

A broader application of the detection response task to cognitive tasks and online environments

Reilly J. Innes^a, Nathan J. Evans^{b,c}, Zachary L. Howard^a, Ami Eidels^a and Scott D. Brown^a

^a School of Psychology, University of Newcastle, Australia

^b Department of Psychology, University of Amsterdam, The Netherlands

^c School of Psychology, University of Queensland, Australia

Word count: 4923

This research was supported by an Australian Government Research Training Program (RTP) Scholarship awarded to RJI and ZLH. NJE was supported by an Australian Research Council Discovery Early Career Researcher Award (DE200101130). Correspondence concerning this article may be addressed to: Reilly Innes, School of Psychology, University of Newcastle, Callaghan NSW 2308, Australia; Email: Reilly.Innes@uon.edu.edu.au

Abstract

Objective

The present research applied a well established measure of cognitive workload in driving literature to an in-lab paradigm. We then extended this by comparing the in-lab version of the task to an online version.

Background

The accurate and objective measurement of cognitive workload is important in many aspects of psychological research. The detection response task (DRT) is a well-validated method for measuring cognitive workload that has been used extensively in applied tasks, for example to investigate the effects of phone usage or passenger conversation on driving, but has been used sparingly outside of this field.

Method

The study investigated whether the DRT could be used to measure cognitive workload in tasks more commonly used in experimental cognitive psychology, and whether this application could be extended to online environments. We had participants perform a multiple object tracking task while simultaneously performing a DRT. We manipulated the cognitive load of the multiple object tracking task by changing the number of dots to be tracked.

Results

Measurements from the DRT were sensitive to changes in the cognitive load, establishing the efficacy of the DRT for experimental cognitive tasks in lab-based situations. This sensitivity continued when applied to an online environment (our code for the online

DRT implementation is freely available at <https://osf.io/dc39s/>), though to a reduced extent compared to the in-lab situation.

Conclusion

The multiple object tracking task provides an effective manipulation of cognitive workload. The DRT is sensitive to changes in workload across a range of settings, is suitable to use outside of driving scenarios, and via online delivery.

Application

Methodology shows how the DRT could be used to measure sources of cognitive workload in a range of human factors contexts.

Keywords:

Cognitive Workload ; Multiple Object Tracking ; Detection Response Task.

Prècis:

The multiple object tracking task is shown as an effective operator of cognitive workload. The sensitivity and reliability of the cognitive workload measure used - the detection response task - is furthered.

Background

The ability to process a useful amount of incoming information from our environment is referred to as “cognitive capacity” (Eidels, Donkin, Brown, & Heathcote, 2010). The most important property of cognitive capacity is that it is limited; that is, we can rarely process all the information available (Kahneman, 1973). Cognitive capacity limitations have been central in almost all major theories of attention (e.g., Broadbent, 2013; Kahneman, 1973), perception (e.g., Treisman & Gelade, 1980), and memory (e.g., Miller, 1956; Baddeley & Hitch, 1974). Cognitive capacity limitations remain an important topic of investigation in many fields of research (Gladstones, Regan, & Lee, 1989; Pashler, 1994; Strayer & Johnston, 2001; Mayer & Moreno, 2003; Townsend & Eidels, 2011). Cognitive load theory plays an important role in learning and education (De Jong, 2010), and in human factors (design) research in engineering, where cognitive load influences the usefulness of systems and human machine interface (Endsley, 1995), and many other areas of applied research. Despite cognitive capacity limitations often forming a key assumption in theories across such diverse domains, objective and well-validated measures of cognitive workload are relatively unexplored.

The NASA Task Load Index (TLX) (Hart & Staveland, 1988) measures workload using a questionnaire delivered after completing a task. However, despite the wide use of the TLX across an array of industries, the subjective nature of the task can compromise findings and fails to give insight into workload changes throughout a task. Without a more quantifiable and objective measurement tool, a serious limitation is presented to the kinds of research questions that can currently be addressed. Research in physiological measures such as electroencephalography (Berka et al., 2007; Stevens, Galloway, & Berka, 2007), galvanic skin response (Patten, Kircher, Östlund, & Nilsson, 2004) and eye tracking (Ahlstrom & Friedman-Berg, 2006; Tsai, Viirre, Strychacz, Chase, & Jung, 2007) have also

aimed to address this workload measurement problem, with some success. These measures are promising, but have other limitations, such as difficulties in interpretation of results, and in accounting for performance changes on the task at hand. An example of this can be seen in work by Brouwer et al. (2012) where EEG power and event related potentials from an n -back task showed inconclusive and contradictory information as workload increased in the main task.

Stimulus–response tasks provide an alternative approach to measuring workload, with different strengths and weakness than physiological and questionnaire-based methods. Stimulus–response tasks have been used extensively within cognitive psychology, often to investigate perceptual processing ability, however, these paradigms have also been used to investigate attention, arousal, fatigue and distraction. Stimulus–response tasks simply require participants to respond as quickly as possible to external stimuli. These measures have been used in research on fatigue and in dual-task research. Examples include the “brake response task” (Strayer & Johnston, 2001; Strayer & Drew, 2004; Strayer, Drews, & Crouch, 2006), the “peripheral detection task” (PDT: Van Winsum, Herland, & Martens, 1999; van Winsum, 2019), the “psychomotor vigilance task” (PVT: Dinges & Powell, 1985; Dinges et al., 1997; Roach, Dawson, & Lamond, 2006; Basner & Dinges, 2011, 2012; Ratcliff & Strayer, 2014), the “tactile detection task” (TDT: Diels, 2011) and the “detection response task” as in ISO:17488 (2016) (DRT: Strayer et al., 2013, 2015; Strayer, 2015; Cooper, Castro, & Strayer, 2016; Castro, Cooper, & Strayer, 2016; Strayer, Cooper, Turrill, Coleman, & Hopman, 2017; Castro, Strayer, Matzke, & Heathcote, 2019; Howard, Evans, Innes, Brown, & Eidels, 2019). The brake response task requires participants to brake as quickly as possible to traffic hazards in simulated, and real world, driving environments. In the braking response task, the frequency of stimuli (hazards) can vary greatly. The PDT, proposed by Van Winsum et al. (1999), requires participants to respond to stimuli

which are peripheral to a main task. In the PDT, a stimulus is presented in the periphery and occurs frequently (every three to five seconds). The PVT differs from these paradigms because it is a single task reaction time measure, where participants have no distractions to induce cognitive load. The PVT has been widely used in the study of fatigue and sleep deprivation (Basner & Dinges, 2012), with applications in health and safety. The PVT requires responses to an increasing timer, which is the only salient stimulus in a display. This PVT stimulus is less frequent than the PDT (elicited every two to ten seconds), and aims to assess vigilance rather than workload.

The DRT originates from the PDT and shares many similarities, however, as Stojmenova, Jakus, and Sodnik (2017) discuss, the main disadvantage of the PDT is that it is in the periphery which can result in increased misses. The DRT uses a more salient stimulus; often a vibration on the skin, or a head-mounted light. The DRT is used to measure the residual attentional resources of a participant in a dual-task design. The DRT was standardized under ISO:17488 (2016) in an attempt to bring consistency to the tasks used in cognitive workload and driver distraction literature, including the PDT, TDT, and DRT. Studies from Van Winsum et al. (1999); Merat and Jamson (2008); Engström, Larsson, and Larsson (2013); Young, Hsieh, and Seaman (2013) and Diels (2011) have shown evidence for these tasks' sensitivity to changes in workload across a variety of stimulus modality and contextual environments, including real world and simulated driving. In a standardized DRT experiment, participants are required to respond to a simple but salient peripheral stimulus (e.g., a small salient light or vibration) that is repeatedly presented at random intervals (between three and five seconds), while they complete other tasks (e.g., driving). Changes in the distribution of response times to the peripheral stimulus indicate cognitive workload change: fast responses suggest lower cognitive workload, and slow responses suggest higher cognitive workload (Strayer et al., 2013, 2015).

The DRT (ISO:17488, 2016) has allowed successful measurement of the cognitive workload of drivers in many situations, with demonstrated external validity (Young et al., 2013; Engström et al., 2013; Strayer et al., 2013, 2015). Despite this success, there have been relatively few attempts to extend the applications of the DRT beyond driving research. Many researchers have pointed out the need for quantified and concurrent workload evaluation in various human factors environments, such as education settings (De Jong, 2010) and workplace environments (De Croon, Sluiter, Kuijer, & Frings-Dresen, 2005). De Jong (2010) notes the potential of dual-task cognitive workload measures in education settings, as a means of evaluating learning environments or the impact of workload on performance, however, there are limited examples of such use in these fields. Brünken, Steinbacher, Plass, and Leutner (2002) and Brünken, Plass, and Leutner (2004) provide examples of successful dual-task cognitive load measures in education settings – using a change detection task and a stimulus-response task respectively – in multimedia learning environments. These studies evaluated the effects of multi-modal materials on learning and workload, and the results indicated higher workload with the presence of more tasks or stimuli. Brock, Stroup, and Ballas (2002) investigated the workload consequences of design changes in US Navy workstations. They employed a dual-task paradigm and found that spatialized auditory displays improved main task performance without a cost to the secondary task. Evidently, workload may have an effect on attention, distraction, learning and performance across a range of environments, but it is not regularly evaluated or considered within these contexts. Using dual-task measures of cognitive workload such as the DRT – which shows reliability and validity in other human-factors contexts – could provide the access to cognitive-workload evaluation needed to motivate the uptake of this practice.

The DRT has been used effectively in driving research to demonstrate the distracting

effects of secondary tasks such as mobile phone usage (Strayer et al., 2013, 2015, 2017, 2019), talking to passengers (Strayer et al., 2015; Tillman, Strayer, Eidels, & Heathcote, 2017), as well as the workload induced by interacting with in-vehicle information systems (Coleman, Turrill, Cooper, & Strayer, 2016; Strayer, Cooper, Turrill, Coleman, & Hopman, 2016; Strayer et al., 2019), and many other examples. Other uses of the DRT outside of driving contexts also exist, such as evaluating user interfaces for forklift operators (Gross, Bretschneider-Hagemes, Stefan, & Rissler, 2018), assessing game design (Thorpe, Nesbitt, & Eidels, 2019), evaluating helicopter pilot information displays (Innes, Howard, Thorpe, Eidels, & Brown, 2019) and evaluating group workload (Xie et al., 2016). These studies highlight the potential of the DRT to provide benefits beyond its usual application in driving research. Each study consistently showed evidence for the sensitivity of the DRT to various applied manipulations. Despite this, there have been few attempts to use the DRT more generally – for example, as a measure of cognitive workload in laboratory-based investigations or furthered as an indicator of cognitive capacity. It is foreseeable that with an appropriate experimental design, the DRT could be practical and useful for measuring workload in controlled laboratory settings.

Previous research has shown some utility in controlled environments, such as Engström et al. (2013) and Young et al. (2013) who demonstrated that changes in the cognitive demands of an n -back task are reflected in corresponding DRT data. Furthermore, several studies have evaluated the effects of stimulus presentation modality (Merat & Jamson, 2008; Conti, Dlugosch, Vilimek, Keinath, & Bengler, 2012; Stojmenova et al., 2017; van Winsum, 2019), with results showing tactile, auditory and visual stimuli are all sensitive to workload change. These results provide a positive start, but are restricted to the atypical stimulus and response modalities used (auditory stimulus presentation with vocal responses). van Winsum (2019) provided an insight into this phenomenon, by demon-

strating the sensitivity and underlying effects of the visual DRT in a series of experiments, demonstrating that cognitive load, manual load, and optic flow all contributed to differences in response times. Several other studies (Bruyas & Dumont, 2013; Schindhelm & Schmidt, 2015) have evaluated the utility and sensitivity of the DRT in more controlled settings, however, in order to improve the generalizability of these studies to real-world problems, it is important to use a continuous task is used to replicate the fluidity and optic flow of applied environments. Examples of these real-world problems include large scale testing of interfaces, user displays, or user-interface – similar to research by Thorpe et al. (2019); as well as other practical uses such as individual cognitive workload comparison for personnel selection – such as in Innes, Howard, Eidels, & Brown, 2018. The use of a continuous task, similar to that of driving, gaming, or machine operation, can minimize the confounding effects related to the peculiarities of lab-based tasks requiring only intermittent attention.

We aimed to expand on this research by evaluating the DRT in a task and situation that shared greater similarity with standard experimental cognitive psychological research than previous DRT assessments. From this, we aimed to implement an easy-to-distribute version of this task, which could potentially be used in applications such as design testing (for example, see Thorpe et al., 2019) and personnel assessment (for example, see Innes et al., 2018). Researchers working in a range of practical environments, such as learning, personnel selection, training, interface evaluation, and gaming, have noted the potential benefits of an approach like ours, which allows for real-time, continuous, and quantified measurement of cognitive workload and task performance (De Jong, 2010; De Croon et al., 2005; Haapalainen, Kim, Forlizzi, & Dey, 2010). Thus, with the aim of allowing quantitative cognitive workload measurement across a broader range of human factors domains, we provide an easy-to-distribute, controlled DRT paradigm which has potential

implementation across a variety of scenarios and tasks.

We used the multiple object tracking task (MOT: Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999) as a primary task to manipulate cognitive workload. The MOT is a widely-used and extensively-studied paradigm which presents participants with a set of smoothly-moving dots on a screen, and requires the participants to keep track of a subset of those dots. The workload of the task increases as the number of dots that the participant is required to track increases. By using a continuous tracking task, whilst also assessing main task performance, the sensitivity and usability of the DRT can be observed, and then compared to a more distributable and widely accessible version of the task. Experiment 1 had participants complete an MOT with three levels of workload manipulated within-subject while simultaneously performing a standard tactile DRT (tDRT), which was performed in a standard lab-based setup. Experiment 2 attempted to extend this to an online environment – which are becoming more common within experimental cognitive research as noted by Zwaan et al. (2017) – via the use of a visual DRT (vDRT) stimulus that was incorporated into the display, in order to increase the potential for easier delivery of the procedure (i.e. over the internet).

Experiment 1

Method

Participants. Participants were 51 undergraduate psychology students (average age of 22; 42 females; 7 left handed) from the University of Newcastle, who were reimbursed with course credit. All participants had normal or corrected-to-normal vision and were able to read English. This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Administration at the University of Newcastle (HREC-2013-0250). Informed consent was obtained from each

participant.

Tasks. Participants simultaneously carried out a MOT and a tDRT, illustrated by the upper and lower timelines (respectively) in Figure 1. The MOT was displayed on a computer in front of the seated participants, and the tDRT was run on a separate computer with tDRT hardware used for responding. The simultaneity of the tasks was done through manual synchronization.

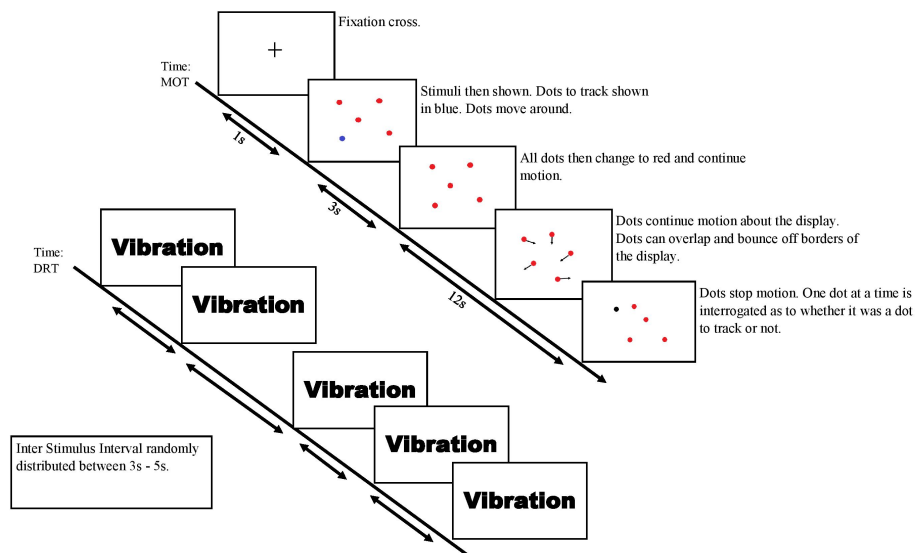


Figure 1. DRT and MOT task operating simultaneously. The upper timeline shows the MOT. Initially, a fixation cross is shown, followed by a display of 10 dots. The dots move smoothly within an aperture on the centre of the screen, with random directions. For the first 3 seconds of movement, a subset of these dots are displayed in a different color, to designate them as targets (to be tracked). After 3 seconds of movement, the colors change so that all dots are identical. Movement then continues for 12 seconds. After this, the dots stop, and the participant is required to indicate (for five of the dots, one dot at a time) whether the dot was a target or not. Constantly during the experiment, the DRT also operates. The DRT delivered a small vibration stimulus to the skin of the shoulder of the participant for 1 second, with 2–4 seconds between stimuli (i.e. an inter-stimulus interval of 3–5 seconds). The participant was required to respond to the stimulus by pressing a button attached to their finger.

The MOT required participants to track multiple moving objects within a circular display of 150 pixels for a short period of time (15s). The target dots were initially blue, and all other dots were red. After 3 seconds of movement, the target dots also turned red. All dots moved within the display area at a frame rate of 15 frames/second for the duration of a trial, and participants were required to track the movements of the indicated target dots. After 15 seconds of movement, the dots stopped and 5 dots were highlighted at random (one at a time). For each of those five dots, the participants were required to indicate whether each had been a tracked target, using the keyboard (either the “P” and “O” keys, or the “Q” and “W” keys, depending on handedness). This was manipulated within-subjects. Participants were given feedback following the completion of the test phase, for example, “Good work! For this trial you identified 3 out of 5 dots correctly”.

There were three levels of workload in the MOT; 1, 2, or 4 dots to track. This was manipulated within-subjects. The dots were circular with a diameter of 14 pixels, corresponding to a visual angle of approximately 6° . The dots’ motion did not follow any uniform direction, and could change direction by up to 15° during each frame. The dots could spatially overlap. If a dots’ motion was about to take it off the edge of the display, its motion was reflected, so that dots appeared to bounce off the sides of the display. A 15inch monitor was used with screen resolution of 1920x1080.

The tDRT implementation adhered to ISO 17488 (2016) guidelines. A finger response button was attached via a velcro strap to the index finger of each participant’s non-dominant hand. The button was used to respond to the tDRT stimuli. A tactile vibration module was fixed to the skin of each participant’s shoulder on their dominant side, using surgical tape. The tactile vibration module elicited a short vibration (lasting 1 second, or until the button was pressed, whichever came first).

Participants were instructed to press the response button on their finger as quickly

as possible each time that they felt a vibration on their shoulder. Participants were able to press the button between their index finger and thumb, or between the index finger and a hard surface, such as the desk. The time between successive vibration stimuli offset to onset was randomly distributed between 3-5 seconds. The stimulus vibration lasted for up to 1 second, but terminated upon a response. Responses entered before the onset of the next vibration stimulus were deemed “hits”, and a failure to respond within a 2.5 second interval (from stimulus onset) was deemed a “miss”. Second (and subsequent) responses entered before the onset of the next stimulus were deemed “false alarms”. Response time was measured as the time between the onset of the vibration stimulus and the pressing of the finger button.

Other implementations of the DRT have used a visual stimulus (a peripheral light) instead of a tactile one (the vibration). We chose the tactile version for two reasons. Primarily, as the MOT is a visual task, the tactile version eliminate potential confounds that could have been caused by multiple visual cues, such as gaze eccentricity. This stems from the logic that underpins multiple resource theory (Wickens, 2002, 2008). Secondly, there is evidence that the tDRT is as sensitive as other methods of DRT implementation in the measurement of cognitive workload (Merat & Jamson, 2008; Conti et al., 2012; Stojmenova et al., 2017).

Procedure. Participants were initially briefed on the experiment and the equipment, after which the DRT equipment was set up. Participants were given instructions on the tDRT, and were initially introduced to the tDRT by five practice trials. After this, a further three practice trials were given with a simultaneous practice block for the MOT. For this block, and subsequent test phases, the tDRT ran continuously, including during the tracking phase, responding phase and in-between MOT trials. At the beginning of each block in the MOT the participant was asked to indicate that they were ready to begin

the task. The experimenter then gave a countdown to indicate when to begin the MOT, ensuring that the tDRT ran simultaneously with the MOT.

Participants completed a practice block followed by nine test phase blocks. Each block consisted of ten trials of the MOT, with the exception of the practice block which only consisted of three trials, of random difficulty. Within each test block, all of the trials used the same number of dots to be tracked: either 1, 2, or 4. Each of these levels of load was used for three blocks, giving a total of 30 MOT trials for each load condition. Blocks took on average five minutes to complete. Participants were given breaks between blocks, and the total time taken to complete the experiment was between 1-1.5 hours.

Results

We treated our study as a one-way design, with the single within-subject variable of load (number of dots to track in the MOT – 1, 2 or 4). For the tDRT we assessed response time and the proportion of missed responses, and for the MOT we assessed the response time and the signal detection measure of sensitivity (d') for responses to each dot interrogation. We used one-way Bayesian ANOVAs for each of the above measures of interest. All analysis was conducted in the statistical program JASP (JASP Team, 2019; van Doorn et al., 2019), using default priors (set at 0.707). Bayes Factors (BF) are a measure of the relative likelihood of the data under the null hypothesis (H_0 ; which assumes equality across conditions) against the data under the alternative hypotheses (H_1). Bayesian inference has become a standard approach in many fields because of its advantages over frequentists methods (Wagenmakers, Lee, Lodewyckx, & Iverson, 2008). In the current design, BFs expressed in favour of the alternate hypotheses written as BF_{10} . We referred to Jeffreys (1961) for interpretation of BF_{10} .

The average proportion of correct responses in the MOT was 83.3% (SE = 6.3%) and average MOT RT was 1.04 seconds (SE = 0.13). The change in performance for the

MOT across different load levels is shown in the top left panel of Figure 2. Mean MOT RT (second panel from the left) and mean d' (top left panel) both deteriorated as load increased. Bayesian ANOVAs showed a strong preference for the model that included the main effect of load for both mean MOT RT ($BF_{10} = 6.4 \times 10^{17}$) and d' ($BF_{10} = 9.3 \times 10^{57}$). Bayesian t-tests revealed that these differences were reliable between all levels of load for both mean MOT RT (1 dot vs. 2 dot, $BF_{10} = 6.4 \times 10^6$; 2 dot vs. 4 dot, $BF_{10} = 39483$) and d' (1 dot vs. 2 dot, $BF_{10} = 3.2 \times 10^{14}$; 2 dot vs. 4 dot, $BF_{10} = 3.6 \times 10^{25}$). These patterns are consistent with the hypothesis that the changing the number of dots to be tracked changed the workload requirements of the MOT.

The average proportion of missed trials in the tDRT was 3.8% (SE = 3.3%) and the average tDRT RT was 0.56 seconds (SE = .07). The second quadrant of Figure 2 shows the change in performance for the tDRT across the different levels of MOT load. Bayesian ANOVAs showed a strong preference for the model that included the main effect of load for both mean tDRT RT ($BF_{10} = 2.1 \times 10^{13}$) and miss proportion ($BF_{10} = 63,343$), with both mean tDRT RT and miss proportion increasing with load. Bayesian t-tests revealed that these differences were present between all levels of load for mean tDRT RT (1 dot vs. 2 dot, $BF_{10} = 30,462$; 2 dot vs. 4 dot, $BF_{10} = 932$), but only reliably between the 1 and 2 dot levels of difficulty for miss proportion (1 dot vs. 2 dot, $BF_{10} = 79.07$; 2 dot vs. 4 dot, $BF_{10} = 2.38$).

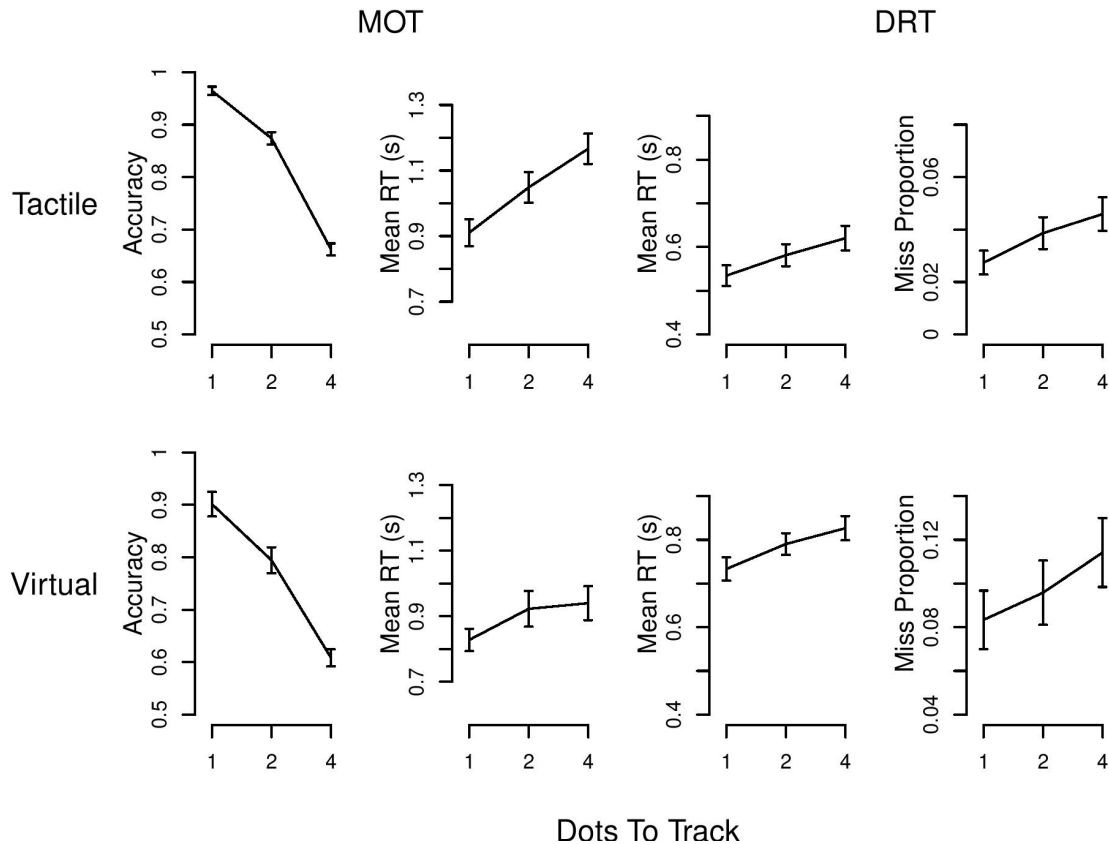


Figure 2. Experiment 1 (top row - tactile) and 2 (bottom row - online) for MOT (left) & tDRT (right). Mean accuracy (MOT) was given as the proportion of correct responses over the total amount of responses. MOT mean RTs indicate the time taken to respond to identifying whether a stimulus was a target. tDRT mean RTs indicate time taken to respond to a vibration (tactile - experiment 1) or visual stimulus (online - experiment 2) in the DRT task. DRT mean proportion of misses were characterized as stimuli elicitations which were not followed by a response from the participant. The proportion of misses was calculated as the number of misses divided by the total number of trials. Error bars show +/- 1 standard error of the mean.

Experiment 1 Discussion

Our first experiment had two key findings. Firstly, within the MOT, increasing the number of dots to track was found to increase mean MOT RT and decrease the accuracy with which dots were identified as target vs. non-target (d'). This indicates that

manipulation of the number of dots to track successfully manipulated the difficulty of the task, increasing the participants' workload. Secondly, the tDRT successfully detected these changes in workload, with an increase observed in both mean tDRT RT and miss proportion as the number of dots to track in the MOT increased. These findings are promising for the use of the DRT as an objective measure of cognitive workload in lab-based experiment cognitive psychology tasks.

Nevertheless, the DRT has some practical limitations. Firstly, the standard DRT can only be applied in lab-based, or controlled applied settings, and requires physical contact with the participant and constant experimenter supervision to operate properly. This can make data collection time-consuming and expensive, as it limits concurrent testing of multiple participants and requires the constant presence of an operator. In addition, the limitation to lab-based settings means that the DRT cannot be applied using the increasing common methods of online data collection, which limits its general appeal to scientists within modern experimental cognitive psychology. We investigated a potential solution for these issues in Experiment 2, using an "online DRT". The online DRT used a visual cue that is presented on the same screen that is used to present the other task (the MOT, in our case). A vDRT stimulus is necessary in this scenario due to the study being presented online.

The participant responds to the vDRT cue with a standard keyboard button press. In all other ways, the online vDRT functions identically to the standard DRT. The online vDRT reduces the marginal equipment costs to zero, and allows for multiple participants to be tested simultaneously, without constant experimenter supervision. Importantly, the online vDRT can also be deployed over the internet using systems such as Amazon Mechanical Turk. Experiment 2 tested the effectiveness of the online vDRT in a similar manner to Experiment 1. Note that the code to implement our online DRT is also freely available

online at <https://osf.io/dc39s/>.

Experiment 2

Method

Participants. Thirty-nine undergraduate psychology students (3 left handed) from the University of Newcastle were reimbursed with course credit. Age and gender were not recorded. Participants were required to have normal or corrected-to-normal vision and be able to read English. Unlike the first study, participants participated online, without supervision from an experimenter. Participants who made more than 100 false alarms on the DRT in any difficulty condition were excluded from analysis. This criterion resulted in the exclusion of three participants. This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Administration at the University of Newcastle (HREC-2013-0250). Informed consent was obtained from each participant.

Procedure. Experiment 2 was an online version of the DRT-MOT study in which the vDRT was integrated into the same display and modality as the MOT. The MOT followed the same procedure as Experiment 1. The vDRT used a visual rather than a tactile stimulus – a red square that was shown above the display area of the MOT. The stimulus was displayed for a longer duration than in Experiment 1, a period of 5 seconds, or until a response was recorded (whichever came first). This longer onset time was the result of technical difficulties and may explain some of the variance. The stimuli had the same frequency as the tDRT in Experiment 1, with each new stimulus elicitation randomly distributed between 3-5 seconds. Participants were instructed to respond by pressing the "T" or "Y" keys (dependent on handedness; "T" for right, "Y" for left) as fast as possible upon detecting the red square. This alleviated the need for DRT hardware. Other than

these changes, the vDRT and MOT ran as per Experiment 1. The design, number of dots to be tracked, number of trials and blocks, were all identical as Experiment 1. An example of the vDRT stimulus during an MOT trial is shown in Figure 3.

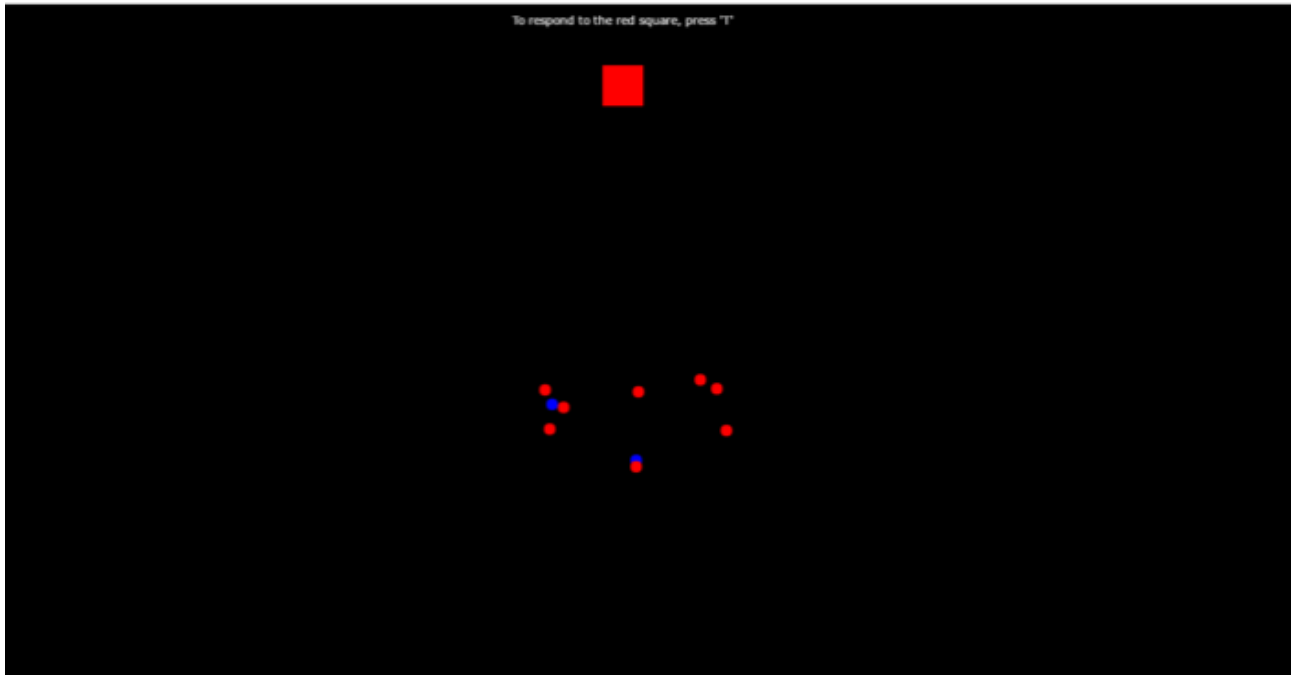


Figure 3. Screen shot of task with vDRT stimuli shown as the red square above the MOT task display area.

Results

Again, the MOT was performed well: average accuracy was 77% (SE = 8.6%), and mean MOT RT was 0.89 seconds (SE = 0.55). The third quadrant of Figure 2 shows the change in performance for the MOT across the different levels of MOT load for experiment 2. A Bayesian ANOVA showed a strong preference for the model that included the main effect of load for d' ($BF_{10} = 1.06 \times 10^{26}$), with d' decreasing as load increased. Bayesian t-tests revealed that these differences were reliable between all levels (1 dot vs. 2 dot,

$BF_{10} = 4.3 \times 10^6$; 2 dot vs. 4 dot, $BF_{10} = 1.0 \times 10^{12}$). However, although mean MOT RT showed a trend consistent with both predictions and the results of Experiment 1, increasing when MOT load increased, a Bayesian ANOVA showed ambiguity for the inclusion of the effect of load ($BF_{01} = 0.6$), making it uncertain whether load reliably influenced mean MOT RT.

The average proportion of misses in the vDRT was 6.1% (SE = 4.9%), and the average vDRT RT was 0.64 seconds (SE = 0.22). The fourth quadrant of Figure 2 shows the change in performance for the vDRT across the different levels of MOT load. A Bayesian ANOVA showed a preference for the model that included the main effect of load for mean vDRT RT ($BF_{10} = 32.38$), with mean vDRT RT increasing as load increased, and Bayesian t-tests revealed that these differences were reliable between the 1 dot and 2 dot conditions ($BF_{10} = 4.36$) and 1 and 4 dot conditions ($BF_{10} = 10.59$), but that the evidence was ambiguous as to whether the effect was present between the 2 dot and 4 dot conditions ($BF_{10} = 0.8$). Miss proportion showed a trend consistent with both predictions and the results of Experiment 1, increasing when MOT difficulty increased, with a Bayesian ANOVA showing evidence for the inclusion of the effect of load ($BF_{10} = 33.21$). These differences were only reliable between the 1 dot and 4 dot conditions ($BF_{10} = 38.68$), with ambiguous evidence as to whether the effect was present between the other conditions (1 dot vs. 2 dot, $BF_{10} = 0.7$; 2 dot vs. 4 dot, $BF_{10} = 1.9$).

Experiment 2 Discussion

As in Experiment 1, increasing the number of dots to track in the MOT task resulted in decreased sensitivity (d') to correctly identifying tracked dots. Although mean MOT RT increased as the number of dots to track increased as predicted, the statistical reliability of this effect was ambiguous. Secondly, the vDRT successfully detected these changes in workload within mean vDRT RT, with increases seen in this measure as the number of

dots to track in the MOT increased. Proportion of misses also showed the expected trend, with more misses as the number of dots to track increased.

Experiment 2 tested whether a minimal-cost, equipment-free version of the vDRT could detect the differences in cognitive workload induced by the the MOT, similar to Experiment 1, to create a baseline to validate an online version of this paradigm. Our online vDRT was more vulnerable to potential perceptual issues arising from using the same modality as the MOT (visual stimuli for both tasks). Furthermore, the online version of the experiment meant that participants completed the experiment on a computer in their own time and unsupervised. This means that the environment is likely less controlled than in a lab setting, with the experimenter watching over them ,and completed on different sized screens. Despite this limitation however, results remained consistent with Experiment 1 findings. This finding is in line with research showing that online psychology studies are effective in these environments (Zwaan et al., 2017), and online studies have the key benefits of allowing efficient collection of larger participant pools.

The online vDRT was only slightly less effective in detecting changes in workload. vDRT RTs and misses were reliably influenced by load, as in Experiment 1, but the differences between some conditions were not reliable. These findings suggest that the online vDRT is effective in detecting changes in cognitive workload. It is difficult however to compare Experiment 1 and Experiment 2 due to the elongated stimulus onset time. As a result of this, both vDRT RT and miss proportion is likely effected, as participants are given more exposure to the stimuli, which is noted by van Winsum (2019). Similarly, mean vDRT RT is likely effected, as the elongated exposure could lead to a decreased importance of responding.

General Discussion

Our study evaluated whether the Detection Response Task (DRT), a dual-task method of measuring cognitive workload primarily applied to driving tasks, would be effective at detecting workload changes in a laboratory setting with a previously unexplored, continuous secondary task. The current research aimed to extend upon similar in-lab research (Young et al., 2013; Hsieh, Seaman, & Young, 2015; Schindhelm & Schmidt, 2015; Thorpe et al., 2019) and establish evidence for the validity of online applications. This has the potential to substantially broaden the scope of cognitive workload evaluation with the DRT, beyond small-sample, high-cost, in-lab or in-simulator experiments. Our results showed that the method can detect cognitive workload changes in a standard cognitive paradigm, highlighting the utility of the DRT as a workload assessment tool which can be applied in a variety of contexts. We manipulated the workload of the participants by changing the number of dots to be tracked in a multiple object tracking (MOT) task. The continuous MOT task was used as the primary task, similar to those used in driving studies, in order to reflect more realistic, fluid environments, compared to the previously researched discrete responding tasks, such as the n -back and surrogate reference task (Hsieh et al., 2015; Stojmenova et al., 2017). Our first experiment revealed that both tDRT (tactile DRT) response times and the proportion of missed responses to the tDRT stimulus were sensitive to changes in workload. Our second experiment investigated an online version of the DRT, using a visual stimulus, which showed that vDRT (virtual, online, DRT) response times and proportion of missed trials increased as the number of dots to track increased. This indicated that the online vDRT was a reliable indicator of changes in cognitive workload, even when applied online and with a DRT stimulus that is the same modality as the other task. The vDRT may be less sensitive to cognitive workload changes as differences between conditions were not as reliable as in Experiment 1, potentially due

to an elongated stimulus onset or due to the small stimulus presentation. However, similar studies regarding DRT stimulus modality have shown similar trends (for more examples see Merat & Jamson, 2008; Stojmenova et al., 2017; Stojmenova & Sodnik, 2018). Our second experiment is of critical importance to the current article as it presents a version of the dual-task which is able to be more broadly distributed with consistent results despite the unsupervised environment.

The implications of our findings are promising for future objective measurement of cognitive workload. The DRT has been predominantly used in driving scenarios, and so demonstrating the sensitivity and reliability of the method outside of an applied scenario may not seem an intuitive step. However, these new results, suggest alternate applications of the DRT using online capability, increasing research reach. Cognitive workload evaluation is important in diverse fields including learning and education (De Jong, 2010; Brünken et al., 2004), personnel selection and training (Innes et al., 2018), gaming (Haapalainen et al., 2010), and large scale interface/display testing (Thorpe et al., 2019). De Jong (2010) clearly explain the difficulty in cognitive load evaluation in educational research. In studies of learning, changes in cognitive load are inferred in a post-hoc manner: if performance increases, then the added load from some extra material is inferred to be useful or necessary, but if performance decreases, then the added load must have been detrimental. Further, De Jong (2010) notes the lack of a concurrent measure of workload which can also measure overload. Using the DRT, one can clearly separate workload and performance factors so that both are measured, and can be measured during the learning process rather than after. Although this does not directly solve the problem posed by De Jong, it does allow insight into the *magnitude* of both performance and workload effects – which includes cognitive overload. For example, in our experiments’ most difficult MOT condition (four dots to track), performance declines dramatically and measured workload increased. In the

easy (one dot) MOT condition, the decline in performance (from the zero dot condition) is minimal – as are workload effects. In an educational setting, this methodology could be applied to titrate precisely which levels of workload lead to performance decline.

Brünken et al. (2004), Brünken et al. (2002) and DeLeeuw and Mayer (2008), provide further examples of the use of such a measure of workload. Those studies used a dual-task methodology like the DRT to observe the effects of multi-modal stimuli on learning and workload. Results from Brünken et al. (2002) showed increases in cognitive load, as predicted by cognitive load theory and in line with our own experimental results. The researchers used a visual change detection task to assess load when participants were presented with audiovisual learning stimuli. Similar results were observed in an auditory stimulus-response paradigm ((Brünken et al., 2004)). DeLeeuw and Mayer (2008) compared results from a dual-task response time measure to subjective workload measures in a study focusing on multimedia learning, finding that the dual-task method was most sensitive to manipulations of extraneous factors. Results from these studies not only show the importance of workload evaluation across contexts, but also show that, similar to the present study, stimulus-response tasks are useful dual-task methods in evaluating workload factors. Evidently, results from the current study are promising for future applications of the DRT in less demanding environments, or for broader distribution – via the online capability.

It should be noted that our findings do not necessarily mean that the DRT will be sensitive to measuring load in all types of tasks. Importantly, the MOT had one key element in common with driving tasks that may have influenced the ability to detect the changes in workload: the task continuity. In both the MOT, and driving research, the task that participants must complete is continuous, with no constant breaks that often occur in standard lab tasks between trials. This means that for continuous tasks the

impact the task has on cognitive capacity is constantly occurring, whereas this is not necessarily the case in tasks with many discrete trials. In these discrete responding tasks the temporal overlap between tasks can potentially compromise the nature of responding in both tasks. Therefore, when selecting a task to pair with the DRT to measure cognitive workload, it is important to attempt to use a task that has some level of continuity, and therefore constantly impact cognitive capacity. Although not reported here, we confirmed this intuition in a series of pilot experiments. Those experiments failed to find any reliable effect on DRT data when cognitive load was manipulated by a random dot kinematogram, pursuit tracking task and a Compensatory Tracking Task. These tasks were found to be relatively incompatible with our design, as the random dot kinematogram required discrete responses, meaning there were minimal periods where participants had to respond to both tasks simultaneously, and the Compensatory Tracking Task and pursuit tracking task showing an insignificant effect on cognitive load, possibly due to the ease of the task. The MOT task we used appears to capture workload changes, however, it should be noted that periods of discrete responding could cause changes in MOT response times and accuracy. Future studies should only elicit the DRT during the tracking phase of the task to avoid potential attentional conflicts.

Future research may more directly compare the standard and online DRT, either within participants or using random assignment across procedures, and may also investigate the workload cost of the DRT procedure itself, by comparison with a null condition (i.e. corresponding to just doing the MOT in our experiment). Further research is also needed in other human factors domains, such as learning and educational research, as De Jong (2010) notes that dual-task designs are under researched, yet could provide meaningful evidence in this area. As outlined above, the methodology we present provides a starting point for this kind of evaluation, highlighting the strengths and applicability of such dual-task

designs. Finally, more research is needed to disentangle the effects of the online vs. in-lab administration with the effects of visual vs. tactile stimulus presentation, which were confounded between our Experiments 1 and 2. Nevertheless, these experiments provide promising results for the use of the DRT, and DRT-like paradigms. These results provide evidence for a simple, and effective, procedure for accurately and objectively measuring cognitive workload in psychological research.

Key Points

- The detection response task is sensitive to workload changes induced by different levels of difficulty in a multiple object tracking task
- The current design shows evidence for the validity of a more accessible version of the detection response task, outside of applied settings.
- An online version of the detection response task - multiple object tracking task showed reliability in trends, despite compromised response times.
- The detection response task is useful for cognitive workload evaluation in less demanding environments.

References

- Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (pvt) to sleep loss. *Sleep*, *34*(5), 581–591.
- Basner, M., & Dinges, D. F. (2012). An adaptive-duration version of the pvt accurately tracks changes in psychomotor vigilance induced by sleep restriction. *Sleep*, *35*(2), 193–202.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... Craven, P. L. (2007). Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, *78*(5), B231–B244.
- Broadbent, D. E. (2013). *Perception and communication*. Elsevier.
- Brock, D., Stroup, J. L., & Ballas, J. A. (2002). Effects of 3d auditory display on dual task performance in a simulated multiscreen watchstation environment. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 46, pp. 1570–1573).
- 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.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, *32*(1-2), 115–132.
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental psychology*, *49*(2), 109.
- Bruyas, M.-P., & Dumont, L. (2013). Sensitivity of detection response task (drt) to the driving demand and task difficulty.
- Castro, S. C., Cooper, J., & Strayer, D. (2016). Validating two assessment strategies for visual and cognitive load in a simulated driving task. In *Proceedings of the human factors and*

- ergonomics society annual meeting* (Vol. 60, pp. 1899–1903).
- Castro, S. C., Strayer, D. L., Matzke, D., & Heathcote, A. (2019). Cognitive workload measurement and modeling under divided attention. *Journal of experimental psychology: human perception and performance*, *45*(6), 826.
- Coleman, J. R., Turrill, J., Cooper, J. M., & Strayer, D. L. (2016). Cognitive workload using interactive voice messaging systems. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 60, pp. 1894–1898).
- Conti, A., Dlugosch, C., Vilimek, R., Keinath, A., & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. *Advances in human aspects of road and rail transport*, 735–743.
- Cooper, J. M., Castro, S. C., & Strayer, D. L. (2016). Extending the detection response task to simultaneously measure cognitive and visual task demands. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 60, pp. 1962–1966).
- De Croon, E., Sluiter, J., Kuijer, P. P., & Frings-Dresen, M. (2005). The effect of office concepts on worker health and performance: a systematic review of the literature. *Ergonomics*, *48*(2), 119–134.
- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science*, *38*(2), 105–134.
- DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of educational psychology*, *100*(1), 223.
- Diels, C. (2011). Tactile detection task as a real time cognitive workload measure. In *Contemporary ergonomics and human factors 2011: Proceedings of the international conference on ergonomics & human factors 2011, stoke rochford, lincolnshire, 12-14 april 2011* (p. 183).
- Dinges, D., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., ... Pack, A. I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, *20*(4), 267–277.
- Dinges, D., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual rt task during sustained operations. *Behavior research methods, instruments, &*

- computers*, 17(6), 652–655.
- Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1), 32–64.
- Engström, J., Larsson, P., & Larsson, C. (2013). Comparison of static and driving simulator venues for the tactile detection response task. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 369–375).
- Gladstones, W. H., Regan, M. A., & Lee, R. B. (1989). Division of attention: The single-channel hypothesis revisited. *The Quarterly Journal of Experimental Psychology*, 41(1), 1–17.
- Gross, B., Bretschneider-Hagemes, M., Stefan, A., & Rissler, J. (2018). Monitors vs. smart glasses: A study on cognitive workload of digital information systems on forklift trucks. In *International conference on digital human modeling and applications in health, safety, ergonomics and risk management* (pp. 569–578).
- Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th acm international conference on ubiquitous computing* (pp. 301–310).
- Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.
- Howard, Z. L., Evans, N. J., Innes, R., Brown, S., & Eidels, A. (2019). Is multitasking just a form of difficulty? *PsyArXiv*.
- Hsieh, L., Seaman, S., & Young, R. (2015). *A surrogate test for cognitive demand: tactile detection response task (tdrt)* (Tech. Rep.). SAE Technical Paper.
- Innes, R. J., Howard, Z. L., Eidels, A., & Brown, S. D. (2018). Biting off more than you can process: Group differences in cognitive workload. In *Proceedings of the 45th annual conference of the australasian society for experimental psychology* (p. 79).
- Innes, R. J., Howard, Z. L., Thorpe, A., Eidels, A., & Brown, S. (2019). The effects of increased

- visual information on cognitive workload in a helicopter simulator.
- ISO:17488. (2016). *Road vehicles—transport information and control systems—detection-response task (drt) for assessing attentional effects of cognitive load in driving*. International Organization for Standardization Geneva, Switzerland.
- JASP Team. (2019). *JASP (Version 0.11.0)[Computer software]*. Retrieved from <https://jasp-stats.org/>
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- Kahneman, D. (1973). *Attention and effort*. Citeseer.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist, 38*(1), 43–52.
- Merat, N., & Jamson, A. H. (2008). The effect of stimulus modality on signal detection: Implications for assessing the safety of in-vehicle technology. *Human factors, 50*(1), 145–158.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81.
- Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin, 116*(2), 220.
- Patten, C. J., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident analysis & prevention, 36*(3), 341–350.
- 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., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic bulletin & review, 21*(3), 577–589.
- Roach, G. D., Dawson, D., & Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiology international, 23*(6), 1379–1387.
- Schindhelm, R., & Schmidt, E. (2015). Evaluation of the tactile detection response task in a laboratory test using a surrogate driving set-up. *IET intelligent transport systems, 9*(7), 683–689.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual

- objecthood. *Cognitive psychology*, 38(2), 259–290.
- Stevens, R. H., Galloway, T., & Berka, C. (2007). Eeg-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In *International conference on user modeling* (pp. 187–196).
- Stojmenova, K., Jakus, G., & Sodnik, J. (2017). Sensitivity evaluation of the visual, tactile, and auditory detection response task method while driving. *Traffic injury prevention*, 18(4), 431–436.
- Stojmenova, K., & Sodnik, J. (2018). Detection-response task uses and limitations. *Sensors*, 18(2), 594.
- Strayer, D. L. (2015). Is the technology in your car driving you to distraction? *Policy insights from the behavioral and brain sciences*, 2(1), 157–165.
- Strayer, D. L., Cooper, J. M., McCarty, M. M., Getty, D. J., Wheatley, C. L., Motzkus, C. J., . . . Horrey, W. J. (2019). Visual and cognitive demands of carplay, android auto, and five native infotainment systems. *Human Factors*, 0018720819836575.
- 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. (2016). Talking to your car can drive you to distraction. *Cognitive research: principles and implications*, 1(1), 16.
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2017). The smartphone and the drivers cognitive workload: A comparison of apple, google, and microsofts intelligent personal assistants. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(2), 93.
- Strayer, D. L., & Drew, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human factors*, 46(4), 640–649.
- Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human factors: The journal of the human factors and ergonomics society*, 48(2), 381–391.
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, 12(6), 462–466.

- Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(8), 1300–1324.
- 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., Strayer, D., Eidels, A., & Heathcote, A. (2017). Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception, & Psychophysics*, 1–9.
- 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.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.
- Tsai, Y.-F., Viirre, 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.
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., . . . Wagenmakers, E.-J. (2019). The jasp guidelines for conducting and reporting a bayesian analysis. *PsyArXiv*.
- van Winsum, W. (2019). A threshold model for stimulus detection in the peripheral detection task. *Transportation research part F: traffic psychology and behaviour*, *65*, 485–502.
- Van Winsum, W., Herland, L., & Martens, M. (1999). *The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance*. TNO Human Factors Research Institute.
- Wagenmakers, E.-J., Lee, M. D., Lodewyckx, T., & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York: Springer Verlag.
- 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.

- Xie, F., Wang, Q., Jin, X., Liao, Y., Zheng, S., Li, L., ... Liu, Z. (2016). Evaluation of the crew workload to quantify typical mission profile special vehicles. In *International conference on man-machine-environment system engineering* (pp. 149–158).
- Young, R. A., Hsieh, L., & Seaman, S. (2013). The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 71–77).
- Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2017). Participant nonnaivete and the reproducibility of cognitive psychology. *Psychonomic bulletin & review*, 1–5.

Reilly J. Innes, University of Newcastle, received his B Psych (hons) from the University of Newcastle in 2016.

Nathan J. Evans, University of Amsterdam, received his PhD in cognitive psychology from the University of Newcastle, Australia in 2017.

Zachary L. Howard, University of Newcastle, received his BS in Psychology from the University of Newcastle in 2014.

Ami Eidels, University of Newcastle, received his PhD in cognitive psychology from Tel Aviv University, Israel in 2006.

Scott D. Brown, University of Newcastle, received his PhD in cognitive psychology from the University of Newcastle, Australia in 2002.