predicted direction but was not statistically significant. Finally, for the Sternberg WM task, the manipulations of WM load and RN trials did affect task performance in differential ways. In particular, five of the eight key predictions were confirmed with strong evidence. However, the performance of NP trials was the one condition across the four DMCC

tasks, in which our predictions were clearly disconfirmed. Specifically, the RT on NP trials in the proactive condition was not significantly faster than the reactive condition, and the NP trial error rate was in fact numerically higher in proactive compared with the reactive condition.

The key results from the Sternberg task, namely, less reliable effects and, in some cases, contrary patterns, may suggest a potential need for further task development and optimisation. Importantly, to our knowledge this is the first time that the proactive and reactive Sternberg task variants have been directly compared. Nevertheless, both the RN interference and the WM load effects were reliably demonstrated in each condition, as longer RT and higher error rate were detected in high WM load trials relative to low-load trials, and in RN trials relative to NN trials. These results suggested that both experimental manipulations were valid in terms of producing the basic effects.

Thus, when considered together, our evaluation of the DMCC task battery suggests that it exhibits substantial convergent and divergent validity. In terms of convergent validity, as just described, the experimental manipulations were generally effective in producing common experimental patterns in all four tasks, suggesting robust cross-task sensitivity to cognitive control demands (summarised in Tables 1 to 4). In terms of divergent validity, there were clear patterns of double dissociation, in that the behavioural markers of proactive and reactive control could effectively be distinguished in all four tasks, with one set of measures showing the predicted proactive > reactive pattern, at least numerically (Stroop congruency cost, AX-CPT A-cue bias, Cued-TS TRCE error interference, Sternberg RN RT effect; see Figures 2 to 5 Panel (a), *left side*), and another set of measures showing the reverse predicted reactive > proactive pattern, again at least numerically (Stroop transfer cost, AX-CPT BX RT interference, Cued-TS non-incentivised C RT, Sternberg NP RT; see Figures 2 to 5 Panel (a), *right side*).

Notably, the online format of data collection proved to be a strength but may have also resulted in some limitations for this study. From a practical standpoint, the nature of this multi-session and multi-task study made frequent laboratory visits less optimal and more time-consuming for data collection of a large sample size. As such, the utilisation of an online format helped to lower the barrier for participation, facilitating the large-scale data collection effort. In particular, both researcher and participant burden were much reduced, since administration demands were largely automated and therefore less time-consuming. For participants, completing each session at their own convenience and from the comfort of their own home made study completion a much more attractive proposition. Nevertheless, by allowing participants to take the tasks in a non-laboratory setting that precluded monitoring by the researchers, it is quite possible that potential distractions could have occurred during participant completion of study sessions. This is a well-known problem with online studies (Skitka & Sargis,

2005), and some results have suggested possible impacts on task performance (Bauer et al., 2012; Skitka & Sargis, 2005). However, in prior online studies in related domains, many key effects have been well replicated and indicate comparable patterns to those observed in laboratory settings (e.g., Crump et al., 2013; Germine et al., 2012; Hicks et al., 2016). Similarly, in the current study, we were able to reproduce some of the same effects, for example, Stroop (Gonthier, Braver, et al., 2016), previously observed in laboratory settings, which provides some reassurance regarding the feasibility and validity of administering the task battery in an online format.

A key advantage of the online format of the DMCC task battery is that it can enable rapid future large-scale replication of the present findings, as well as additional investigation of cognitive control modes in different labs and in various populations. Indeed, in our opinion, one of the most potentially promising applications of the task battery is to test for potential changes in proactive and reactive control across different age groups (e.g., adolescents, older adults), or in various clinical populations (e.g., depression, attention deficit hyperactivity disorder, anxiety disorder). In fact, a novel aspect of this task battery is that it can provide valid and reliable proactive and reactive control assessments and measures across four different task domains, enabling a test for consistency of cognitive control profiles in different groups. We believe this is an important use of the task battery that will be of broad interest to the wider research community.

Although the goal of this report was to describe and evaluate the validity of the DMCC task battery in terms of group effects, another important area of focus relates to the utility of the battery for individual differences analyses. A potential fruitful research direction is to examine individual differences in proactive control and reactive control modes, exploring putative state and trait factors that may influence cognitive control biases and task performance. In a companion paper, we provide extensive analyses of the psychometric characteristics of the DMCC task battery, in particular, focusing on test-retest reliability (Snijder et al., 2022). Indeed, in that paper, we were able to detect good to excellent reliability of all tasks and conditions within the battery, using newer hierarchical Bayesian modelling, which recent work has suggested provides a more statistically appropriate means of estimating reliability and individual differences in cognitive control tasks (Haines et al., 2020; Rouder & Haaf, 2019). Furthermore, in the current study, we also collected an additional set of individual difference measures—including both tasks of cognitive ability and self-report questionnaires indexing personality traits and psychological well-being—to enable analyses testing for linkages between these measures and the DMCC task battery. For instance, we provide a first case study illustration of the utility of such analyses in another companion paper, demonstrating the selective and specific relationship between the proactive condition of

the AX-CPT and individual differences in WM capacity (Lin et al., 2022). We plan to make the full DMCC dataset publicly available to enable further analyses of this type by interested researchers.

Another important use of the task battery is to elucidate the neural mechanisms associated with proactive and reactive control. Indeed, in a large neuroimaging study, we have been utilising an almost identical version of the DMCC battery, with all the same tasks and conditions included and just slightly adapted for the neuroimaging task environment (Braver et al., 2021). Although this study is still ongoing, we have validated that the battery can be used to identify consistent neural activity patterns, as well as proactive and reactive effects, across the four DMCC task domains within the fronto-parietal and cingulo-opercular brain networks that are thought to be the key substrates of cognitive control. Similarly, we have found that the DMCC battery provides a useful task context from which to explore: genetic similarity in cognitive control neural activity patterns among monozygotic twins (Tang et al., 2021); dissociability of neural control representations (Freund et al., 2021); and whole-brain neural modelling of cognitive control (Singh et al., 2022). We have already begun to publicly release portions of the DMCC neuroimaging dataset (Etzel et al., 2022), so that others can explore brain-behaviour relationships, potentially making use of the proactive and reactive control metrics that we have validated here.

It is worth noting that our interest in optimising the study design for future investigations of individual difference led to the use of a fixed condition order across participants. For detection of individual differences, it is beneficial to have every participant perform the tasks in the same order, so that order effects do not serve as a between-individual confound variable. However, for studies of group and condition differences, as in the current article, counterbalancing of condition order is often a key feature, to enable examination and control of systematic order effects. For example, it is possible that some of the predicted patterns in proactive versus reactive comparisons that we did not observe in the current study (e.g., Stroop congruency cost, Sternberg NP effects) may have been affected by the order in which conditions were performed (i.e., all participants performed the proactive condition after reactive). Conversely, it is also possible that some of the observed effects with strong evidence may have been weaker, if aggregating data from other condition orders. Thus, a useful extension of the present study would be to run additional waves of data collection, with participants performing the battery with a different condition order to both replicate the current findings and also to test the impact of a different condition order (e.g., proactive before reactive) on the pattern of data. In this case, the online feature of the battery lowers the barrier for future waves of data collection in which to examine the effect of these types of design changes.

A central tenet of the DMCC framework is the domain-generality of proactive and reactive control modes. The current findings provide support for domain-generality in that consistent shifts in control mode could be induced in each of the four tasks. Nevertheless, stronger evidence for domain-generality will require in-depth analyses of relationships among the tasks and indices of each control mode, potentially through multi-level, Bayesian, or latent-variable modelling. Future investigations concerning the DMCC task battery will need to more systematically evaluate this domain-general hypothesis regarding cognitive control modes. As described above, we have made the DMCC task battery data publicly available (<https://osf.io/pqyga/>), to encourage interested investigators to conduct further explorations. Similarly, to facilitate future development of the battery, all task scripts and additional information about the task battery are available at the DMCC project website (<https://sites.wustl.edu/dualmechanisms/>). It is our hope that the richness of the DMCC battery and associated dataset will open new avenues of research and assist other investigators in addressing key questions regarding the mechanisms of cognitive control.

### Acknowledgements

We would like to thank Erin Gourley for her contributions to task programming and data collection for the DMCC project.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Institutes of Health grant (R37 MH066078) and through funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis.

### ORCID iDs

Rongxiang Tang  <https://orcid.org/0000-0003-0406-9024>

Julie M Bugg  <https://orcid.org/0000-0002-0969-186X>

Jean-Paul Snijder  <https://orcid.org/0000-0001-9341-1447>

Todd S Braver  <https://orcid.org/0000-0002-2631-3393>

### Supplementary material

The Supplementary material is available at: [qjep.sagepub.com](http://qjep.sagepub.com).

### Note

1. The astute reader might notice that the use of green font cue to signal punishment incentive trials might be misaligned with the typical cultural/symbolic association of punishment with red font. Indeed, it was our original intention to

program the task this way. The font assignment was in fact a programming error that was only detected after data collection had been completed. Nonetheless, we expect that any potential effects of font colour were quite minor, as we observed the expected reactive control effects in Cued Task-Switching.

## References

- Barch, D. M., & Ceaser, A. (2012). Cognition in schizophrenia: Core psychological and neural mechanisms. *Trends in Cognitive Sciences*, *16*(1), 27–34. <https://doi.org/10.1016/j.tics.2011.11.015>
- Barch, D. M., Yodkovik, N., Sypher-Locke, H., & Hanewinkel, M. (2008). Intrinsic motivation in schizophrenia: Relationships to cognitive function, depression, anxiety, and personality. *Journal of Abnormal Psychology*, *117*(4), 776–787. <https://doi.org/10.1037/a0013944>
- Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, *27*(3), 362–373. <https://doi.org/10.1093/arclin/acs027>
- Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, *23*(9), 769–783. <https://doi.org/10.1016/j.tics.2019.07.002>
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In C. Jarrold (Ed.), *Variation in working memory* (pp. 76–106). Oxford University Press.
- Braver, T. S., Kizhner, A., Tang, R., Freund, M. C., & Etzel, J. A. (2021). The dual mechanisms of cognitive control project. *Journal of Cognitive Neuroscience*, *33*(9), 1990–2015. [https://doi.org/10.1162/jocn\\_a\\_01768](https://doi.org/10.1162/jocn_a_01768)
- Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351–7356. <https://doi.org/10.1073/pnas.0808187106>
- Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*(4), 713–726. [https://doi.org/10.1016/s0896-6273\(03\)00466-5](https://doi.org/10.1016/s0896-6273(03)00466-5)
- Braver, T. S., Satpute, A. B., Rush, B. K., Racine, C. A., & Barch, D. M. (2005). Context processing and context maintenance in healthy aging and early stage dementia of the Alzheimer's type. *Psychology and Aging*, *20*(1), 33–46. <https://doi.org/10.1037/0882-7974.20.1.33>
- Bugg, J., & Crump, M. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*, Article 367. <https://doi.org/10.3389/fpsyg.2012.00367>
- Bugg, J. M. (2014a). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587. <https://doi.org/10.1037/a0035032>
- Bugg, J. M. (2014b). Evidence for the sparing of reactive cognitive control with age. *Psychology and Aging*, *29*(1), 115–127. <https://doi.org/10.1037/a0035270>
- Bugg, J. M., & Braver, T. S. (2016). Proactive control of irrelevant task rules during cued task switching. *Psychological Research*, *80*(5), 860–876. <https://doi.org/10.1007/s00426-015-0686-5>
- Bugg, J. M., & Chanani, S. (2011). List-wide control is not entirely elusive: Evidence from picture–word Stroop. *Psychonomic Bulletin & Review*, *18*(5), 930–936. <https://doi.org/10.3758/s13423-011-0112-y>
- Bugg, J. M., & Dey, A. (2018). When stimulus-driven control settings compete: On the dominance of categories as cues for control. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(12), 1905–1932. <https://doi.org/10.1037/xhp0000580>
- Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 433–449. <https://doi.org/10.1037/a0029145>
- Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011). Why it is too early to lose control in accounts of item-specific proportion congruency effects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 844–859. <https://doi.org/10.1037/a0019957>
- Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: Effects of interference expectancy and fluid intelligence. *PLOS ONE*, *5*(9), Article e12861. <https://doi.org/10.1371/journal.pone.0012861>
- Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, *106*(14), 5529–5533. <https://doi.org/10.1073/pnas.0810002106>
- Chun, C. A., Ciceron, L., & Kwapil, T. R. (2018). A meta-analysis of context integration deficits across the schizotypy spectrum using AX-CPT and DPX tasks. *Journal of Abnormal Psychology*, *127*(8), 789–806. <https://doi.org/10.1037/abn0000383>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*, *8*(3), Article e57410. <https://doi.org/10.1371/journal.pone.0057410>
- De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing: An International Journal*, *69*(10–12), 1322–1326. <https://doi.org/10.1016/j.neucom.2005.12.100>
- Dey, A., & Bugg, J. M. (2021). The timescale of control: A meta-control property that generalizes across tasks but varies between types of control. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 472–489. <https://doi.org/10.3758/s13415-020-00853-x>
- Etzel, J. A., Brough, R. E., Freund, M. C., Kizhner, A., Lin, Y., Singh, M. F., Tang, R., Tay, A., Wang, A., & Braver, T. S. (2022). The Dual Mechanisms of Cognitive Control dataset, a theoretically-guided within-subject task fMRI battery. *Scientific Data*, *9*(1), Article 114. <https://doi.org/10.1038/s41597-022-01226-4>

- Freund, M. C., Bugg, J. M., & Braver, T. S. (2021). A representational similarity analysis of cognitive control during color-word Stroop. *Journal of Neuroscience*, *41*(35), 7388–7402. <https://doi.org/10.1523/JNEUROSCI.2956-20.2021>
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. <https://doi.org/10.3758/s13423-012-0296-9>
- Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*(5), 778–788. <https://doi.org/10.3758/s13421-016-0591-1>
- Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R., & Braver, T. S. (2016). Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology*, *7*, Article 1822. <https://doi.org/10.3389/fpsyg.2016.01822>
- Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., Ahn, W., & Turner, B. (2020). *Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox*. PsyArXiv. <https://doi.org/10.31234/osf.io/xr7y3>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d. *Behavior Research Methods, Instruments & Computers*, *27*(1), 46–51. <https://doi.org/10.3758/BF03203619>
- Hicks, K. L., Foster, J. L., & Engle, R. W. (2016). Measuring working memory capacity on the web with the online working memory lab (the OWL). *Journal of Applied Research in Memory and Cognition*, *5*(4), 478–489. <https://doi.org/10.1016/j.jarmac.2016.07.010>
- Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 851–860. <https://doi.org/10.1037/a0023437>
- Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638–644. <https://doi.org/10.3758/bf03196526>
- Janowich, J. R., & Cavanagh, J. F. (2018). Delay knowledge and trial set count modulate use of proactive versus reactive control: A meta-analytic review. *Psychonomic Bulletin & Review*, *25*(4), 1249–1268. <https://doi.org/10.3758/s13423-018-1502-1>
- Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193. <https://doi.org/10.1016/j.neuroscience.2005.06.042>
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. <https://doi.org/10.3758/bf03196323>
- Lin, Y., Brough, R. E., Tay, A., Jackson, J. J., & Braver, T. S. (2022). Working memory capacity preferentially enhances implementation of proactive control. <https://doi.org/https://doi.org/10.31234/osf.io/wvvpbn>
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 219–234. <https://doi.org/10.1037//0096-1523.20.2.219>
- Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*(3), 166–174. <https://doi.org/10.3758/BF03197535>
- Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, *94*(2), 145–188. [https://doi.org/10.1016/0001-6918\(95\)00050-X](https://doi.org/10.1016/0001-6918(95)00050-X)
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442. <https://doi.org/10.1037/0278-7393.22.6.1423>
- Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 137–157. <https://doi.org/10.1037/0096-1523.34.1.137>
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Minear, M., & Shah, P. (2008). Training and transfer effects in task switching. *Memory & Cognition*, *36*(8), 1470–1483. <https://doi.org/10.3758/MC.336.8.1470>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral Cortex*, *18*(5), 1010–1028. <https://doi.org/10.1093/cercor/bhm135>
- Richmond, L. L., Redick, T. S., & Braver, T. S. (2015). Remembering to prepare: The benefits (and costs) of high working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1764–1777. <https://doi.org/10.1037/xlm0000122>
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>
- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. <https://doi.org/10.3758/s13423-018-1558-y>
- Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. *Archives of General Psychiatry*, *53*(12), 1105–1112. <https://doi.org/10.1001/archpsyc.1996.01830120037008>
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. <https://doi.org/10.1037/0033-295X.84.2.127>
- Singh, M. F., Wang, A., Cole, M., Ching, S., & Braver, T. S. (2022). Enhancing task fMRI preprocessing via individualized model-based filtering of intrinsic activity dynamics. *NeuroImage*, *247*, Article 118836. <https://doi.org/10.1016/j.neuroimage.2021.118836>

- Skitka, L. J., & Sargis, E. G. (2005). The Internet as psychological laboratory. *Annual Review of Psychology*, *57*(1), 529–555. <https://doi.org/10.1146/annurev.psych.57.102904.190048>
- Snijder, J., Tang, R., Bugg, J., Conway, A. R. A., & Braver, T. S. (2022). *On the psychometric evaluation of cognitive control tasks: An investigation with the Dual Mechanisms of Cognitive Control (DMCC) Battery*. PsyArXiv. <https://doi.org/10.31234/osf.io/z7ury>
- Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(3), 155–167. <https://doi.org/10.3758/cabn.3.3.155>
- Spinelli, G., Perry, J. R., & Lupker, S. J. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture–word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(8), 995–1014. <https://doi.org/10.1037/xhp0000656>
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654. <https://doi.org/10.1126/science.153.3736.652>
- Tang, R., Etzel, J. A., Kizhner, A., & Braver, T. S. (2021). Frontoparietal pattern similarity analyses of cognitive control in monozygotic twins. *NeuroImage*, *241*, Article 118415. <https://doi.org/10.1016/j.neuroimage.2021.118415>