



How is multi-tasking different from increased difficulty?

Zachary L. Howard¹ · Nathan J. Evans² · Reilly J. Innes¹ · Scott D. Brown¹ · Ami Eidels¹

Published online: 21 May 2020
© The Psychonomic Society, Inc. 2020

Abstract

With the advancement of technologies like in-car navigation and smartphones, concerns around how cognitive functioning is influenced by “workload” are increasingly prevalent. Research shows that spreading effort across multiple tasks can impair cognitive abilities through an overuse of resources, and that similar overload effects arise in difficult single-task paradigms. We developed a novel lab-based extension of the Detection Response Task, which measures workload, and paired it with a Multiple Object Tracking Task to manipulate cognitive load. Load was manipulated either by changing within-task difficulty or by the addition of an extra task. Using quantitative cognitive modelling we showed that these manipulations cause similar cognitive impairments through diminished processing rates, but that the introduction of a second task tends to invoke more cautious response strategies that do not occur when only difficulty changes. We conclude that more prudence should be exercised when directly comparing multi-tasking and difficulty-based workload impairments, particularly when relying on measures of central tendency.

Keywords Cognitive workload · Bayesian modeling · Computational models · Multitasking

Introduction

Multitasking is the way of modern life. At work, we have several screens and can communicate with collaborators all over the globe. On the go, we can stay connected through in-car technologies like hands-free mobile phones, touch screens, heads-up-displays, and voice-controlled emails. However, Strayer and colleagues (Cooper, Ingebretsen, & Strayer, 2014; Strayer et al., 2013) have shown that these technologies are associated with a significant increase in mental effort, or cognitive workload. Effortful secondary-task use is associated with diminished performance in simulated driving tasks (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Strayer, Drews, & Johnston, 2003) with impairments roughly equivalent to those experienced by drunk drivers (Strayer, Drews, & Crouch, 2006). A recent investigation pointed to driver distraction from the use of cognitively demanding

secondary tasks as one of the leading causes of vehicular accidents on US roads (Dingus et al., 2016). The increasing prevalence of cognitively demanding technology motivates a deeper consideration of the factors that underlie increases in mental workload. We use the term “workload” to refer to the total amount of a finite-processing resource that a task or set of tasks uses (thus “high workload” is induced by a more demanding set of tasks).

Cognitive workload and capacity

Human capacity for processing information is inherently limited (Kahneman, 1973; Townsend & Eidels, 2011), so increases in workload due to multi-tasking have been assumed to result from a reduction in the resources available for each task (see, e.g., Strayer, Cooper, Turrill, Coleman, & Hopman, 2017). This results in a diminished ability to react to changes in all concurrent tasks. For example, when talking on a mobile phone, drivers are typically slower to react to hazards (Strayer et al., 2003), and also exhibit decreased speech complexity (Drews, Pasupathi, & Strayer, 2008). Reduced performance in both tasks suggests multi-tasking involves a general reduction of available cognitive resources rather than a trade-off effect. The (over)utilization of cognitive resources is known as “cognitive workload,” and is often defined as the “amount

✉ Zachary L. Howard
zachary.howard@uon.edu.au

¹ School of Psychology, University of Newcastle,
Callaghan, NSW 2308, Australia

² Department of Psychology, University of Amsterdam,
Amsterdam, The Netherlands

of attention that must be directed to a task” (Lively, Pisoni, Van Summers, & Bernacki, 1993, p. 2962), such that tasks requiring more attentional resources impose a higher cognitive workload.

Importantly, if workload is defined by the amount of attentional resources currently allocated, there is no distinction, in terms of the workload effects, between one demanding task or several less demanding tasks (multitasking) that capture the same total amount of cognitive resources. However, multitasking often affects cognitive processing in unique ways, such as in Franz, Zelaznik, Swinnen, and Walter (2001), who found that the degree of dual-task interference in a finger movement paradigm was related to incongruity between the task instructions. The aim of the present investigation was to directly compare “workload” increases resulting from a more demanding single task (difficulty increases) versus additional tasks (multi-tasking).

Multi-tasking versus single-task difficulty

There is substantial evidence that cognitive workload increases both due to multi-tasking and increasing the demands of a single task (e.g., through a difficulty increase). Multi-tasking has been shown to cause a general increase in cognitive workload across a range of domains using a variety of measures. Goldberg et al. (1998) found that performing the Wisconsin Card Sorting Task and a verbal shadowing task simultaneously led to a significant increase in errors on both tasks, and resulted in a reduction of prefrontal cortex activity, which the authors attributed to the task demands exceeding available capacity. Participants in Rubio, Díaz, Martín, and Puente (2004) self-reported greater workload when performing both a tracking task and a memory search task compared to performing either task alone. Importantly, the lowest reported dual-task load was still greater than the highest reported single-task load, with this increase in workload resulting in poorer performance (in both tasks) when performing both tasks concurrently. Tsai, Viirre, Strychacz, Chase, and Jung (2007) found increased blink rates when performing a Serial Addition Task whilst driving, compared to a driving-only condition. And, as discussed earlier, numerous studies have reported slowed response times from drivers in a Detection Response Task to highlight the load that mobile phone or other technology use imposes on drivers (see, e.g., Strayer et al., 2013; Strayer, Cooper, Turrill, Coleman, & Hopman, 2015; Strayer et al., 2017; Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014).

These examples represent just a small fraction of a body of research demonstrating that completing multiple tasks simultaneously has a measurable impact on cognitive workload. However, in laboratory studies, workload is often manipulated by altering the difficulty of a single task, such as the “n-back” task (see Ayaz et al., 2010; Mehler, Reimer, Coughlin, & Dusek, 2009), without requiring any multi-tasking or task-

switching. For example, Knoll et al. (2011) manipulated load through the difficulty of a reading task and found an increase in workload using EEG (with gamma power increasing with difficulty). Nourbakhsh, Wang, Chen, and Calvo (2012) reinforced these findings using galvanic skin responses. EEG results, particularly suppression of power in the alpha band and reduced P300, have also shown workload increases with the difficulty level of an n-back task (Brouwer et al., 2012), and have distinguished learning methods that require more cognitive resources (Antonenko, Paas, Grabner, & Van Gog, 2010). These same measures have shown dual-task-induced workload increases (Holm, Lukander, Korpela, Sallinen, & Müller, 2009; Lin, Chen, Chiu, Lin, & Ko, 2011). Some of these studies manipulated workload through single-task difficulty as well as through combining tasks (e.g., Rubio et al., 2004).

Many studies implicitly equate the workload effects of difficulty increases with dual-task performance impairments. For example, Strayer et al. (2006) compared drivers using cell-phones (a dual-task) with drivers who were affected by alcohol (compromising performance in the task without adding a second task). However, it is unclear whether dual-task-induced load increases and single-task difficulty manipulations (hereafter “within-task load”) should be equated, and thus whether studies of cognitive workload relying on one or the other kind of workload are comparable (or even investigating the same underlying mechanisms). For clarity, we consider multi-tasking, or dual tasking, to include the addition of a task(s) that requires additional behavioral responses, whereas a difficulty increase keeps the task instructions the same but makes it harder. Dual-task performance often incurs a cost relative to single tasks (Franz et al., 2001), and there is some evidence for neurological processing differences between single and dual tasks (Isreal, Wickens, Chesney, & Donchin, 1980; see Kok, 2001, for a review). In addition, theories that postulate multiple cognitive resource pools (e.g., Wickens, 2002, 2008) support a fundamental distinction between single- and dual-task load increases. We designed a task that would allow simultaneous manipulation of workload *within* a single task (by increasing difficulty) and *between* multiple tasks (by adding additional tasks), which we describe shortly.

The detection response task and evidence accumulation models

To measure workload, we implemented a novel variation of the Detection Response Task (DRT; Strayer et al., 2014, 2015; ISO, 2016). Note that while the typical DRT is administered according to an ISO standard (17488), our task was our own adaptation in an online JavaScript environment, similar to the version validated by Innes, Howard, Evans, Eidels, and Brown (Under Review). We use the term DRT due to the many similarities between our variation and the ISO standard version, but wish to note the important distinction between the

two. The DRT is a useful tool for assessing changes in cognitive resource usage in an objective and quantifiable way. The DRT involves a short, repetitive stimulus (often a light presented in the periphery) that a participant must respond to whilst also completing another task(s). When cognitive workload increases, response times to the DRT stimulus reliably increase, as do the proportion of missed DRT signals (Strayer et al., 2013). DRT response time increases relative to a single-task (DRT only) baseline have been interpreted as an index of cognitive workload. The DRT aligns with the “diminished processing resources” account of cognitive workload, and thus provides a natural measure for our purposes. However, although workload is often assumed to reduce processing resources, the account has rarely been tested directly. Cognitive process-models have been used widely in other areas to understand the effects of aging (Ratcliff, Spieler, & McKoon, 2000), alcohol (van Ravenzwaaij, Dutilh, & Wagenmakers, 2012), reading impairments (Ratcliff, Perea, Colangelo, & Buchanan, 2004) and other factors on human performance. Specifically, these models allow researchers to test whether changes in the observed speed and accuracy are driven by sheer information-processing speed or by other factors, such as caution. For example, Ratcliff et al. (2000) showed that older adults tend to be slower than their younger counterparts in cognitive tasks because they are careful to respond, and that older adults are not necessarily slower at processing information in these tasks. Importantly, accumulator models can tease apart the different mechanisms underlying response times, allowing inferences on the cognitive process instead of the observed performance.

Changes in response latency across conditions and subjects could be the outcome of numerous factors, such as the quality of the signal relative to noise, variability in the time for stimulus encoding and motor preparation, or the tendency of some individuals to be more cautious than others. Each of these mechanisms could cause a similar change in mean response time, but have profoundly different implications. For example, increases in DRT response times with an increase in the number or complexity of tasks have been assumed to reflect a reduction in available cognitive resources (e.g., Strayer et al., 2017; Thorpe, Nesbitt, & Eidels, 2019), implying that the primary task (i.e., not DRT) has become more cognitively demanding. In an evidence accumulation framework (e.g., Evans & Wagenmakers, 2019; Ratcliff, Smith, Brown, & McKoon, 2016) this would be reflected by a diminished processing speed (often referred to as “drift rate”; e.g., Brown & Heathcote, 2008), since the drift rate defines how efficiently evidence is sampled from a stimulus. All things held constant, a higher drift rate leads to a faster process, which might imply there are more resources available. However, increases in response time can also be caused by more cautious responding, or a change in non-decision-related process time (e.g., stimulus encoding; Brown & Heathcote, 2008; Ratcliff,

1978), or different combinations of the above. Therefore, increases in DRT mean response time with an increasing number or complexity of tasks may not necessarily reflect an increase in cognitive demand, as cognitive demand is only one of the different theoretical mechanisms that can cause a change in mean response time.

A recent investigation of the effects of different driving conditions on workload concluded that changes in DRT mean response time did not actually reflect changes in information processing, suggesting that changes in DRT mean response time may not reflect changes in workload (Tillman, Strayer, Eidels, & Heathcote, 2017). Using a driving simulator, Tillman et al. found that drift rate changes were not needed to explain increased DRT response times resulting from conversations with either passengers or via a mobile phone, and instead, only threshold and non-decision time varied between the conditions. This result is surprising given the traditional interpretation of the DRT as picking up residual resources, which implies that any changes in DRT mean response time should be the result of increases in workload, and therefore, decreases in drift rate. As such, a secondary goal of the present work was to directly test whether this common measure of cognitive workload actually captured changes in processing speed. We provide evidence that it does so, confirming the DRT as a suitable tool for the present endeavor. We discuss Tillman et al. further in the *Discussion*.

In the current study we employ evidence accumulation modeling to distinguish these effects separately for dual-task versus difficulty-based workload changes. Our chosen model (linear ballistic accumulation (LBA); Brown & Heathcote, 2008) allows changes in response time to be partitioned into separate components and thus give a more complete account of the cognitive impacts of particular manipulations. We apply a modified one-boundary model in Study 1, and adapt our tasks in Study 2 to be amenable to full two-choice modeling for added robustness. To address whether the cognitive causes of increases in cognitive workload differed between within- and dual-task manipulations, we developed a novel task in which participants tracked moving objects in the Multiple Object Tracking (MOT; e.g., Pylyshyn & Storm, 1988) task, while simultaneously responding to a peripheral stimulus that allows measurement of left-over resources, the DRT (Strayer et al., 2013). The MOT task had three levels of difficulty – no tracking (i.e., do not track any objects), low (track 1 object), and high (track 3 or 4 objects). Critically, the no-tracking condition requires no participant involvement, so DRT performance on the no-tracking MOT condition measures single-task load. Conversely, DRT performance in the low and high MOT load conditions measures multi-tasking load (and the two can be compared to directly assess the effect of difficulty increase, with multi-tasking held constant). Adjusting the number of dots-to-track in an MOT task is a commonly applied difficulty manipulation (see e.g., Cavanagh & Alvarez,

2005; Drew, Horowitz, & Vogel, 2013). Notably, the DRT component remained the same for all trials, and responses to the DRT component were the dependent variable for the modelling analyses. Thus any changes to the response time or lapses (misses) to the DRT can be attributed to the manipulation of load via the tracking task. We applied the LBA (Brown & Heathcote, 2008) model to our data.¹

Our design allowed the mechanisms accounting for the increase in workload to be quantified, and compared between single and dual tasks. We used both a standard DRT (respond when the DRT is displayed, Study 1) and an adapted DRT that required a choice between two DRT signals (red or blue, Study 2). The latter task was introduced to facilitate the application of more robust evidence accumulation modeling, among other benefits. When choices involve two decisions, parameter estimates are better identified (Ratcliff & Strayer, 2014). To foreshadow the outcome, our modelling results suggest the following. First, that completing a simultaneous second task affected workload in a qualitatively different way to changing the difficulty within the existing set of tasks. Second, that changes in DRT response time capture changes in workload by reflecting decreases in processing rate – a finding that has often been assumed, but has not been borne out empirically to date.

Method – Study 1

Participants

Fifty-two undergraduate psychology students from the University of Newcastle were reimbursed with course credit. Participants were required to have normal or corrected-to-normal vision and be able to read English. Participants completed the experiment online, without supervision from an experimenter, which they could only access through the university's online study participation system. In order to minimize the potential for the identification of individual participants within our data set (openly available on the OSF at <https://osf.io/eb5ap/>), our online data collection system was anonymous, and did not involve the collection of any potentially identifying demographic information.

Tasks

Participants tracked objects moving on the screen (MOT) while also responding to a red frame that appeared at random intervals (DRT), as illustrated in Fig. 1. The timelines of the MOT and DRT are illustrated in Fig. 2. Both tasks were administered on

the participant's computer concurrently, with the DRT stimulus displayed as a red border around the MOT display area. The MOT required participants to track the movement of zero (baseline), one (low load), or four (high load) small discs for a period of 15 s. These discs were initially colored blue to identify them as “targets” to be tracked, and turned red after 3 s so they were no longer distinctive from the crowd of other red discs (non-target foils). This procedure ensured participants had to effortfully track the moving objects rather than the color. There were always ten discs in total, and non-target discs were colored red for the entire 15 s. Each disc was circular with a diameter of 14 pixels, corresponding to a visual angle of approximately 6° at 60-cm viewing distance. All discs moved within the display area at a frame rate of 15 frames/s for the duration of a trial, and movement direction was random (instantiated by randomly selecting locations for the dot to move to). Discs could overlap briefly if their paths crossed, and discs bounced randomly away if they reached the edge of the display area (a 150 x 150-pixel square centered in the middle of the screen). At the beginning of the experiment, instructions made clear that when zero objects were to be tracked participants only needed to respond to the DRT and could ignore the moving dots.

The DRT stimulus was a red rectangle that formed a border around the MOT display area. Stimulus presentation adhered to ISO 17488 (Young, Hsieh, & Seaman, 2013), with the DRT being displayed every 3–5 s (sampled uniformly). The maximum time to respond to the DRT stimulus was 2.5 s (button presses recorded after this time were considered misses). The stimulus displayed until the participant responded or for 1 s, whichever was shorter. The DRT display was not shown outside of the 15-s tracking periods, so there were no DRT presentations during instructions, breaks, or MOT decision time.

Procedure

Participants were recruited through an online system for course credit. After signing up, participants completed a 45-

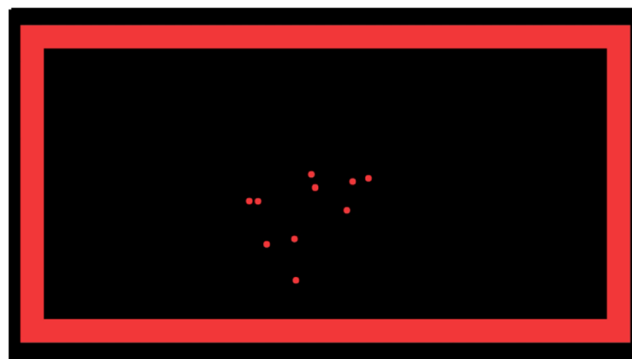


Fig. 1 Illustration of the concurrent display of the Multiple Object Tracking (MOT; red dots) and Detection Response Task (DRT; red border) displays. The DRT component flashed on intermittently and the red dots moved around the display randomly throughout a trial

¹ Note: We also applied the Racing Diffusion Model (Tillman & Logan, 2017), which was applied to similar data by previous authors (Tillman et al., 2017). The results of this model were qualitatively the same as the LBA, and the LBA was preferred by WAIC in all cases. Hence, for brevity, we only report the LBA analyses.

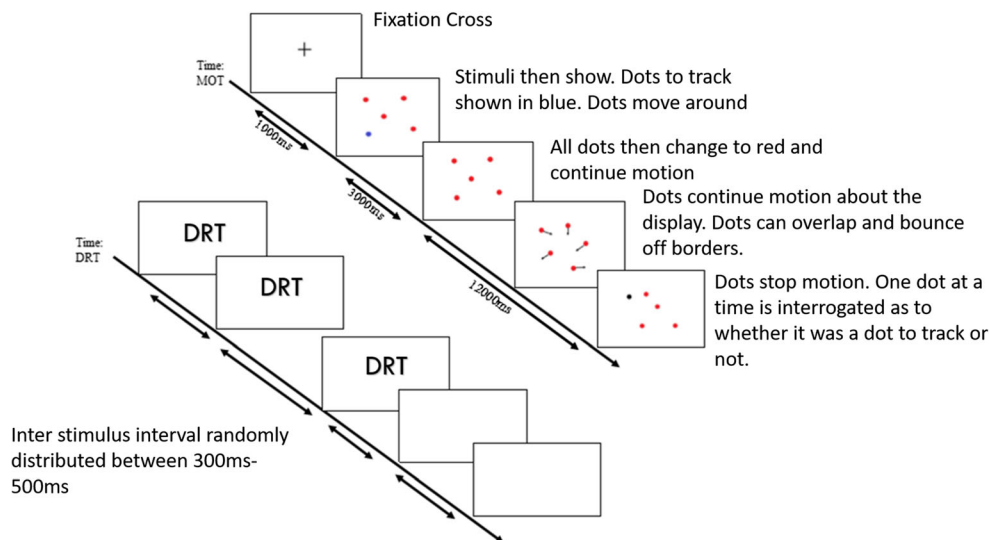


Fig. 2 Timeline of both Multiple Object Tracking (MOT) and Detection Response Task (DRT) tasks for all studies. Note the DRT stops at the MOT interrogation phase to ensure there is no response competition

min online experiment. Participation required a computer with access to an Internet browser and a keyboard, but other factors (e.g., screen size, light levels) were outside of the experimenter's control. After initial instructions regarding both tasks, a practice block consisting of three MOT trials (of 15 s each), with a tracking load of two moving dots, was presented. On each MOT trial, initiated by the participant, the dots to be tracked (targets; zero, one, or four during the experiment, or two in the practice block) were initially colored blue, and the other dots were all colored red. The dots then moved smoothly around the display area for 15 s. After 3 s from movement onset the target dots' color changed from blue to red, and the participants were required to keep track of the targets. During the 15 s, the DRT (a red square bordering the MOT display area) was presented every 3–5 s, and required a key-press response ("T"). After the 15-s tracking period finished, the DRT presentation stopped, the dots stopped moving, and the interrogation stage commenced: a single dot was colored white and participants were asked "was the white dot one of the targets?", with a yes/no response indicated using the keyboard ("O" and "P," respectively). This question was repeated for five randomly selected dots, one at a time.

After three practice trials, participants completed a total of nine blocks with ten MOT trials in each. There were three blocks for each level of load (zero, one, four). The initial order of these three levels was randomized (e.g., four, zero, one), then this order was repeated three times. Participants were encouraged to take short breaks between trials and blocks. The total number of DRT trials in a block could vary slightly depending on the random sampling of presentation times within the 3- to 5-s intervals.

Results – Study 1

Effect of load on DRT measures

Of the 52 participants in Study 1, we excluded ten for inadequate performance. Three of these participants had mean response times of less than 200 ms, suggesting they were "button mashing" (recall there are no incorrect responses). Six participants achieved less than 25% accuracy on the DRT (three of whom did not respond at all), and one participant achieved less than 50% MOT accuracy in the non-tracking condition, suggesting they did not understand the instructions. To confirm we successfully manipulated workload as assessed by the DRT, we compared the accuracy and response time of DRT responses across load levels. There was a very strong effect of load level on DRT response accuracy (proportion of hits; $BF_{10} > 10^{20}$), with accuracy decreasing with load. The same trend held for MOT accuracy ($BF_{10} > 10^{35}$), showing there was no trade-off between tasks. These results are summarized in Fig. 3a. The reverse trend was observed for response times, with a very strong increase in response time across load levels to both the DRT ($BF_{10} > 10^{120}$) and MOT ($BF_{10} > 10^{150}$) tasks. These results are summarized in Fig. 3b.

Modeling approach

We applied an evidence accumulation model to assess which components of the DRT decision-making process changed across workload conditions. We used the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008), allowing up to three parameters to vary between tracking conditions to account for changes in load. Those parameters were the rate of evidence accumulation for the DRT stimulus (*drift rate*),

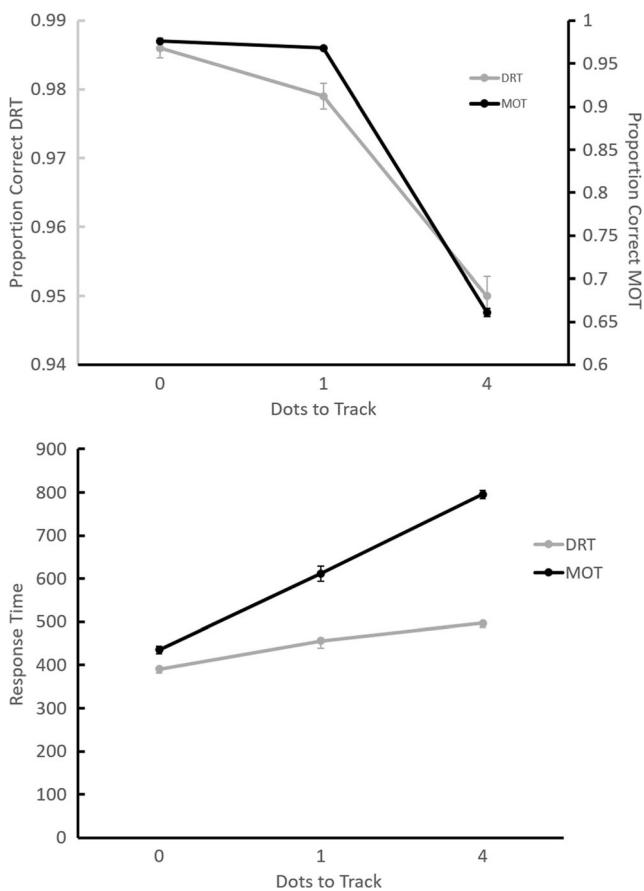


Fig. 3 Accuracy (**top**) and response time (**bottom**) across workload manipulations in the standard Detection Response Task (DRT) detection paradigm and Multiple Object Tracking (MOT) task from Study 1. Manipulating the number of dots to be tracked had a marked effect on both outcome measures in both tasks. Error bars represent the standard error of the mean statistic. NB: Note there are two y-axes in the left panel due to the scale differences between DRT and MOT accuracy

the amount of evidence required to respond to the DRT (*decision threshold*), and the time taken for perceptual and motor processes relating to the DRT (*non-decision time*). In Study 1, we modelled the DRT responses using a single accumulator, since the standard DRT has no choice of response to make. We assumed that missed responses (either no response recorded, or a response time exceeding 2.5s) were the result of an accumulation process that did not reach the threshold before the termination of the DRT trial (Evans, Steyvers, & Brown, 2018; Ulrich & Miller, 1993). This approach allowed misses to inform the predictions of the models, and differs from other recent applications that assume missed responses result from a separate process, such as the accumulator failing to start (e.g., Tillman et al., 2017).

To assess the question of within-task difficulty versus multi-tasking-based workload changes, we compared experimental conditions in two different ways. To assess the effect of difficulty increase, we compared response time in the conditions requiring one and four dots to be tracked. These conditions maintained the same multi-tasking load (in both,

participants were responding to both the DRT and MOT), whereas the *difficulty* of the MOT changed. In contrast, to assess the effect of an additional task we compared response times in the conditions requiring no tracking and four dots to be tracked. In the zero dots condition participants only completed the DRT component of the task, whereas in the four dots condition they were required to complete both the MOT and DRT. For inference, we fit all three load conditions simultaneously, then tested whether specific parameters changed between the conditions. Our interest was in whether drift rate, decision threshold, and/or non-decision time differed between the lower and higher workload conditions, for either the additional or within-task manipulations. For example, we tested for a drift rate decrease under load when adding a task (0 dots to 4 dots), and separately when increasing difficulty (1 dot to 4 dots).

To determine which parameters changed over manipulations, we used a novel combination of a well-known AIC weighting procedure (Wagenmakers & Farrell, 2004) and model averaging, allowing us to obtain a WAIC weight for whether each parameter changed over each manipulation, taking into account potential uncertainty in the true model (Vehtari, Gelman, & Gabry, 2017). We performed this analysis both at the group-level, where inferences have been focused in previous studies (e.g., Tillman et al., 2017), and the individual level, in order to assess individual differences in the effects of parameters. Note that the evidence at the hierarchical level can often be inflated by relatively few participants (Evans, 2019; Evans, Bennett, & Brown, 2019; Evans, Hawkins, & Brown, *in press*). We present these hierarchical level posterior probabilities to be consistent with previous authors, though we tend toward interpreting the individual data to give a more representative overview of our results.

To determine how the parameters changed, we found the posterior mean value of the parameter in each sub model, for each load condition, and determined whether the parameter changed in a positive or negative direction. We then weighted the evidence for each change by that model's posterior WAIC weight for each subject \times sub-model combination. For efficiency, we plot these changes in Fig. 4 as bar plots that show both the total posterior evidence for any parameter change (which can be discerned by the height of the gray bars), as well as the evidence for the direction of changes (represented by the relative amounts of dark to light gray, where more of one shade indicates higher certainty of the parameter's direction of change). We use the darker gray to represent evidence for an *increase* in a parameter over load (i.e., the parameter has a higher value in the 4 dots condition), lighter gray to represent evidence for a parameter *decrease*, and the remainder (white) for evidence that the parameter does not change. For clarity, we have ordered the subject-level posterior plots by the total evidence for any change, such that the right-most participants show stronger evidence, and the left-most

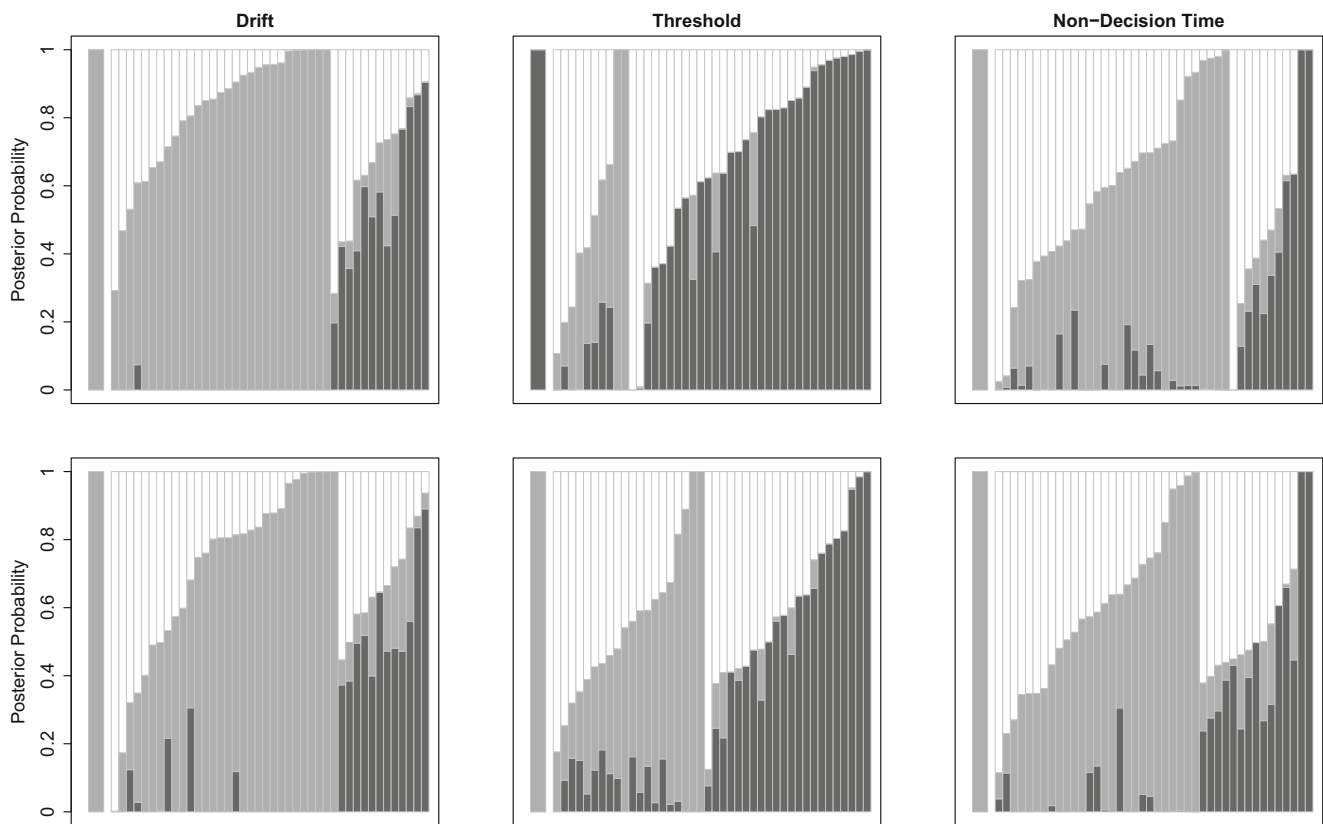


Fig. 4 Hierarchical and individual parameter changes for the LBA model for both additional-task and difficulty-load manipulations in Study 1. The left bar of each panel represents the hierarchical weight, and all other bars show individual subjects ordered by strength of evidence against no-change (but grouped by their strongest direction of change). Each bar represents the total posterior weight (summing to 1), with the proportion

of light gray, dark gray, and white reflecting posterior evidence for a decrease, increase, or no-change in the parameter respectively. Evidence for “a change” can be observed by taking the total light+dark gray area. These results do not speak to the magnitude of change, which is addressed in the following Fig. 5

subjects show evidence *against* a change. However, we have split the subjects into two groups based on whether they show more evidence for a positive or negative change. To illustrate, take the top left panel of Fig. 4a. Most subjects show an almost entirely light gray bar, which means most subjects show (a) strong evidence that the parameter (drift rate) changes, and (b) that almost all this evidence points to a *decrease* under higher workload. To contrast, the threshold parameter in Fig. 4b (middle panel) shows an almost equal split of light versus dark gray, and these bars tend to be shorter. This indicates inconsistent evidence at best that threshold changes between the 1- and 4-dot conditions.

Modeling results

Figure 4 shows group and individual effects of load on model parameters. If a parameter (say drift rate) shows a consistent change with the workload manipulation, we would expect to see the majority of the bars mostly filled with the one shade (suggesting most participants have strong evidence for the same change). For Study 1, there was strong evidence at the group level (posterior weight > .99) that drift rate decreased

with load for both within-task-difficulty and dual-task manipulations. Most individual subjects showed strong evidence that drift changed, and this change was generally a decrease with either type of workload. This confirms that the DRT is reflecting the diminished processing resources that are generally attributed to increased workload. For non-decision time both manipulations showed strong evidence for a change at the group level (posterior weight > .99) in both models, but at the individual level evidence tended to be equally split between “no-change” evidence, and evidence for positive and negative changes. These results suggest some individual variability in the non-decision-time parameter. Examination of the posterior parameter distributions showed the change in non-decision time was very small compared to the other parameters, and as such, it is likely only capturing minor variations in the response-time distributions as other parameters are allowed to change.

Interestingly, opposite changes in threshold were observed at the group level for the additional task manipulation (increase – posterior weight > .99) and difficulty manipulation (decrease – posterior weight > .99). The individual subject results show a clear threshold increase with the additional task

manipulation, but relatively little evidence for any threshold change in the difficulty-only manipulation (with posterior evidence for and against a change ~ 0.5 , and the direction of change approximately 50–50 split between positive and negative across subjects). These results suggest that a threshold change is only really apparent when workload is manipulated by the addition of a task – and the threshold parameter increases in that case.

The above results speak to the certainty and direction of the parameter changes, but tell us little about the relative magnitude of those changes. To understand the parameter changes better, we fit a single-accumulator LBA model where each of the three parameters (drift rate, threshold, non-decision time) were allowed to freely vary between each of the three tracking levels. This was based on the above finding that all three parameters varied between at least two of the conditions. We then examined the group-level posterior parameter distributions directly. We plot the posterior distribution of standardized changes in parameters. Thus, if most of a distribution is far from zero, there is evidence for a larger change (which is comparable between parameters).

Figure 5 shows that drift rate steadily decreased with increased levels of load regardless of the manipulation (i.e., rate decreases with difficulty (1 vs. 4 dots) and an additional task (0 vs. 4 dots)). Posterior p-values² showed high certainty for a decrease in drift rate from tracking 0 to 1 ($p = .044$) to 4 ($p < .0001$), meaning a large proportion of posterior samples show drift rate decreasing with any increase to the number of dots to be tracked. In Fig. 5 we also observe a less pronounced but clear threshold effect, such that threshold increases from the “no-tracking” condition to either level of tracking, but does not substantially differ between those levels.³ Posterior p-values showed high certainty for an increase in threshold for both additional task comparisons, 0–1 ($p < .0001$) and 0–4 ($p = 0.016$), but $p = .64$ for the 1–4 difference (suggesting no particular threshold change in either direction was supported). This is consistent with the mixed individual level posterior weights reported above, and also shows that the threshold change occurs with the addition of both an easy (1 dot) and a difficult (4 dot) task. For non-decision time, the standardized parameter distributions were very small, with almost no discernable deviation compared with the other parameters, so we omit them from the figure. We can conclude that multi-tasking and difficulty increases seem to differentially affect the threshold, but not drift rate, of processing in the DRT,

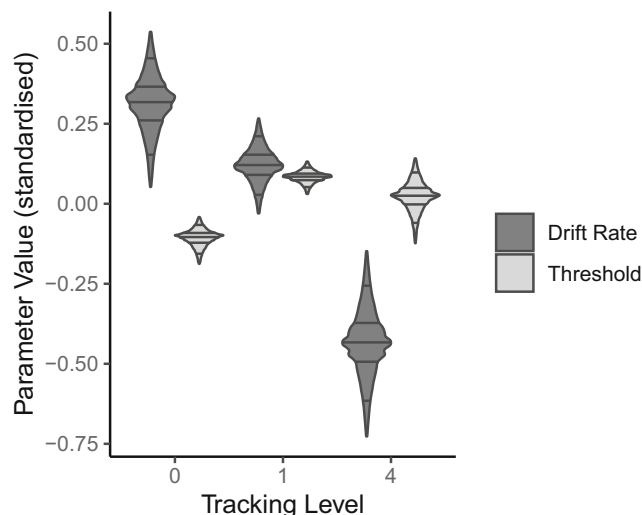


Fig. 5 Posterior parameter distributions for each load level for drift (dark gray) and threshold (blue) parameters. These results were derived from a single-bound LBA model allowing each parameter to vary freely between all three tracking conditions. Parameters were standardized by subtracting the mean value across the combined posterior (for drift and threshold separately) from each posterior point. Each violin reflects the density of the standardized posterior, and the 5th, 25th, 50th, 75th, and 95th quantiles (from bottom to top) are marked with solid lines

suggesting a small but important strategy difference between these manipulations.

Interim discussion

In Study 1 we found evidence for a load-induced decrease in drift rate regardless of the type of load manipulation. This aligns with the “residual capacity” interpretation of both the DRT (Strayer et al., 2013, 2014, 2015; Thorpe et al., 2019) and workload more generally (Kahneman, 1973), and affirms the use of the DRT to address workload-based questions. We also uncovered an interesting distinction between workload induced by the addition of a task (multi-tasking; in our case the addition of a tracking task to the standard detection DRT), and workload increased only through increasing the difficulty within a single task (in our case increasing the number of dots to be followed within a tracking task). In the former, the increase in response time as workload increased was associated with both a decrease in processing speed *and* an increase in the decision threshold; commonly considered a measure of caution (Brown & Heathcote, 2008). This threshold change was not found for the difficulty increase manipulation. Importantly, adding both a relatively easy (1-dot) or more difficult (4-dot) task had the same effect on threshold, suggesting we have observed a multi-tasking specific strategy adjustment. The implication of this finding is that the cognitive mechanisms underpinning workload changes are *not* equivalent between manipulations. Instead, our findings suggest that adding a task results in participants changing their

² Note we frame posterior p-values as a test *against* the direction reported, so that a small p-value corresponds to high certainty. We report this way to be consistent with traditional frequentist p-values.

³ We plot standardized parameter distributions, so the magnitude of changes between drift and threshold is directly comparable.

response strategy, whereby they increase their caution to compensate for the expected load associated with an additional task. In contrast, increasing difficulty within a task seems to result in most participants remaining equally cautious in their DRT responses, suggesting that they do not explicitly react to this increase in difficulty by changing their strategy. This strategy change may reflect some trade-off between maintaining adequate performance *over all tasks* versus maintaining adequate performance *within each task*.

Following the encouraging results of Study 1 we turn to address in Study 2 a few potential concerns. First, although we identified a clear distinction between additional and within-task load manipulations in terms of the effect each had on the threshold parameter of our models, it is possible that presenting all three conditions (i.e., our within-subjects design) may have changed the way participants would have responded to these same manipulations in isolation. Second, modeling single-response decisions can be problematic due to the lack of choice data to precisely constrain parameter estimates. Finally, it could be argued that the baseline “no-tracking” condition in Study 1 was not an ideal comparison because the relative size and salience of the DRT stimulus ensured the task required little cognitive effort from participants. In Study 2 we modified the DRT task such that it required a choice response (color of the stimulus). This allowed parameter estimates to be constrained by both response time and choice data, and also required a more cognitively engaging decision to be made. We also degraded the DRT stimulus by significantly reducing its size and embedding it within noise, with the aim of ensuring even the no-tracking condition required cognitive effort from participants. To address potential concerns over the within-subjects design of Study 1, we also separated the additional and within-task load manipulations into separate participant groups for Study 2, using a between-subjects design. This ensured that the effects of one manipulation were independent of the other, and would provide a useful replication between the two designs.

Method – Study 2

Participants

A total of 127 undergraduate psychology students from the University of Newcastle participated and were reimbursed with course credit. Two versions of Study 2 were run, one with an “additional-task” ($n=57$) workload manipulation, and one with a “within-task-difficulty” ($n=70$) workload manipulation. Participants were randomly assigned to one task or the other upon loading the experiment page. Participants were required to have normal or corrected-to-normal vision and be able to read English. Participants participated online, without supervision from an experimenter.

Tasks

The MOT task was similar to Study 1, with one important change. The type of load (within or dual-task) was a between-subjects manipulation. One group of participants tracked either one or four dots for the duration of the experiment. Another group tracked zero or three dots. Piloting demonstrated that the jump from zero to four (with no “easier” tracking condition as a comparison) led to participants disengaging from the task, thus we reduced the number of dots shown in that condition.

The DRT task was modified to require choices, rather than simple detection. The DRT signal was changed to a single square presented just above the MOT display (see Fig. 6). This stimulus more closely approximates the hardware DRT units often used in driving studies, and also facilitated further manipulations that are not reported here. This stimulus was embedded in noise consisting of uniformly sampled grey or black pixels. The mask was designed both to make the choice component more difficult and to ensure the “no-tracking” condition was still cognitively demanding. The major change to the DRT task was the introduction of the choice component itself. The DRT signal, presented with the same frequency as in Study 1, was red 50% of the presentations, and blue 50% of the presentations (RGB: (255 0 0), and (0 0 255), respectively).

Procedure

Participants were recruited online, and randomly allocated to either the within- or dual-task-load condition. Otherwise, the procedure was mostly similar to Study 1. Practice was identical to Study 1 apart from the DRT task requiring participants to decide if the DRT was red or blue. During the experiment

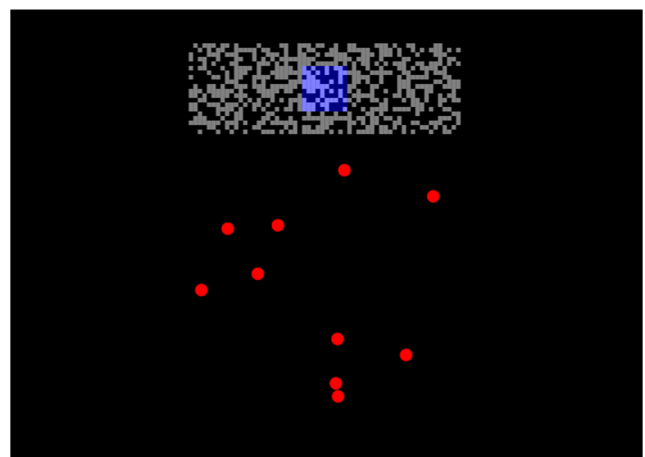


Fig. 6 Illustration of the concurrent display of the Multiple Object Tracking (MOT; red dots) and Detection Response Task (DRT; blue square) displays in Study 2. The DRT component flashed on intermittently and the red dots moved around the display randomly throughout a trial. The DRT stimulus required a choice response (red or blue) and was embedded in visual noise to make the task more demanding

participants were presented with eight blocks of 12 MOT trials; four blocks of each load level. The number of DRT trials varied depending on the participant's response time and the random sampling of trials. Blocks alternated between high and low load, with the initial block randomly selected. Participants in the within-task load type tracked either one or four dots, participants in the dual-task load type tracked either zero or three dots. Participants responded to the DRT stimulus on a standard QWERTY keyboard by pressing the “Q” button if the signal was red, or the “W” button if the signal was blue. Figure 6 shows an example of a trial in progress, with a blue DRT signal embedded in noise.

Results – Study 2

Effect of load on DRT measures

Online data collection again required stringent inclusion criteria, to exclude non-compliant participants. Of the 127 participants who completed Study 2, 19 were excluded. Six participants did not respond to the DRT, and a further four had less than 25% accuracy. Three participants had a mean DRT response time of less than 200 ms. The remaining six participants showed less than 50% accuracy on the lowest tracking level in their respective study. After exclusions there was a total participant pool of 47 participants in the additional task group, and 61 participants in the difficulty group.

To confirm that we increased the baseline demand of the DRT stimulus as intended, we first compared the equivalent tracking levels between Study 1 and Study 2. For the no-tracking condition, response time to the DRT was slower in Study 2 ($M = 627$ ms, $SD = 213$ ms) than in Study 1 ($M = 390$ ms, $SD = 203$ ms), $BF_{10} > 10^{100}$. The same pattern held for the higher tracking conditions, showing the choice-variant of the task increased baseline workload substantially. As with Study 1, we compared the accuracy and response time of DRT responses across load levels as a manipulation check. There was very strong evidence for a decrease in response accuracy (proportion of misses) in both the additional task manipulation ($BF_{10} > 10^{42}$) and the difficulty manipulation group ($BF_{10} > 10^{12}$; see Fig. 7). For both manipulations the corresponding trend again held for response time, with very strong evidence for an increase in response time across tracking levels in the additional task manipulation ($BF_{10} > 10^{242}$), and moderate evidence for an increase in response time across tracking levels in the difficulty manipulation ($BF_{10} > 10^{14}$).

Modeling approach

The modeling approach for Study 2 was simplified by the between-subjects design. The introduction of the choice component (“DRT signals red or blue”) allowed the full

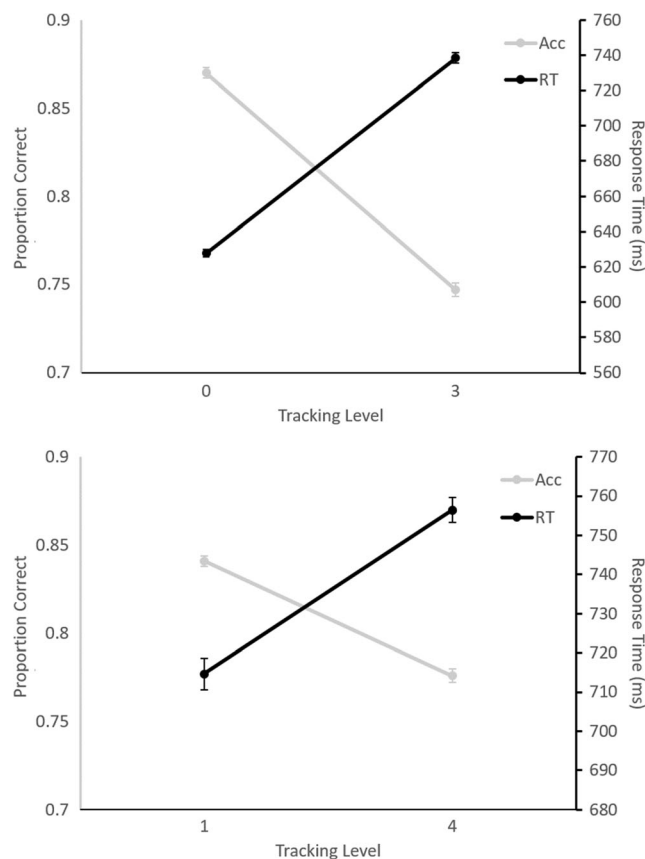


Fig. 7 Accuracy (left axis) and response time (right axis) of both the additional task (**top panel**) and difficulty (**bottom panel**) workload manipulations in the choice and Detection Response Task (DRT) paradigm from Study 2. Manipulating the number of dots to be tracked had a marked effect on both outcome measures. Error bars represent the standard error of the mean statistic. NB: Only DRT results are shown here

specifications of the LBA specified by Brown and Heathcote (2008) to be implemented. This manifested by the introduction of a second accumulator which corresponded to the erroneous response (the blue response, if the stimulus was red, or the red response given a blue stimulus). The red and blue outcomes were not treated separately for the purpose of modelling; instead we allowed a separate drift rate parameter for the target and non-target accumulation processes (regardless of which color was correct on a given trial). We repeated the posterior model weighting process reported for Study 1 here with the caveat that we did not have to fix the level of the “other” load manipulation due to the between subjects design. The below results, although combined for clarity, contain entirely separate participant groups for the “between-” and “within-”task-load manipulations that were modelled independently. The interpretation of the plots is identical to Study 1.

Modeling results

Figure 8 shows group and individual effects of load on model parameters in Study 2. At the group level there was strong

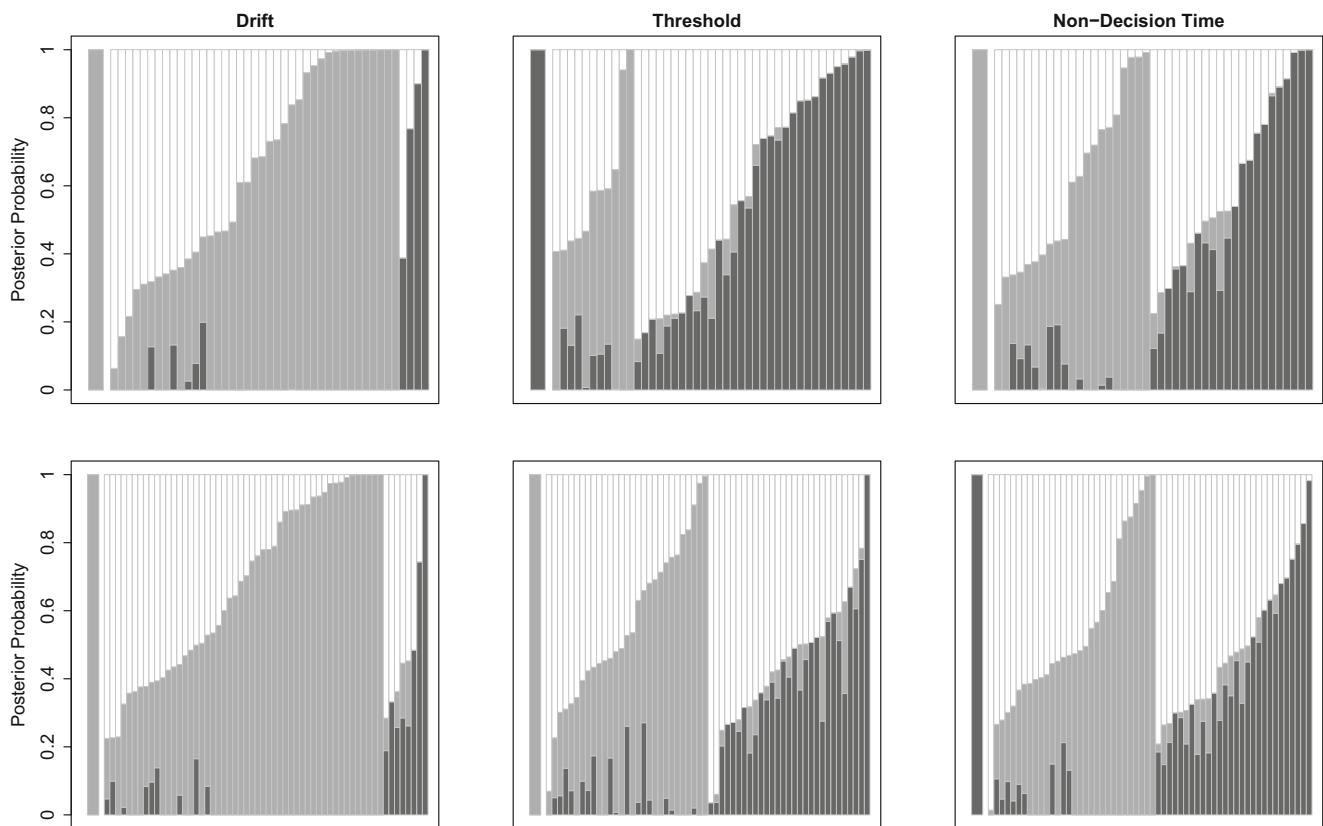


Fig. 8 Hierarchical and individual parameter changes for the LBA model for both additional task and difficulty load manipulations in Study 2. The left bar of each panel represents the hierarchical weight, and all other bars show individual subjects ordered by strength of evidence against no-change (but grouped by their strongest direction of change). Each bar represents the total posterior weight (summing to 1), with the proportion

of light gray, dark gray, and white reflecting posterior evidence for a decrease, increase, or no-change in the parameter, respectively. Evidence for “a change” can be observed by taking the total light+dark gray area. These results do not speak to the magnitude of change, which is addressed in Fig. 9

evidence that drift rate decreased with load for both manipulations (posterior weight $p > .99$ for both tasks), and this trend held for most individuals across both tasks. For non-decision time the manipulations showed opposite trends at the group level, with non-decision time decreasing with the addition of a task (posterior weight $> .99$ for both tasks), but increasing with a difficulty-based increase in load (posterior weight $> .99$ for both tasks). However, these group weights are overstated as the most complex model accounted for almost all of the posterior weight, thus only one parameter direction could be exhibited. At the individual level both manipulations showed relatively little evidence for any change, and evidence was equally split between increase, decrease and no-change. These results suggest significant variability in the non-decision time parameter across individuals, and again the magnitude of these changes was very small.

Importantly, in Study 2 we replicated the finding that increased workload leads to increases in response caution (threshold) when induced by an additional task, but not for the within-task difficulty manipulation. At the group level the additional task manipulation (posterior weight $> .99$) showed

evidence for an increase in threshold with load. The group trends for the difficulty manipulation replicated Study 1, with strong evidence for a decrease in threshold (posterior weight $> .99$) – although again the individual subject results show this is an overstatement. Only the additional task manipulation showed strong evidence for threshold changes at the individual level, and this change was, again, predominantly an increase. As stated earlier, group-level posterior weights can sometimes be strongly influenced by individual subjects with extreme evidence for a particular model (see Evans, 2019, for a brief discussion, and Evans et al., *in press*, for an example), and the strong posterior evidence for *a change* does not speak to the magnitude of any change (which we address in the following paragraph). It should also be noted that within the additional-task group the threshold change was not found in all participants, leaving room for individual variability in the effect.

As with Study 1, we examined the posterior parameter distributions to assess the magnitude of the observed parameter changes. These results generally supported the trends reported in Study 1 and in the weighting analysis of Study 2

reported above. Figure 9a and b show that drift rate decreases substantially with increased levels of load regardless of the manipulation (i.e., rate decreases with difficulty (one vs. four dots) and an additional task (1 vs. 3 dots)). Posterior p-values showed high certainty for a decrease in drift rate from tracking 1 to 4 dots ($p = .012$), and also from 0 to 3 dots ($p = 0.018$), meaning a large proportion of posterior samples show drift rate decreasing with any increase to the number of dots to be tracked. In Fig. 9a we also observe an apparent threshold effect, such that threshold increases from the “no tracking” (0 dots) condition to the “tracking” condition (3 dots). The posterior p-value for this difference was only $p = 0.2$, probably due to the relative uncertainty of the estimated parameters (note the overlap in the tails of the distribution), potentially driven by the individual variability noted above. However, there was no observable difference at all in thresholds for the difficulty manipulation ($p = 0.8$), so the qualitative threshold trends are consistent with Study 1. The non-decision time parameter again had much smaller effects in both manipulations and is not presented in the figure.

General discussion

In our experiments participants were asked to respond to a brief visual stimulus. In Study 1 this was a simple detection task, in Study 2 participants were asked to make a choice between two color options. While responding to this detection/choice task we manipulated participants workload by simultaneously introducing a visual tracking task with several levels, with the specific aim of comparing workload shifts from the “no tracking” condition to either of the tracking conditions (i.e., load induced by the addition of a task) against

workload shifts from the low to high level tracking (i.e., load induced by increasing difficulty only). Studies 1 and 2 identified and replicated several key findings. First, as anticipated, the primary driver of increased response time to the DRT stimuli as workload increased was slower processing rates. In Study 1, the standardized effect of drift rate was much greater than the equivalent threshold effect (see Fig. 5). This finding supports the general interpretation of the DRT (see, e.g., Strayer et al., 2013, 2017), i.e., that increased “workload” reflects an increasingly diminishing pool of processing resources.

It is entirely expected that the primary driver of response time changes in the DRT task is processing speed. The theoretical underpinning of the task is that of a shared, limited capacity pool of cognitive resources (see, e.g., Strayer, Watson, & Drews, 2011; Thorpe et al., 2019). The drift rate of an accumulator model “maps the speed of information uptake” (Voss, Nagler, & Lerche, 2013, p.4), and would naturally be expected to decrease if there were less available processing resources. In fact, drift rate has previously been linked to cognitive workload using the theoretical framework of Systems Factorial Technology (SFT; Eidels, Donkin, Brown, & Heathcote, 2010; Endres, Houpt, Donkin, & Finn, 2015; Townsend & Eidels, 2011). Despite this, the only previously reported analysis of DRT response times using an accumulator model did not implicate drift rate effects (Tillman et al., 2017). As such, our identification (and replication) of lower drift rates as the primary driver of both multi-tasking and difficulty-based workload is an important contribution to the theoretical landscape of attention and workload. We also found that workload increases attributable to the addition of a task resulted in an increased response threshold. This change was not observed when workload resulted only

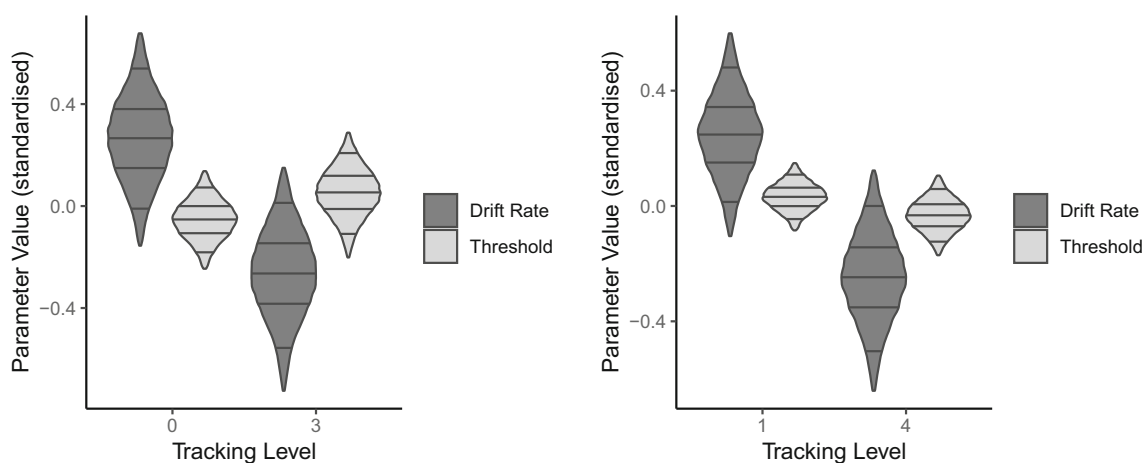


Fig. 9 Posterior parameter distributions for each load level for drift (dark gray) and threshold (blue), for both additional task (a) and difficulty based (b) workload manipulations. These results were derived from an LBA model allowing each parameter to vary freely between the two tracking conditions (the two panels show models fit to different data).

Parameters were standardized by subtracting the mean value across the combined posterior (for each parameter separately) from each posterior point. Each violin reflects the density of the standardized posterior, and the 5th, 25th, 50th, 75th and 95th quantiles (from bottom to top) are marked with solid lines

from an increase in difficulty of the same task, but was found whether the additional task was relatively easy (track 1 dot) or relatively difficult (track 3–4 dots). In Study 2 we found this trend held even where the difficulty of the baseline task was increased (by switching to a choice task) but the tracking task remained the same.

Our threshold results align with a recent study by Tillman et al. (2017). Tillman et al. performed a simulated driving performance study where workload was manipulated by adding a conversation task (with either a passenger in the room, or via mobile phone), and measured using the ISO standard DRT task. While those authors did not find evidence implicating drift rate in workload changes, their findings do support our multi-tasking-specific threshold findings. Specifically, Tillman et al found that compared to the baseline “driving only” condition, speaking with a passenger or via mobile phone was associated with increased threshold. However, there was no evidence that threshold changed between the type of conversation (in room or via mobile phone) - a manipulation that would only affect difficulty in their paradigm. This suggests participants fundamentally adjust their processing strategy to account for changes in workload when additional tasks are to be completed concurrently, but not when difficulty of the current task palette increases.

It is unclear why such a strategy difference arises. One potential explanation (helpfully raised by a reviewer) is that participants might be more salient to an addition of a task, whereas difficulty changes might go undetected. However, in both of our Experiments two features preclude this explanation. First, the wording changes between 0, 1, and 3–4 dots were identical. Participants were not specifically told that there would be an additional task. Second, no two consecutive blocks had the same tracking load. Therefore, the participants task demands changed at every block, yet they only adjusted their thresholds when moving from not tracking to tracking. It is possible that more salient instructions (e.g. “now you will work on a more difficult task”) might change the effect, and future work could consider this. Another account may be that truly concurrent performance of tasks is unlikely – instead “multi-tasking” may in practice be rapid switching between tasks (Salvucci, Taatgen, & Borst, 2009). Rapidly switching between tasks comes with a performance cost (Karayanidis, Coltheart, Michie, & Murphy, 2003), so our observed threshold adjustments might be counteracting these effects. Whatever the explanation, the differential cognitive adjustments made to multi-tasking and difficulty-based workload increases have important implications for studies of workload, particularly those intending to compare manipulations of different types. Both decreased drift rates and increased threshold can increase response times (Ratcliff et al., 2016) - thus response time alone should not be used to quantitatively compare absolute workload levels without deeper investigation (such as the accumulator models applied in the present article).

Why do our drift rate results disagree with Tillman et al. (2017)? Applying the same accumulator model to conceptually similar tasks, our results strongly implicate diminished processing resources as the primary driver of response-time differences between load levels, while Tillman et al found workload was entirely captured by threshold. Tillman et al. suggested their results were consistent with the DRT and driving/conversation tasks tapping different resource pools. Our results may then suggest the tasks tapped a shared resource pool. The primary difference between our own task and Tillman et al.’s conversation manipulations are that our manipulations share modality with the detection task (all our tasks are visual), whereas conversations and visual detection tap primarily different modalities (working memory versus visual). It is plausible that rather than a single shared capacity pool, a multiple resource pool account of cognitive workload (Wickens, 2002, 2008), in which different modalities access different cognitive resources, is more suitable. Under the assumption that there are multiple resource pools, the visual DRT may be less sensitive when the primary task involves a non-visual modality, as the visual DRT is unlikely to be effectively measuring the residual processing resources left over from a non-visual task. However, further research is needed to properly assess whether the measurement properties of the DRT depend on the primary task modality.

A final note is that our reported results rely on our definition of increasing tracking levels as only influencing the difficulty of the task. The number of dots-to-track in an MOT paradigm has often been applied as a difficulty manipulation in the literature (see, e.g., Cavanagh & Alvarez, 2005; Drew et al., 2013, for two examples). Although this treatment is commonplace, an argument could be made that each additional dot-to-track constitutes another “task.” We favor the view that additional tasks require additional behavioral responses (e.g., responding to the DRT *and* tracking any number of dots, compared with tracking one or four dots). In a similar vein we recognize that our difficulty changes may not be linear, i.e. each additional dot may not alter workload to the same degree. We believe the results of Experiment 1 mitigate this issue, given the threshold change was found when comparing either 0–1 or 0–4 dots (suggesting the change was inherent to the addition of the tracking task rather than the difficulty of that task).

Conclusion

The primary question addressed in this manuscript was whether multi-tasking is simply a form of difficulty increase. Our results strongly suggest that this is not the case. There are fundamental differences in the way the cognitive processing system accounts for these two load manipulations. The major implication of this finding is that the relative workload effects

of some manipulations should not be directly compared. Returning to Strayer et al. (2006), they compared the effect of mobile-phone usage on driving performance (an additional task) against alcohol consumption (an effect that may make driving more difficult, but does not add an additional set of behaviors to perform). Our results suggest such direct comparisons may be unwise as the observed response-time changes may not solely result from changes in “workload,” but also strategic adjustments. Importantly, we have confirmed that the DRT captures elements of processing capacity through changes in drift rate, thus the traditional interpretation of the measure is largely unaffected. Our results also hint at deeper theoretical conceptualizations of workload (e.g., single vs. multiple resource pools); however, further work is needed to concretely distinguish these theoretical accounts. Future research should also investigate more nuanced kinds of workload. For example, people with high depression scores often fail to exhibit typical cognitive control mechanisms when exposed to emotionally valenced stimuli (Compton et al., 2008; Williams, Howard, Ross, & Eidels, 2018), which might suggest the emotional stimuli induce “cognitive overload” in that group. How such effects fit within the present discussions remains unknown. Further exploration of how other capacity limitations, such as the “within-task” capacity measures of Systems Factorial Technology (such as those observed by Eidels, Houpt, Altieri, Pei, & Townsend, 2011; Garrett, Howard, Houpt, Landy, & Eidels, 2019; Howard, Garrett, Little, Townsend, & Eidels, Under Review) relate to more general “system-level” workload are also needed.

Author Note This research was supported by an Australian Government Research Training Program (RTP) Scholarship awarded to Zachary Howard.

Data Availability The data and materials for all experiments are available at <https://osf.io/eb5ap/> and neither experiment was preregistered.

References

- Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22(4), 425–438.
- Ayaz, H., Willems, B., Bunce, B., Shewokis, P. A., Izzetoglu, K., Hah, S., ... Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations*, 21–31.
- Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4), 045008.
- Brown, S. D., & Heathcote, A. J. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in cognitive sciences*, 9(7), 349–354.
- Compton, R. J., Lin, M., Vargas, G., Carp, J., Fineman, S. L., & Quandt, L. C. (2008). Error detection and post error behavior in depressed undergraduates. *Emotion*, 8(1), 58.
- Cooper, J. M., Ingebreetsen, H., & Strayer, D. L. (2014). Mental workload of common voice-based vehicle interactions across six different vehicle systems.
- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113(10), 2636–2641.
- Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14(4), 392.
- Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human factors*, 51(5), 762–770.
- Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition*, 126(2), 213–223.
- Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic bulletin & review*, 17(6), 763–771.
- Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of mathematical psychology*, 55(2), 176–190.
- Endres, M. J., Houpt, J. W., Donkin, C., & Finn, P. R. (2015). Working memory capacity and redundant information processing efficiency. *Frontiers in psychology*, 6, 594.
- Evans, N. J. (2019). Assessing the practical differences between model selection methods in inferences about choice response time tasks. *Psychonomic Bulletin & Review*.
- Evans, N. J., & Wagenmakers, E. J. (2019). Evidence accumulation models: Current limitations and future directions. *The Quantitative Methods for Psychology*.
- Evans, N. J., Bennett, A. J., & Brown, S. D. (2019). Optimal or not; depends on the task. *Psychonomic Bulletin & Review*, 26(3), 1027–1034.
- Evans, N. J., Hawkins, G. E., & Brown, S. D. (in press). The role of passing time in decision-making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Evans, N. J., Steyvers, M., & Brown, S. D. (2018). Modeling the covariance structure of complex datasets using cognitive models: An application to individual differences and the heritability of cognitive ability. *Cognitive science*, 42(6), 1925–1944.
- Franz, E. A., Zelaznik, H. N., Swinnen, S., & Walter, C. (2001). Spatial conceptual influences on the coordination of bimanual actions: When a dual task becomes a single task. *Journal of motor behavior*, 33(1), 103–112.
- Garrett, P. M., Howard, Z., Houpt, J. W., Landy, D., & Eidels, A. (2019). Comparative estimation systems perform under severely limited workload capacity. *Journal of Mathematical Psychology*.
- Goldberg, T. E., Berman, K. F., Fleming, K., Ostrem, J., Van Horn, J. D., Esposito, G., ... Weinberger, D. R. (1998). Uncoupling cognitive workload and prefrontal cortical physiology: a pet rcbf study. *Neuroimage*, 7(4), 296–303.
- Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. (2009). Estimating brain load from the eeg. *The Scientific World Journal*, 9, 639–651.
- Howard, Z. L., Garrett, P. G., Little, D. N., Townsend, J., & Eidels, A. (Under Review). Nice Guys Check Twice: No-Signal Processes in Systems Factorial Technology. *Psychological Review*
- Innes, R. J., Howard, Z. L., Evans, N. J., Eidels, A., & Brown, S. (Under Review). A broader application of the detection response task to cognitive tasks and online environments. *Human Factors*.

- International Standards Organization. (2016). Road vehicles—Transport information and control systems—Detection Response Task (DRT) for assessing attentional effects of cognitive load in driving (ISO 17488).
- Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human factors*, 22(2), 211–224.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Prentice-Hall Englewood Cliffs, NJ.
- Karayanidis, F., Coltheart, M., Michie, P. T., & Murphy, K. (2003). Electrophysiological correlates of anticipatory and post stimulus components of task switching. *Psychophysiology*, 40(3), 329–348.
- Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., & Zarjam, P. (2011). Measuring cognitive workload with low-cost electroencephalograph. In *Ijfi conference on human-computer interaction* (pp. 568–571).
- Kok, A. (2001). On the utility of p3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3), 557–577.
- Lin, C.-T., Chen, S.-A., Chiu, T.-T., Lin, H.-Z., & Ko, L.-W. (2011). Spatial and temporal eeg dynamics of dual-task driving performance. *Journal of neuroengineering and rehabilitation*, 8(1), 11.
- Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *The Journal of the Acoustical Society of America*, 93(5), 2962–2973.
- Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*(2138), 6–12.
- Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian computer-human interaction conference* (pp. 420–423).
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3), 179–197.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2), 59.
- Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain and Cognition*, 55(2), 374–382.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4), 260–281.
- Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, 7(1), 1–25.
- Ratcliff, R., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic bulletin & review*, 21(3), 577–589.
- Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Applied Psychology*, 53(1), 61–86.
- Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multi-tasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1819–1828).
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile.
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015). *Measuring cognitive distraction in the automobile iii: A comparison of ten 2015 in-vehicle information systems*. AAA Foundation for Traffic Safety.
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2017). The smartphone and the driver's cognitive workload: A comparison of apple, google, and microsoft's intelligent personal assistants. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(2), 93.
- Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human factors*, 48(2), 381–391.
- Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied*, 9(1), 23.
- Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). Measuring cognitive distraction in the automobile ii: Assessing in-vehicle voice-based. *Accident Analysis & Prevention*, 372, 379.
- Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multi-tasking in the automobile. In *Psychology of learning and motivation* (Vol. 54, pp. 29–58). Elsevier.
- Thorpe, A., Nesbitt, K., & Eidels, A. (2019). Assessing game interface workload and usability: A cognitive science perspective. In *Proceedings of the Australasian computer science week multiconference* (p. 44).
- Tillman, G., & Logan, G. D. (2017). The racing diffusion model of speeded decision making.
- Tillman, G., Strayer, D., Eidels, A., & Heathcote, A. (2017). Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception, & Psychophysics*, 79(6), 1795–1803.
- Townsend, J. T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic bulletin & review*, 18(4), 659–681.
- Tsai, Y.-F., Viire, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(5), B176–B185.
- Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, 37(4), 513–525.
- van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2012). A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology*, 219(4), 1017–1025.
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5), 1413–1432.
- Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental psychology*.
- Wagenmakers, E.-J., & Farrell, S. (2004). Aic model selection using akaike weights. *Psychonomic bulletin & review*, 11(1), 192–196.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2), 159–177.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50(3), 449–455.
- Williams, P., Howard, Z., Ross, R., & Eidels, A. (2018). Cognitive dysfunction under emotional exposure: When participants with depression symptoms show no cognitive control. *Australian Journal of Psychology*, 70(4), 378–387.
- Young, R. A., Hsieh, L., & Seaman, S. (2013). The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load.