International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

# Mouse Movement Trajectories as an Indicator of Cognitive Workload

Alexander Thorpe, Jason Friedman, Sylvia Evans, Keith Nesbitt & Ami Eidels

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2021.2002054



# Mouse Movement Trajectories as an Indicator of Cognitive Workload

Alexander Thorpe [a], Jason Friedman [b], Sylvia Evans [a], Keith Nesbitt [c], and Ami Eidels [a]

[a]School of Psychological Sciences, University of Newcastle, Australia; [b]Department of Physical Therapy, Tel Aviv University, Tel Aviv, Israel; [c]School of Information and Physical Sciences, University of Newcastle, Australia

**ABSTRACT**

Assessing the cognitive impact of user interfaces is a shared focus of human-computer interaction researchers and cognitive scientists. Methods of cognitive assessment based on data derived from the system itself, rather than external apparatus, have the potential to be applied in a range of scenarios. The current study applied methods of analyzing kinematics to mouse movements in a computer-based task, alongside the detection response task, a standard workload measure. Sixty-five participants completed a task in which stationary stimuli were tar;geted using a mouse, with a within-subjects factor of task workload based on the number of targets to be hovered over with the mouse (one/two), and a between-subjects factor based on whether both targets (exhaustive) or just one target (minimum-time) needed to be hovered over to complete a trial when two targets were presented. Mouse movement onset times were slower and mouse movement trajectories exhibited more submovements when two targets were presented, than when one target was presented. Responses to the detection response task were also slower in this condition, indicating higher cognitive workload. However, these differences were only found for participants in the exhaustive condition, suggesting those in the minimum-time condition were not affected by the presence of the second target. Mouse movement trajectory results agreed with other measures of workload and task performance. Our findings suggest this analysis can be applied to workload assessments in real-world scenarios.

## 1. Introduction

We live in a time when a driver can speak to their car, and the car can speak back. New technology has been developed to deliver information in new ways in everyday life. But does this information make life easier, or simply bombard us with stimuli? In some forms of human-computer interaction the ability to process data and make rapid decisions may depend critically on the user's level of cognitive workload. When this criteria is important for performance of the interface it can be useful to apply previously validated cognitive measures to these scenarios. This allows researchers to assess users' cognitive states using robust measures, and can enable user interface designers to evaluate and improve the usability of their designs. In this paper we combine a traditional dual-task measure with analysis of the user mouse movements to provide an enhanced interpretation of cognitive workload in a simple targeting task. This measurement is demonstrated without the need for additional apparatus, lending it ecological validity for human-computer interaction applications.

### *1.1. Background*

User interfaces (UIs) appear in many forms, from computer displays to in-vehicle assistants in cars and heads-up displays in aircraft. UIs are designed to present information that users need in order to complete tasks, but the amount of information a UI presents, and the manner in which it is presented, can affect users' experience negatively (Haapalainen et al., 2010). For example, in-car voice-based assistants are designed to make the driving experience easier by providing voice control for various systems in the vehicle, but their use can distract drivers, leading to driver errors (Strayer et al., 2017). The paradox of more information hindering users stems from the limited attentional resources users have at their disposal (Bach et al., 2009). The effect of UIs on user's cognitive state is therefore of particular interest in the discipline of human-computer interaction.

Excessive or distracting information affects users because of the limited *cognitive capacity* people have to process stimuli (Kahneman, 1973; Townsend & Eidels, 2011). Tasks that require some resource such as *working memory* can overload this capacity, leading to deterioration in performance (Causse, Peysakhovich et al., 2016). The *cognitive workload* of a task or stimulus refers to the effect it has on the user's cognitive capacity. As workload increases, the resources available to the user diminish, until they are overloaded, and performance suffers. However, workload and performance are not correlated one-to-one. Some sources of workload may have more of an effect than others. For example, according to Multiple Resource Theory (Wickens, 1980), two tasks in a multitasking setting that appeal to the user's visual sense are theorized to cause greater interference than tasks that do not share a sensory modality, as processing stimuli of different modality may utilize different cognitive resources. However, multitasking is likely to impose greater cognitive workload than a single

**CONTACT** Alexander Thorpe alexander.thorpe@uon.edu.au University of Newcastle, University Dr, Callaghan NSW 2308, Australia

task, as the planning and execution of responses draws on a single pool of resources (Wickens, 2002).

The effects of increased sensory information can also be understood using Lavie's (1995) two-stage model of processing, which distinguishes between perceptual demands and cognitive demands. Tasks with high *perceptual load* affect users by presenting high amounts of stimuli to be processed (Lavie, 1995). Excessive perceptual load can lead to *inattentional blindness*, a state in which users fail to perceive new, potentially important stimuli (Macdonald & Lavie, 2008; Simons & Chabris, 1999). By contrast, environments that require users to multitask, or share cognitive resources between concurrent tasks, impose high *cognitive control load* (Lavie et al., 2004), which can lead to decreased performance in each of the tasks, and even to a loss of situational awareness (Endsley, 1993). The distinction between these forms of load is subtle, but meaningful. Perceptual load is increased merely by the presence of stimuli, whereas cognitive control load is increased by demands on the user to process and interact with stimuli. Increased cognitive workload is also associated with higher levels of stress (Or & Duffy, 2007). Completing a cognitive task at a high level of stress carries with it a higher physiological cost than completing the same task while calm (Mandrick et al., 2016), leading to an overall negative physiological impact including fatigue (Patel et al., 2016). Stress and cognitive workload therefore have a complex effect on performance in the context of computer-based tasks. However, the current study focuses on the specific effect of cognitive workload on computer-based task performance.

Researchers have assessed the effect of additional perceptual load on users' cognitive state when that additional information is not relevant to the task. This kind of experimental design, known as a distractor task, assesses the negative effect of the irrelevant stimuli on task performance (Forster & Lavie, 2007). In contrast, a task where the additional information is relevant is known as a *redundant target task*; additional information is expected to improve performance in the form of faster responses, facilitation that is termed the redundant target effect (Townsend & Nozawa, 1995).

Another method of assessing the cognitive workload of a computer-based system is by presenting a secondary measure. This can take the form of a physiological measure, such as heart rate tracking (Rajan et al., 2016; Rottger et al., 2009; Ryu & Myung, 2005), or eye tracking (Causse, Imbert et al., 2016; Kim & Wohn, 2011; Kujala & Saariluoma, 2011). These measures are useful as they serve as indicators of physiological arousal, which in turn reflects the user's stress and attention capacity (Bach et al., 2009). A drawback of these measures is the need for measurement devices, which can be expensive and obtrusive. An alternative secondary measure is a dual task experimental paradigm. Such designs present a secondary task whose demands do not change, alongside a primary task with differing levels of cognitive workload. Any changes in performance on the secondary task are inferred to be the result of changing demands in the primary task. This design has been used to assess the cognitive impact of interacting with synthetic talking head systems (Stevens et al., 2013) and the effects of multitasking on driving ability (Salvucci & Beltowska, 2008). The latter application has also given rise to the development of a standardized dual task measure, the detection response task.

### 1.2. Detection response task

The detection response task (DRT) is a commonly used dual task measure. It is a standardized task designed to assess cognitive workload in real-world settings (International Organization for Standardization, 2016). It requires participants to respond whenever they detect the presence of a predesignated stimulus such as a tone, or a light. It is presented alongside a continuous primary task, and indexes *residual capacity*, or those resources not used by the primary task. Because the DRT's demands do not change, it can be inferred that any changes in DRT performance, either in the form of response speed or accuracy, are due to changes in the workload demands of the primary task. Slow, inaccurate responses on the DRT indicate low residual capacity, and therefore a state of high workload, whereas fast, accurate DRT responses indicate high residual capacity, and low workload.

The DRT has been used to estimate the workload demand of mobile phone use while driving (Strayer et al., 2015; Tillman et al., 2017), the effect of voice-based assistants in driving (Strayer et al., 2017), and the relative workload imposition of 2D and 3D information presented in the heads-up displays of helicopter pilots (Innes et al., 2018). The DRT has also been applied in the domain of human-computer interaction, to assess the cognitive impact of an ambient display (Shelton et al., 2020). Because the DRT is a standard measure, it can be delivered in a computer-based task, provided the task follows the methodology laid out in the ISO standard (Thorpe et al., 2019). The overall workload demand of the DRT is low, so it has a small, but non-zero impact on primary task performance (Thorpe et al., 2020).

The DRT is well suited to measure cognitive workload in real-world tasks, such as driving. Since it is a secondary measure, accompanying performance in some other, main task, workload can be monitored via the DRT while the human operator can be engaged with the task of interest. This is an advantage, but also a challenge. Real life tasks offer little experimental control, and workload can vary inadvertently without the control of the experimenter. By contrast, traditional experimental designs offer improved experimental control, where the researcher can control and manipulate the workload. The two differ on yet another important aspect: while real-life tasks are often continuous (e.g., drivers must constantly monitor the road) most traditional psychological experiments, such as redundant target tasks, are not continuous. Instead, they have trial-by-trial designs, where trials begin with the presentation of stimuli, and end with the participant's response to that stimuli. The resulting data is discrete, commonly in the form of response times (RT). These tasks are well-controlled, but do not reflect the continuous nature of real-world tasks.

Trial-by-trial designs are also not always compatible with the DRT, as the latter should be presented alongside tasks with constant workload. Consider a hypothetical experimental trial, depicted in Figure 1, in which participants need to identify whether a string of letters is an English word or

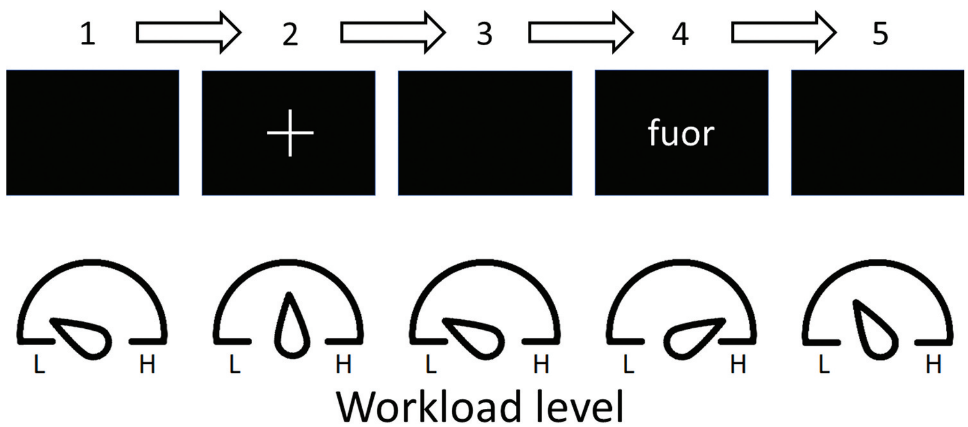


**Figure 1.** Screens from an example trial of a generic experiment, with corresponding gauges illustrating how workload levels change over time. Different stages of the trial have different workload demands.

a non-word. The gauges below each screen in the figure represent the fluctuation in participants' workload over the stages of the trial. These gauges do not represent any empirical data, and they are intended to be taken only as illustrative of the way workload changes over time. At various points in the trial, participants are simply presented a blank screen, which imparts no workload in itself. Screens 2 and 4 impart some workload as participants must perceive the stimuli presented, though screen 4 imparts the highest workload as participants must also process the string "fuor" and decide whether it is a word or a non-word. As workload does not remain consistent throughout this trial, a DRT probe presented at different times throughout the trial would detect different levels of workload. The DRT therefore cannot accurately estimate residual capacity on this task.

Nevertheless, the control gained in trial-by-trial experiments has allowed for the development of sophisticated modeling techniques that can assess the effects of workload on participants' cognitive state. These techniques generally analyze RT distributions to draw their inferences. To harness the power of these analytic techniques, it is therefore valuable to have a primary task that can produce discrete data, but is otherwise more ecologically valid.

### 1.3. Arm reaching trajectories

In arm reaching experiments, participants do not simply select a response with a button press, but rather reach to a target or response point with their hand. These experiments therefore produce richer data on each trial than traditional trial-by-trial experiments – the participant's response is not a single data point, but a time series of the position of their hand in space. This allows performance to be assessed throughout a trial, and changes in performance within a single trial to be detected. Participants monitor their own performance and can update the execution of arm movements, either consciously or unconsciously, during a movement (Friedman et al., 2013). Such experiments therefore represent a means of presenting choice-based experiments in a way more similar to real-world behaviors than traditional trial-by-trial experiments, making them more amenable to use with the DRT.

When applied to the study of cognitive processes, arm reaching experiments can test predictions in ways that may not be possible using RT data. Finkbeiner and Heathcote (2016) presented response boards on the left and right of participants, with each board corresponding to a response to one of two color-coded targets. These targets could be presented on the same side of a display as its corresponding response board (congruent), the opposite side (incongruent), or in the center (neutral). Data was collected based on how participants reached for the response boards, and the latency of the first movement toward a response board, also labeled movement *onsets*. By looking at the directions of movements made at different movement onset times, it was possible to support one explanation of the choice conflict effect being studied, and reject another explanation, in a way that RT data could not.

Arm reaching studies have been used to study the effect of distractors on perceptual processing (van Zoest & Kerzel, 2015), similarly to previous studies that collected RT data (Lavie, 1995; Lavie et al., 2009; Macdonald & Lavie, 2008). The presence of distractors was found to affect the trajectory of arm movements by literally attracting the participants' hands toward the distractor's position (see also Tillman et al. (2017)). Arm reaching data can also be used in similar applications to RT data, such as in cognitive modeling (Friedman et al., 2013). Arm reaching data have also been used to study the effect of task difficulty (MacKenzie, 2018), effort (Balasubramanian et al., 2015), and distraction and cognitive workload (Dounskaia & Goble, 2011) on trajectories. Arm reaching data can also be used to investigate how responses are planned and executed. The execution of a response need not begin only after the planning stage has finished, but rather execution can begin based on interim perception and planning, while further perception and planning continues (Cisek & Kalaska, 2010). Similarly, because trials do not finish until the end of the execution of a response, participants can update their responses and change their arm movement throughout a response (Friedman et al., 2013; Resulaj et al., 2009).

The same designs and analysis techniques used in arm reaching studies can be applied to scenarios more relevant to UI design, such as mouse movements on a computer screen (Freeman, 2018; Senanayake et al., 2015; Stillman et al., 2018) or finger movements on a touch screen (MacKenzie, 2018). Fitts' law is used to calculate the index of difficulty of a hand movement in terms of bits of information, based on the size of the target and distance the hand is required to travel (Fitts, 1954). The throughput of a process, in terms of bits per second, can also be calculated to assess the efficiency of the process. Although this law was originally devised based on direct body movements such as finger pointing, it can be generalized to scenarios where an object is moved using an input device (Jiang et al., 2015). Applying arm-reaching measures to a computer-based environment would enhance their applicability in the space of UI assessment, given the ubiquity of computer mice and touch screens in computer-based systems.

It is necessary to consider which features of movement trajectories best lend themselves to analysis in the context of UIs or other computer-based environments. As described above, Finkbeiner and Heathcote (2016) analyzed their data pooled by *movement onsets* to enrich their findings beyond what RTs could achieve. The curvature of movement trajectories could also be analyzed to measure the impact of distractors on performance (Bundt et al., 2018; Erb et al., 2016; Tillman et al., 2017). The discrepancy between "ideal" movement directly toward a target and the participant's actual movement trajectory could indicate the extent to which the distractor has interfered with movement execution. Movement time can also be used to assess movement execution, with faster movement times representing less distraction during movement execution; this measure has also been found to correlate with trajectory curvature as a measure of distraction (Bundt et al., 2018; Erb et al., 2016). Another possible avenue is the analysis of *submovements*. Motor control theories propose that rather than continuously controlling arm trajectories, they are instead controlled at discrete points in time (Miall et al., 1993), such that most movements are made up of the superposition of a number of component movements (submovements).

Arm movements toward a target are reassessed and optimized throughout the execution of the movement (Meyer et al., 1988), and the onset times of submovements in a targeting task have been shown to be temporally linked to electrophysiological markers in the action-monitoring system (Pereira et al., 2017). It follows that a participant under higher workload has less available capacity to dedicate to this optimization process, resulting in a less optimal movement characterized by slower responses with more submovements (Grimes & Valacich, 2015) due to an increase in execution noise (van Beers et al., 2004). Submovement data can also address the stages of a decision making process outlined above. Friedman et al. (2013) found evidence that when participants began executing a movement before completing the planning stage, the direction of that movement would be somewhere between the two options, and the extent to which it targeted one option reflected the currently accumulated evidence in favor of that option. Further submovements reflected the updating of evidence, suggesting the two stages of planning and executing a movement could occur concurrently, if not in parallel.

### 1.4. Current study and aims

The current body of literature suggests we can use movement trajectories to assess the effect of cognitive workload in a computer-based environment. Trajectory data may shed light on which components of a process affect behavior. For example, we would expect trials under higher workload to be slower than those under lower workload, but this difference could be driven by changes in the planning of movements, or in their execution. It is necessary to identify the stage or stages in which workload impacts performance, both to build an accurate theoretical account of workload demands on mouse movements, and to better inform practices that could optimize computer-based tasks and limit their cognitive demands. By utilizing the DRT alongside a mouse-based targeting task, a comprehensive picture of users' cognitive states across different levels of workload can be built, without the need for additional apparatus. Our study therefore aimed to assess workload in a computer-based task using movement trajectories alongside the DRT. We aimed to validate potential analyses of trajectories that could be applied beyond lab-based experiments. Specifically, we aimed to use movement onset time to assess the effect of workload on movement planning, and submovements to assess the effect of workload on movement execution. Submovement measurement was chosen over movement time as the latter is strongly influenced by factors including vigor and distance to target which are not necessarily related to workload (Shadmehr et al., 2019). Instead, we used the number of submovements, which may reflect the planning horizon – when participants are able to plan ahead, they will make less submovements, which may reflect lower workload and are less influenced by distance to target and vigor. For further analysis using alternative measures, see Appendix A.

It was expected that high *primary task workload*, that is the cognitive workload imposed by the targeting task, would be associated with slower and less accurate DRT responses, as well as slower responses on the targeting task. Three potential outcomes relating to mouse movement data were identified.

(1) Primary task workload affects movement execution only.
(2) Primary task workload affects movement planning only.
(3) Primary task workload affects both movement planning and execution.

If only movement execution is affected by primary task workload, we expected to see less optimal mouse trajectories with more submovements, while onset time would not be affected by primary task workload. This prediction was made due to previous studies of reaching trajectories finding that greater curvature was associated with distraction or incomplete movement planning. More submovements would therefore indicate a negatively impacted execution phase. Support for this prediction would imply that increased workload draws on attentional resources otherwise required for executing movements. However, if the planning stage of responses must be completed before execution can begin, it was expected that movement onset would be later under high load, with no change in the number of submovements. This prediction was based on the assumptions of a response model in which planning must be completed before execution – unlike our first prediction, this prediction implies all processing that requires cognitive resources will be complete before movement execution begins, leading to no effect of workload on movement execution. If both movement planning and execution were affected by primary task workload, we expected to see both onset time and submovements increase with primary task workload. Increased cognitive workload has been associated with slower information processing in previous cognitive research, and with greater trajectory curvature in arm-reaching research. Support for this prediction would therefore indicate that both stages of the response process require attentional resources that may be impacted by increased workload. Based on previous studies presented above, we hypothesized that the third outcome, in which both movement planning and execution are affected by workload, was most likely.

## 2. Materials and methods

### 2.1. Participants

Sixty-five participants from the University of Newcastle (F = 47, M = 18) completed the experiment in the Newcastle Cognition Lab. Mean age of participants was 24.55 years (*SD* = 8.20 years). Fifty-seven participants were right-handed. Participants were recruited through SONA, the online recruitment system, and remunerated with course credit. The study was approved by the University of Newcastle Human Research Ethics Committee, and participants signed an informed consent form before starting the experiment.

### 2.2. Design

The current study used a dual-task paradigm, in which a mouse-based task was presented alongside a computer-based DRT. The primary task was a targeting task, whereby stationary targets were presented and participants were required to hover the mouse over the targets as they appeared on screen. The primary task had a 2 (Workload: high, low) x 2 (Stopping rule: minimum time, exhaustive) mixed factors design. The first, primary task workload, was a within-subjects factor. primary task workload determined how many targets were presented on each trial. In the low-load condition, one stationary target was presented, while in the high-load condition two stationary targets were presented concurrently. Participants were required to hover the mouse over these targets, though they were not required to click targets to register a response.

A between-subjects factor of stopping rule was also presented, which related to how participants were required to respond to each trial. Low-load trials were identical for minimum-time and exhaustive conditions – with only one target present, participants were only required to hover the mouse over the target to complete the trial. However, high-load trials differed by stopping rule. Minimum-time trials were completed by hovering the mouse over only one target, which would complete the trial, while exhaustive trials were completed by hovering the mouse over first one target then the other. The trial would not be complete until participants hovered the mouse over both targets. Figure 2 shows the differences in required responses based on the workload and stopping rule conditions. It should be noted that, in the high-load minimum-time cell in the top-right of Figure 2, the trial could be completed by hovering the mouse over either target, and in the high-load exhaustive cell in the bottom-right,

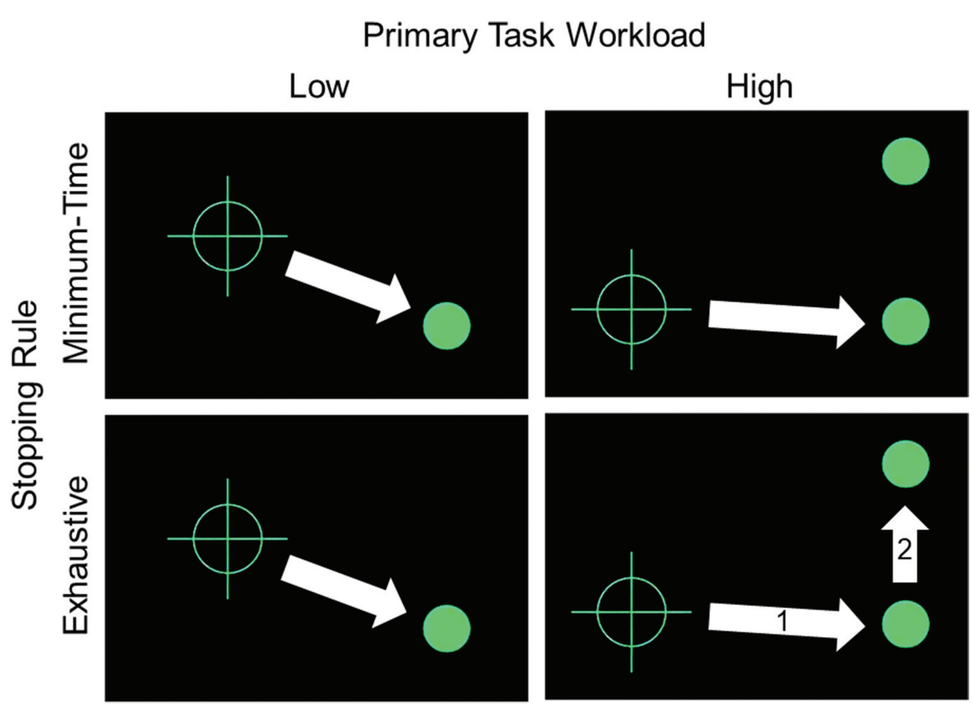


Figure 2. Expected responses to trials in each condition of primary task workload and stopping rule. Arrows show the expected movement of the mouse-controlled crosshair toward the target(s). In the minimum-time condition, high-load trials can be completed by hovering the mouse over just one target, while in the exhaustive condition, both targets must be targeted, as indicated by the numbered arrows.

either target could be chosen as the first target to hover the mouse over, though both must be targeted for the trial to complete. Changes to stopping rule across tasks were designed to manipulate the redundancy of the targets. In the minimum-time condition, processing either target was sufficient to complete the trial – participants had the opportunity to process the other target, but it was not necessary. In contrast, both targets needed to be processed in the exhaustive condition. The exhaustive condition forced participants to dedicate cognitive resources to the second target where the minimum-time condition did not. The impact of the second target on cognitive resources can therefore be examined by comparing performance across the two stopping rules, whereas this impact may not be accounted for by either stopping rule alone. Thirty-three participants completed minimum-time versions of the tasks, while thirty-two completed exhaustive versions.

Alongside this task, a secondary computer-based DRT was presented throughout the experiment. Participants were presented a visual signal and were required to respond as quickly as possible by pressing the space bar on a keyboard. Primary task dependent variables were response time, or the time taken to complete the trial, movement onset time, defined as the time from the beginning of a trial to the first mouse movement toward the target, and scaled submovements, defined as the number of submovements in each mouse trajectory scaled by the length of the trial. Further details of how these dependent variables were calculated is presented below in the Data Analysis section. Three dependent variables were collected from the DRT – mean RT, hit rate and false alarm rate.

## 2.3. Stimuli

The primary task was presented in the center of the screen, in a square area subtending 13 × 13 degrees of viewing angle. Low-load trials presented one target dot, while high-load trials presented two targets dots. All the dots had a radius of 0.5 degrees. In the primary targeting task, participants controlled a crosshair object with a radius of 0.75 degrees. Participants responded to targets by moving this crosshair over the stimulus with a mouse using their right hand. The precision of mouse movements required to complete a trial was optimized during pilot testing to ensure rapid, random movements would not be registered as completed trials. Upon completing a targeting task trial, the next trial began immediately. Trials did not time-out, but rather continued until participants completed the trial. This was done so as to require continuous task performance, without inter-stimulus intervals. The placement of targets on low-load trials was random, as was one target in high-load trials, with each co-ordinate drawn from independent uniform distributions. On high-load trials, the second target was placed such that it would be no more than 1.5 degrees of viewing angle closer or further from the mouse cursor to the first target. This was done so targets would be at least somewhat equivalent in their distance from the cursor at the beginning of each trial, without restricting potential target positions to the extent that valid positions could not be found.

Wall and floor objects surrounded the central area, and lines on these objects moved to simulate forward motion through a trench. This was done to present a continuous visual environment, similar to what would be seen in a video game. DRT stimuli were presented by illuminating the walls of the trench as shown in the right panel of Figure 3. DRT stimuli were presented for one second, or until participants responded to the stimuli. Responses slower than 2.5 seconds were classified as late misses, and multiple responses in a single trial were classified as false alarms. DRT trials were presented with inter-stimulus intervals of 2–4 seconds, per the ISO standard (International Organization for Standardization, 2016). All stimuli were presented in green, of RGB values 8, 219, 78, on a black background.

## 2.4. Apparatus

Figure 3 depicts an example screen from the experiment. The dimensions of stimuli are hereafter reported in degrees of viewing angle. Given the monitor size and resolution, with participants seating at a roughly fixed distance of 80 cm, one degree of viewing angle was equal to approximately 51 pixels or approximately 1.4 cm of screen space. Degrees of viewing angle were used to calculate stimuli size to ensure perceptual control across participants. The experiment was presented using PsychoPy 1.85.3 on a Dell S2240Lc LCD monitor, with screen dimensions of 53x30cm and a resolution of 1920 × 1080 pixels.

## 2.5. Procedure

Demographic information was collected from participants, and they were seated in a quiet room in front of the test

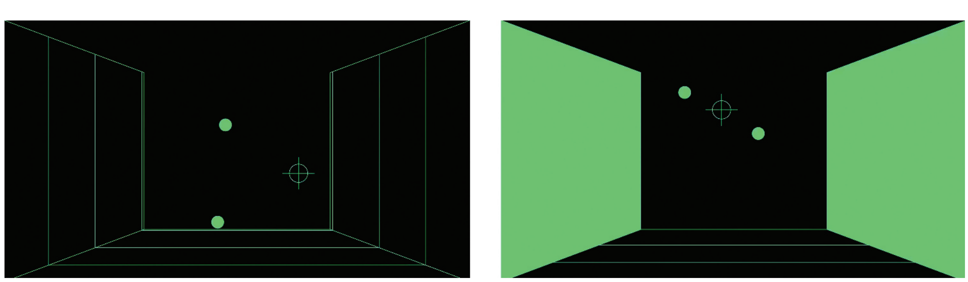

Figure 3. Example screens from high-load trials with DRT stimulus absent (left) and present (right).

monitor. Their distance from the monitor was measured to ensure 80 cm distance. An interactive instruction phase was presented to explain the various tasks. A practice block was then presented. Following this, fifteen blocks of 60 second duration were presented. Low- and high-load trials were randomly presented within each block. There was not a set number of trials within a block, instead participants completed trials until the block ended. Each block was separated by 15 second breaks to manage fatigue.

### 2.6. Data analysis

Mouse movements on targeting trials were analyzed by first collecting a time series of mouse motion in relation to the target(s), at a time resolution of 60 Hz. Two data sets were derived from this motion data. *Movement onset* was calculated by locating the first movement of the mouse toward the target, defined as any movement within 90° of the target. The *number of submovements* on each trial was approximated by identifying and counting peaks (Hogan & Sternad, 2009) in mouse speed in each trajectory, from the first movement toward the target to the completion of the trial. This value was divided by the time taken to complete the trial. This division was done for normalization, to control for the random distance the participant needed to move the mouse on each trial, since a longer distance to travel may lead to longer duration trials, which would likely result in more submovements due to the strong correlation between number of submovements and movement duration (Park et al., 2017). This adjusted dependent variable is therefore defined as *scaled* submovements.

Figure 4 shows the relationship between some example mouse movements and the resulting data. In panel 1, the movement of the crosshair to the chosen target, represented by the white line, has relatively few submovements, while the movement in panel 2 has more submovements. The results in panel 3 illustrate the difference between the first two panels. Assuming performance in the single-target condition is similar for both data sets, the number of submovements in the data from panel 2 increases with primary task workload, suggesting a greater increase in workload. The data from panel 1 shows only a small increase in submovements, suggesting little effect of primary task workload. For further analyses utilizing alternative measures of performance, see Appendix A.

As primary-task events and DRT signals were presented independently, there was no guarantee that a DRT trial would occur during only one condition of primary task workload. Given the DRT must be presented alongside a primary task of consistent workload, it was necessary to remove DRT trials that occurred during both low- and high-load trials. To maximize observation numbers, DRT trials were considered valid if at least 90% of the trial occurred during only one primary task workload condition. Consider the example of a DRT trial with a response time of 500 ms. If the DRT stimulus was presented during a single-target trial on the primary targeting task, and the participant responded to the DRT during a subsequent double-target trial, the DRT trial may be invalid as it occurred during two levels of primary task workload. However, if at least 450 ms of the DRT trial occurred during one of those two targeting task trials, it would be considered valid as 90% of the trial's time took place alongside only one level of primary task workload.

Data from five participants were excluded due to low response rates (missing more than 50% of trials on either task). Very fast trials ($<$ 150 ms) were removed from both primary task and DRT data as they would not represent conscious responses to the current trial, but rather an accidental response or carry-over response from a previous trial. Group-level comparisons were carried out using paired-samples *t*-tests and mixed-design ANOVA. The latter tests were chosen to detect potential interactions between the two factors, and because one factor, primary task workload, was within-subjects while the other, stopping rule, was between-subjects. Equivalent Bayesian tests in cases of non-significant results, using JASP (JASP Team, 2018). The latter analyses were carried out because frequentist tests cannot provide evidence in favor of null results, whereas Bayesian tests can (Lee & Wagenmakers, 2014). Where Bayesian analysis is reported, two kinds of Bayes Factors are used – $BF_{10}$, which provides evidence for an effect against a null model, similarly to a *t*-test, and $BF_{\text{Inclusion}}$, which represents evidence in favor of including a given factor in an explanatory model of the data. Despite this subtle difference, both kinds of Bayes Factors are interpreted the same way. We used Jeffreys (1961) classification scheme to interpret Bayes Factors. In this convention, a *BF* of 1 to 3 represents anecdotal evidence, 3 to 10 represents moderate evidence, 10 to 30 represents strong evidence, 30 to 100 represents very strong evidence, and greater than 100 represents extreme evidence. Each of these values applies to the alternative hypothesis. Values of

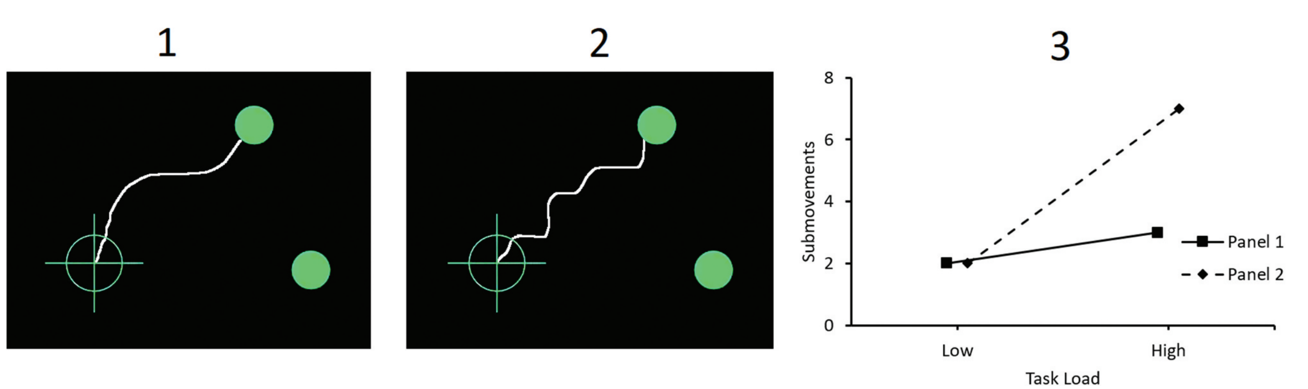


Figure 4. Example mouse movements with few submovements (panel 1) and many submovements (panel 2), with the resulting trends in data (panel 3).

less than one provide evidence for the null hypothesis in the same manner, so a *BF* of 0.1, or 1/10, represents the same strength of evidence for the null that a *BF* of 10 would for the alternative.

## 3. Results

Before investigating mouse trajectory data, it is necessary to first apply a more traditional analysis of response time data to provide a baseline for the new analyses. When two targets were presented, participants in the exhaustive group were required to hover the mouse over first one target, then the other, as described in the Method section. The signal detection and movement decision processes are assumed to happen during the first of these two targeting legs, as they must be at least partly completed before the first leg could be completed. If the presence of a second target increased the workload required to perceive the targets, plan a movement and execute it, the effect of this increased workload is likely to be seen in response times on the first leg of the trial, but not the second. Indeed, because the legs must be completed in succession, some of the planning and execution for the second leg may have taken place during the first leg of the trial, making the second leg faster. The first leg of the high-load trials was otherwise identical to the low-load trials, with participants hover the mouse over a single target. Any difference from the low-load condition to the first leg of the high-load condition would therefore be due to the effect of processing the second target. For this reason, mouse movements from the first leg only are used for the comparisons below.

### 3.1. RT analysis

The left panel of Figure 5 shows mean targeting task RTs across levels of primary task workload and stopping rule. A main effect of primary task workload was found, with mean targeting RT faster on low-load trials ($M$ = 568 ms, $SD$ = 118 ms) than on high-load trials ($M$ = 576 ms, $SD$ = 118 ms), $F(1, 58) = 18.69$, $p < .001$. No main effect of stopping rule was found, with Bayesian analysis providing anecdotal evidence against stopping rule's inclusion in an explanatory model, $BF_{\text{Inclusion}} = 0.96$. A significant interaction was observed, $F(1, 58) = 42.14$, $p < .001$, due to a greater effect of primary task workload for the exhaustive condition. Indeed, an analysis of simple effects shows that RT increased significantly in the exhaustive condition, $F(1, 28) = 50.91$, $p < .001$, but not in the minimum-time condition, which showed anecdotal evidence in favor of no difference, $BF_{10} = 0.64$. It must also be noted that the observed main effect was very small, with a mean difference of just 8 ms. Considering the display refreshed at a rate of 60 Hz, this difference represents less than one frame's worth of time. This finding should therefore be interpreted with some caution. It should be noted that each of the panels in Figure 5 represents an independent task. Although they share a similar time-scale, they are not directly comparable.

The right panel of Figure 5 shows mean DRT RT across primary task workload and stopping rule conditions. No main effect of primary task workload on mean RT was found, with anecdotal evidence against including primary task workload as an explanatory factor, $BF_{\text{Inclusion}} = 0.53$. No main effect of stopping rule was found, again with anecdotal evidence against including the factor, $BF_{\text{Inclusio}}$n = 0.70. A significant interaction effect was found, $F(1, 58) = 5.89$, $p = .018$. An analysis of simple effects shows this interaction was driven by an increase in mean RT with increasing primary task workload for the exhaustive condition, $F(1, 28) = 6.48$, $p = .017$, but not for the minimum-time condition, with moderate evidence in favor of no difference in this condition, $BF_{10} = 0.23$. This pattern of results indicates that increased primary task workload had a deleterious effect on participants' cognitive capacity in the exhaustive condition only.

No differences in mean DRT hit rate was found, with moderate evidence against including primary task workload as a factor, $BF_{\text{Inclusion}} = 0.27$, and anecdotal evidence against including stopping rule, $BF_{\text{Inclusi}}$on = 0.63. No differences in DRT false alarm rate was found, again with moderate evidence against including primary task workload, $BF_{\text{Inclusion}} = 0.19$, and anecdotal evidence against including stopping rule, $BF_{\text{Inclusion}} = 0.46$.

### 3.2. Mouse trajectories

A main effect of primary task workload was found on mouse trajectories – when participants were presented with two targets, onset times were slower ($M$ = 231 ms, $SD$ = 33 ms) than when only one target was presented ($M$ = 223 ms, $SD$ = 34 ms), $F(1, 58) = 83.62$, $p < .001$. No significant

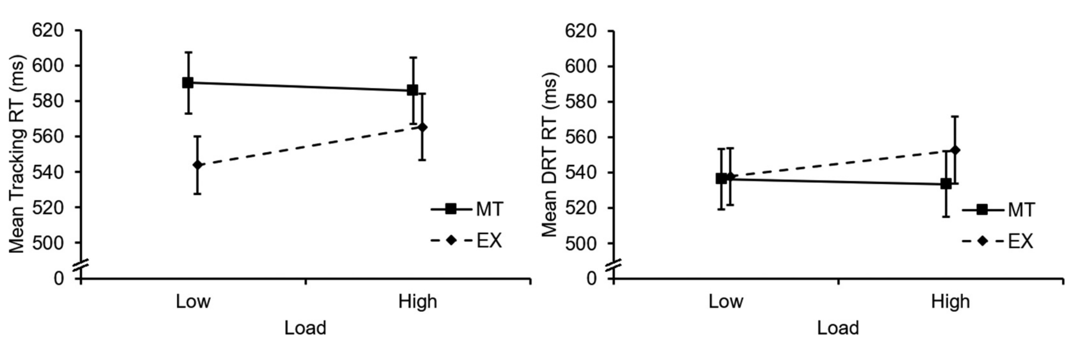


Figure 5. Mean RT on the targeting task (left) and DRT (right) across levels of primary task workload and stopping rule (MT = Minimum Time, EX = Exhaustive). Error bars represent one standard error. Data points are offset to aid readability.

main effect of stopping rule was found, with Bayesian analysis indicating anecdotal evidence against including stopping rule as a factor in an explanatory model, $BF_{Inclusion} = 0.46$. As the left panel of Figure 6 shows, a significant interaction effect was found between primary task workload and stopping rule, $F(1, 58) = 284.26$, $p < .001$. An analysis of simple effects shows that, in the minimum-time condition, onset times were faster under high primary task workload than under low workload, $F(1, 30) = 66.59$, $p < .001$. In contrast, in the exhaustive condition onset times were slower under high primary task workload than under low workload, $F(1, 28) = 207.62$, $p < .001$. It is worth noting that the low-load trials were identical for both stopping rules. Any observed difference would be due to the context effect of the high-load conditions, which had different demands across the two stopping rules. An analysis of simple effects shows the difference between onsets in low-load trials approached significance, with Bayesian analysis presenting anecdotal evidence for this difference, $BF_{10} = 1.14$. Given the relatively low Bayes Factor and the lack of a similar effect in other measures, this represents little evidence for a context effect.

When scaled by trial duration, participants made more submovements on trials with two targets ($M = 3.52$ peaks/sec, $SD = 0.23$ peaks/sec) than on trials with only one target ($M = 3.38$ peaks/sec), $SD = 0.17$ peaks/sec), $F(1, 58) = 127.72$, $p < .001$. A significant main effect of stopping rule was also found, $F(1, 58) = 11.67$, $p = .001$, and a highly significant interaction between the two factors, $F(1, 58) = 62.87$, $p < .001$. As the right panel of Figure 6 shows, the latter two effects were driven by the exhaustive high-load condition, which was greater than the other three conditions. Taken together, these analyses indicate that performance decreased under high load in the exhaustive condition, as predicted. However, this decrease was not observed in the minimum-time condition, which did not reflect our hypothesis, or any of the predicted outcomes.

### 3.3. Target choice

On high-load trials participants were presented concurrently with two targets on the screen. If the relative distance from the starting point to each of the targets had no bearing on which target was chosen, the likelihood they chose the nearest of the two was 50%, while a deviation might suggest they invested some processing resources in identifying the closer target, to minimize mouse travel distance and time. One-sample $t$-tests were used to assess these deviations. As Figure 7 shows, participants chose the closer target on a high-load trial significantly more often than chance, denoted by the dashed line in Figure 7, on both the minimum-time ($M = 60\%$, $SD = 3\%$), $t(30) = 17.16$, $p < .001$, and the exhaustive conditions ($M = 55\%$, $SD = 3\%$), $t(28) = 11.06$, $p < .001$. The two groups were compared using an independent-samples $t$-test. The closer target was chosen significantly more in the minimum-time condition than in the exhaustive condition, $t(58) = 5.97$, $p < .001$. These findings

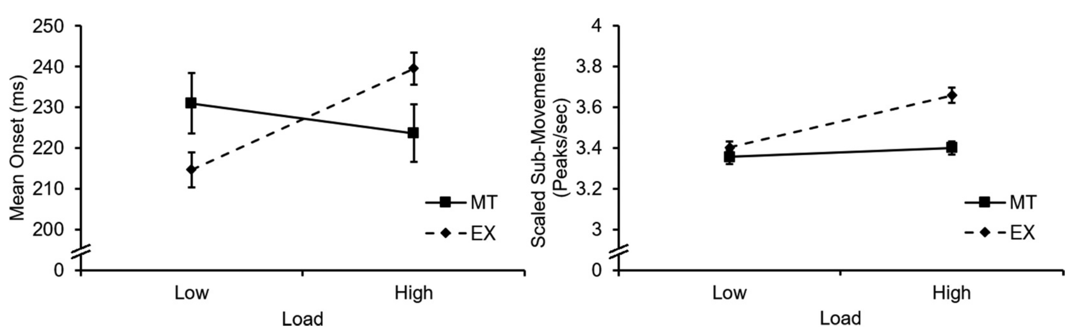


**Figure 6.** Mean movement onset times (left) and mean scaled submovements (right) across levels of task load and stopping rule. Error bars represent one standard error.

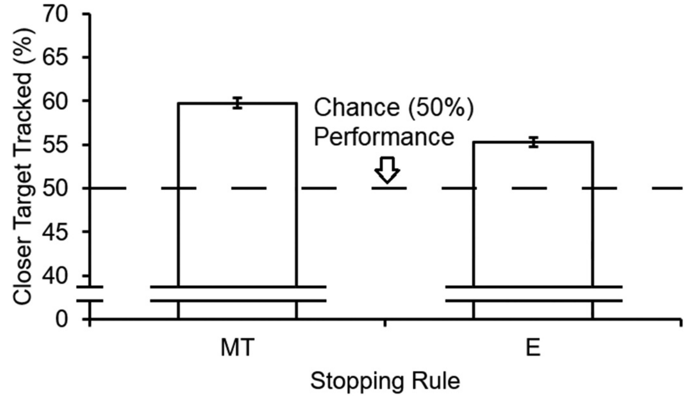


**Figure 7.** Proportion of trials in which the closer target was chosen across stopping rule conditions. Dashed line represents chance performance.

indicate participants were able to identify and choose the closer target, suggesting some planning of movements.

We can also assess how participants' target choices were affected by onset time. If participants who waited longer to begin a movement were more accurate than those who began movements earlier, it would suggest participants were planning their movements based on which target was closer. However, a negative correlation was found between onset time and target choice accuracy, when accuracy was defined as the proportion of trials on which the closer target was chosen, $r(58) = -.26$, $p = .042$. Given that the correlation was weak, with a $p$-value close to the .05 significance level, and that the relationship was negative when we would expect a positive correlation between onset time and accuracy, this finding does not support the prediction that participants planned their movements based on target proximity. Taken together, these findings offer mixed evidence for movement planning based on choosing the closer target.

### 3.4. Variable trial numbers

Due to the relatively continuous design of the task, there was no set number of trials per block; instead, participants would complete as many trials as they could in the allotted time. This leads to the potential confounding factor of trial numbers on performance and workload. If completing more trials increased workload or led to fatigue, then participants who completed more trials could have experienced a more demanding task than those who completed relatively few trials.

Participants in the minimum-time group completed significantly more trials ($M = 1262$, $SD = 155$) than those in the exhaustive group ($M = 922.90$, $SD = 79$), $t(58) = 10.59$, $p < .001$. This difference was also significant for low primary task workload, $t(58) = 9.61$, $p < .001$, and high workload, $t(58) = 11.21$, $p < .001$. Despite completing more trials, no main effects of stopping rule were found on any measure of targeting task performance or workload, as discussed above. This suggests the increased trial numbers did not affect performance on either the targeting task or DRT. The discrepancy in trial numbers between the groups may be due to the differing nature of the two stopping rules – in the minimum-time group, participants could begin moving toward a target as soon as they perceived one, whereas those in the exhaustive group were sometimes required to perceive two targets, leading to longer onset times on these trials. However, with the lack of evidence of a relationship between trial numbers and either task performance or workload, further exploration of this relationship is beyond the scope of the current study.

## 4. Discussion

As predicted, participants exhibited slower RTs on both the primary task and DRT as primary task workload increased in the exhaustive condition, though DRT accuracy was not affected by workload. Furthermore, mouse movement onsets were later and exhibited more submovements under high workload. However, this pattern was not observed for the minimum-time condition. This suggests adding a second target did not increase the workload demand of the task when that target was redundant, whereby participants could choose to ignore it. By contrast, adding a second target which needed to be targeted along with the first target did increase workload.

Our analysis of mouse trajectories agreed with the more traditional RT analysis in its findings, but added nuance to the findings. We found that, when primary task workload was associated with slower RTs, both the planning and execution stages of mouse movements were affected. This finding was predicted by our hypothesis that the third possible outcome presented in the Introduction was most likely. If only the planning stage was affected, as in the second possible outcome, we would expect only onset times to be affected, as there would be no difference in the way movement was optimized across primary task workload levels. In contrast, if only execution was affected as in the first possible outcome, we would expect to see movements begin just as quickly or even quicker when more targets were present, but trajectories would contain more submovements. Additionally, our finding of more submovements under high workload indicates participants were beginning their movements before the planning stage of the movement had finished. This suggests the two stages could be completed concurrently, with further movement planning being undertaken even as mouse movements were being executed. We note that these differences in submovements were observed even though the motor task was essentially the same between workload levels for the part of the movement we analyzed (moving from one target to another).

Stopping rule had little overall effect on performance. The only main effect of stopping rule was driven by the high-load condition, which was where the two stopping rules differed in their demands. In the minimum-time condition, simply perceiving a single target was sufficient to complete the trial. Indeed, the presence of multiple targets led to faster onsets, suggesting a redundant target effect whereby added information leads to faster performance. In the exhaustive condition, each target needed to be hovered over before the trial was complete. It is therefore unsurprising that participants in this condition took longer in the planning stage, as there was a strategic benefit to targeting the closer target first, as well as in the execution phase, because the target not chosen in the first leg of the trial was relevant to the second leg. The latter was reflected in the increased submovements with increased primary task workload in the exhaustive condition, but not in the minimum-time condition. The former was addressed by assessing target choice data. Our finding that participants chose the optimal target at greater than chance supports previous findings that under time pressure, subjects are able to choose the better option (Brenner & Smeets, 2015). However, our findings did not support the prediction that the slower onsets in the exhaustive condition were driven by participants spending more time choosing the optimal target. Indeed, slower onsets were associated lower, not higher accuracy. More generally, our findings offer mixed evidence that participants reliably chose a target before moving – above-chance accuracy suggests participants were able to identify and hover the mouse over the closer target, but in some cases participants may have simply chosen targets as they were processed.

### 4.1. Implications

The finding of both increased movement onsets and increased submovements under high workload lends support to model of movement responses whereby planning and execution can happen concurrently or even in parallel, with different processing channels dedicated to each stage of the response. Under such a model, the planning stage need not be completed before the execution stage begins. Under the predictions of Multiple Resource Theory, the two stages may partially share cognitive resources, as planning and executive functioning is theorized to utilize the same resources as response execution. However, the perceptual component of the planning stage, in which the target is located, would utilize a resource dedicated to visual processing, and would therefore interfere less with execution. The agreement between the submovement and DRT analyses suggests the former can be used to assess workload in experimental paradigms that traditionally use trial-by-trial designs, and are hence unsuitable for deployment alongside the DRT. By altering trial-by-trial designs such as those used for discrete choice or signal detection experiments to require a physical response beyond a button press, as was done in the current study, new insight can be gained into the workload demands of these tasks.

The use of trajectory data to assess cognitive workload also has applications beyond the laboratory. The workload imposition of a computer-based system has an impact on the usability of that system and user performance. Measuring this workload imposition using data from the system itself, rather than an external apparatus such as a physiological measure, would allow developers and researchers to estimate cognitive workload *in situ* in an unobtrusive way. Our findings imply mouse trajectories can be used in this way to assist UI in development to ensure UIs do not impose more cognitive workload than necessary, or to assess existing systems to assess how users respond to different computer-based environments. This could be done by preparing candidate configurations of UIs and comparing the relative utility and cognitive demands of these configurations to identify optimal configurations for particular scenarios. Such research has previously been carried out using the DRT (Innes et al., 2018), but submovement analysis could also be used in this application.

Aside from assessing UIs, submovement analysis could also be used to assess users – a user's cognitive state could be tracked over several sessions to assess how effectively they are learning to use an interface or evaluate the effectiveness of training. Users could also be compared to identify individual differences for the purposes of selection or recruitment. There is also a potential application of movement trajectories in touchscreen use, given the increased use of smartphones and tablets in recent years. User behavior during web site navigation could be analyzed to evaluate web page designs, or to evaluate the effectiveness of web advertising to capture users' attention. However, further research is required to investigate whether the current study's findings can be applied to touchscreen-based tasks.

A measure based on performance within a system also has the advantage of potentially operating as an online, real-time measure of workload. *Adaptive systems* can classify a user's workload state based on task performance or other measures, such as physiological measures, and change features of the system to minimize the workload imposition on the user (Durkee et al., 2015). Such a system requires objective, real-time measures of workload, which can operate while the user completes their primary task – neurological measures such as functional magnetic resonance imaging, for instance, could not be applied in an in-car system, simply because the apparatus is too intrusive. There is therefore an advantage to a measure that can be built in to a computer-based system or UI. Our findings indicate that, in a computer-based task where the user uses a mouse as an input device, user behavior could be analyzed to diagnose user workload. Whether the temporal resolution of these measures is high enough to be usable as a real-time measure is an open question, and further research is required to address it.

### 4.2. Considerations

In applying psychological experimental designs to real-world scenarios, some compromises are made to benefit applicability at the expense of experimental control. Nevertheless, the current study's design was less controlled than other redundant target or choice tasks, even compared to other arm reaching experiments. Other studies present targets in constant positions, and may require participants to return their hand or mouse to a constant starting position (Finkbeiner & Heathcote, 2016; Friedman et al., 2013; Grimes & Valacich, 2015). In the current study, each trial began where the previous trial ended, and the position of the targets was random, within some limits. The distance between each target and the mouse cursor was also not equal – indeed, the above analysis would not be possible if targets were equidistant from the cursor. This loss of control affects our ability to draw some of the conclusions available to previous researchers. For example, the curvature of the trajectory can be used as a measure of distraction, because it is assumed the arm or mouse is drawn toward the position of the distractor stimulus (Dias da Silva & Postma, 2020). This measure is only reliable when the angle formed by the two targets and the starting position is constant. In the current study, two targets could be placed on opposite sides of the mouse cursor, forming a 180 angle, or they could be placed adjacent to one another. This renders the use of trajectory curvature in the current study less reliable, as the relationship between the curvature and a theoretical explanation for the curvature would be unclear. There was also no "reset" at the end of each trial where the mouse cursor would be moved to a constant starting position, nor was there an inter-trial interval. These features may affect the size of the effect observed (Kieslich et al., 2020). The decisions not to include these features were made in an effort to present a more continuous task, but this could have introduced the potential for sequential effects from trial to trial, which would not be present in a more traditional arm reaching experiment.

Another potential issue with our measurement of performance is its sensitivity. Movement trajectories were chosen as a dependent variable as it reflects real-world behavior and could be used to differentiate the effect of cognitive workload at each stage of the response. However, it may not be sufficiently sensitive to detect subtle changes in workload, especially on a short time scale. This could limit its applicability in real-world settings. Nevertheless, it was sufficiently sensitive to detect the same difference in performance on the primary task that the DRT detected in residual capacity, as seen in the right panels of Figures 5 and Figures 6. This measure also demonstrated its value in addressing theoretical questions about cognitive workload in mouse-based responses, which was the original focus of the current study. Further research is required to determine how quickly a reliable estimate of participant workload can be established, which would inform how applicable the measure is to such scenarios as adaptive user interfaces, which require very high temporal resolution. It may be the case that, even with relatively low sensitivity, submovements could be used to detect individual differences between users, allowing for user evaluation, or for use in an adaptive system that adapts a UI on a user-by-user basis, rather than adapting the UI over time for the same user. They could also be used for evaluating UI designs in an objective way, by comparing the cognitive impacts of different configurations of a UI on a user.

We speculated above that the faster onset times in the minimum-time condition indicate a redundant target effect. However, this was not replicated in the other measures of primary task performance. It is therefore unclear whether we actually observed a redundant target effect. Further investigation of a possible redundant target effect is possible using systems factorial technology (SFT), a non-parametric methodology that utilizes RT distributions (Townsend & Nozawa, 1995). Specifically, the *workload capacity*, or processing efficiency, of a process can be estimated (Townsend & Eidels, 2011) by comparing empirical data to a benchmark model, such as Miller's Race Model (Miller, 1982), which assumes the two targets are processed independently and in parallel, and the faster of the two processes to finish drives the participant's response. Given that we also inferred that our pattern of results indicated no increased workload demand in the minimum-time condition, analysis using SFT could further enrich our analysis and answer currently opaque questions, including the efficiency of target processing.

Another issue relates to the manipulation of primary task workload. The addition of a second target was assumed to increase primary task workload. Following on from the previous paragraph, if we assume the perception of a target and the planning and execution of a response requires some amount of cognitive resources, such as working memory, then the addition of a second target would lead to higher primary task workload in a *limited-capacity* system (Townsend & Eidels, 2011). Only in the case of an *unlimited-* or *super-capacity* system, which could process multiple targets with no loss in efficiency, would the addition of a second target not impact the participant more than a single target alone. This was the foundation of our choice to manipulate primary task workload with the addition of a second target. However, without further analysis using SFT, we cannot be sure this extra processing demand affected participants' performance. For example, if they simply ignored the second stimulus on high-load trials, they would not experience the expected increase in workload on those trials, and this would not be detectable in the current analysis. There is some evidence against such a confounding strategy – the finding that participants chose the closer target at a rate greater than chance suggests they processed both targets on at least some trials. This, in combination with the slower onset times in the exhaustive condition, indicates some planning of movements, which in turn suggests more cognitive resources would have been required to complete these trials.

Additionally, the increased demand of perceiving and responding to a second target may not have sufficiently increased primary task workload to be an effective workload manipulation. Given the lack of a significant difference in DRT performance across primary task workload conditions in the minimum-time group, the high-load condition did not appear to increase participants' overall workload to an extent the DRT could detect. It may still be the case that participants processed the two targets using a limited-capacity process, that is to say at some cognitive cost, but this cost cannot be quantified using the current analysis. Additional analysis using SFT could further illuminate this question, while future research could also utilize a more demanding workload manipulation, such as the introduction of a new task, to greater differentiate the workload conditions.

### 4.3. Conclusion

We presented a mouse-based targeting task alongside a computer-based DRT to assess the effect of increased cognitive workload on users' cognitive state. We applied an analysis of mouse movement trajectories to the resulting data set to investigate whether such analysis could be applied to more real-world data sets. Increased primary task workload had a detrimental effect on users' cognitive capacity as measured by the DRT, and their performance on the primary task in the form of slower RTs, but not in conditions where redundant targets were presented. Trajectory analysis indicated that, in cases of slower RTs under high primary task workload, mouse movements began later and contained more submovements than under low workload. This analysis provided greater nuance to our findings, and suggests the analysis of movement trajectories can be applied to real-world tasks such as UIs. Because the data were derived from the application itself, no additional apparatus were required, allowing for the use of this methodology beyond lab-based experiments. Movement trajectories could be used in the assessment of user workload to assist in UI development or research into the workload imposition of computer-based systems.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was supported by an Australian Research Council grant [DP160102360] to AE, and by an Australian Government Research Training Program (RTP) Scholarship awarded to Alexander Thorpe

## ORCID

Alexander Thorpe http://orcid.org/0000-0001-9267-1175

## References

Bach, K. M., Jaeger, M. G., Skov, M. B., & Thomassen, N. G. (2009). Interacting with invehicle systems: Understanding, measuring, and evaluating attention. In *Proceedings of the 23rd British hci group annual conference on people and computers: Celebrating people and technology*. ACM.

Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., & Burdet, E. (2015). On the analysis of movement smoothness. *Journal of Neuroengineering and Rehabilitation*, *112*(1). https://doi.org/10.1186/s12984-015-0090-9

Brenner, E., & Smeets, J. B. J. (2015). Quickly making the correct choice. *Vision Research*, *113*(Part B), 198–210. Retrieved April 27, 2015, from http://www.sciencedirect.com/science/article/pii/S0042698915001571

Bundt, C., Ruitenberg, M. F. L., Abrahamse, E. L., & Notebaert, W. (2018, 5). Early and late indications of item-specific control in a stroop mouse tracking study. *PLOS ONE*, *13*(5), 1–14. https://doi.org/10.1371/journal.pone.0197278

Causse, M., Imbert, J. P., Giraudet, L., Jouffrais, C., & Tremblay, S. (2016). The role of cognitive and perceptual loads in inattentional deafness. *Frontiers in Human Neuroscience*, *10*, 344. https://doi.org/10.3389/fnhum.2016.00344

Causse, M., Peysakhovich, V., & Fabre, E. F. (2016). High working memory load impairs language processing during a simulated piloting task: An erp and pupillometry study. *Frontiers in Human Neuroscience*, *10*, 14. https://doi.org/10.3389/fnhum.2016.00240

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*(1), 269–298. (PMID: 20345247). https://doi.org/10.1146/annurev.neuro.051508.135409

Dias da Silva, M., & Postma, M. (2020, November). Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering. *Computers in Human Behavior*, *112*, 106453. https://doi.org/10.1016/j.chb.2020.106453

Dounskaia, N., & Goble, J. A. (2011, 2 3). The role of vision, speed, and attention in overcoming directional biases during arm movements. *Experimental Brain Research*, *209*, 299–309. https://doi.org/10.1007/s00221-011-2547-9

Durkee, K. T., Pappada, S. M., Ortiz, A. E., Feeney, J. J., & Galster, S. M. (2015). System decision framework for augmenting human performance using real-time workload classifiers [Conference Proceedings]. In *2015 IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision* (p. 8–13). IEEE. http://ieeexplore.ieee.org/document/7107968/

Endsley, M. R. (1993). Situation awareness and workloadflip sides of the same coin. In *International symposium on aviation psychology, 7 th* (pp. 906–911). Columbus, OH: The Laboratory.

Erb, C., Moher, J., Sobel, D., & Song, J.-H. (2016, 7). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, *152*, 114–126. https://doi.org/10.1016/j.cognition.2016.03.015

Finkbeiner, M., & Heathcote, A. (2016, 4). Distinguishing the time-and magnitude-difference accounts of the simon effect: Evidence from the reach-to-touch paradigm. *Attention, Perception & Psychophysics*, *78*(3), 848–867. https://doi.org/10.3758/s13414-015-1044-9

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *74*(6), 381–391. https://doi.org/10.1037/h0055392

Flash, T., & Henis, E. A. (1991). Arm trajectory modification during reaching towards visual targets. *Journal of Cognitive Neuroscience*, *3*(3), 220–230. https://doi.org/10.1162/jocn.1991.3.3.220

Forster, S., & Lavie, N. (2007). High perceptual load makes everybody equal: Eliminating individual differences in distractibility with load: Research report. *Psychological Science*, *18*(5), 377–381. https://doi.org/10.1111/j.1467-9280.2007.01908.x

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315–323. (Publisher: SAGE Publications Inc). https://doi.org/10.1177/0963721417746793

Friedman, J., Brown, S., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, *57*(3), 140–151. https://doi.org/10.1016/j.jmp.2013.06.005

Grimes, G., & Valacich, J. (2015). Mind over mouse: The effect of cognitive load on mouse movement behavior. In *2015 international conference on information systems: Exploring the information frontier, icis 2015*. Association for Information Systems. (2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015; Conference date: 13-12-2015 Through 16-12–2015).

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load [Conference Proceedings]. In *Proceedings of the 12th acm international conference on ubiquitous computing* (p. 301–310). ACM. https://doi.org/10.1145/1864349.1864395

Hogan, N., & Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of Motor Behavior*, *41*(6), 529–534. https://doi.org/10.3200/35-09-004-RC

Innes, R., Howard, Z., Eidels, A., & Brown, S. (2018). *Cognitive workload measure and analysis* (Tech. Rep.). University of Newcastle.

International Organization for Standardization. (2016). *Road vehicles. transport information and control systems. detection-response task (drt) for assessing attentional effects of cognitive load in driving.* (ISO Standard No. 17488:2016). https://www.iso.org/standard/59887.html

JASP Team. (2018). *JASP (Version 0.9)*[Computer software]. https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press. *MR0187257*.

Jiang, X., Zheng, B., Bednarik, R., & Atkins, M. S. (2015). Pupil responses to continuous aiming movements. *International Journal of Human-Computer Studies*, *83*, 1–11. https://doi.org/10.1016/j.ijhcs.2015.05.006

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020, February). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, *52*(1), 317–341. Retrieved June 21, 2020, from. https://doi.org/10.3758/s13428-019-01228-y

Kim, K. H., & Wohn, K. Y. (2011). Effects on productivity and safety of map and augmented reality navigation paradigms. *IEICE Transactions on Information and Systems*, *E94-D*(5), 1051–1061. https://doi.org/10.1587/transinf.E94.D.1051

Kujala, T., & Saariluoma, P. (2011). Measuring distraction at the levels of tactical and strategic control: The limits of capacity-based measures for revealing unsafe visual sampling models. *Advances in Human-Computer Interaction*, *2011*, 1–13. https://doi.org/10.1155/2011/594353

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology. General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1346–1358. https://doi.org/10.1037/a0016454

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Macdonald, J. S. P., & Lavie, N. (2008). Load induced blindness. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1078–1091. https://doi.org/10.1037/0096-1523.34.5.1078

MacKenzie, I. S. (2018). Fitts' law. In K. L. Norman & J. Kirakowski (Eds.), *The wiley handbook of human computer interaction* (Vol. 1, pp. 349–370). Wiley.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, *121*, 62–73. https://doi.org/10.1016/j.biopsycho.2016.10.002

Meyer, D., Abrams, R., Kornblum, S., Wright, C., & Smith, J. (1988, 8). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, *95*(3), 340–370. https://doi.org/10.1037/0033-295X.95.3.340

Miall, R. C., Weir, D. J., & Stein, J. F. (1993). Intermittency in human manual tracking tasks. *Journal of Motor Behavior*, *25*(1), 53–63. https://doi.org/10.1080/00222895.1993.9941639

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Or, C. K. L., & Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics*, *7*(2), 83–94. https://doi.org/10.3233/OER-2007-7202

Park, S.-W., Marino, H., Charles, S. K., Sternad, D., & Hogan, N. (2017, March). Moving slowly is hard for humans: Limitations of dynamic primitives. *Journal of Neurophysiology*, *118*(1), 69–83. https://doi.org/10.1152/jn.00643.2016

Patel, N., Vytal, K., Pavletic, N., Stoodley, C., Pine, D. S., Grillon, C., & Ernst, M. (2016). Interaction of threat and verbal working memory in adolescents. *Psychophysiology*, *53*(4), 518–526. https://doi.org/10.1111/psyp.12582

Pereira, M., Sobolewski, A., & Millán, J. D. R. (2017). Action monitoring cortical activity coupled to submovements. *eNeuro*, *4*(5), e0241–17. https://doi.org/10.1523/ENEURO.0241-17.2017

Rajan, R., Selker, T., & Lane, I. (2016). Task load estimation and mediation using psychophysiological measures [Conference Proceedings]. In *21st international conference on intelligent user interfaces, iui 2016* (Vol. 07-10-March-2016, p. 48–59). Association for Computing Machinery. http://dl.acm.org/citation.cfm?doid=2856767.2856769

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

Rottger, S., Bali, K., & Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, *52*(5), 512–523. https://doi.org/10.1080/00140130802379129

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991–1009. https://doi.org/10.1016/j.ergon.2005.04.005

Salvucci, D. D., & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: Experiment and theoretical account. *Human Factors*, *50*(5), 834–844. https://doi.org/10.1518/001872008X354200

Senanayake, R., Goonetilleke, R. S., & Hoffmann, E. R. (2015). Targeted-tracking with pointing devices. *IEEE Transactions on Human-Machine Systems*, *45*(4), 431–441. https://doi.org/10.1109/THMS.2015.2408260

Shadmehr, R., Reppert, T., Summerside, E., Yoon, T., & Ahmed, A. (2019, May). Movement vigor as a reflection of subjective economic utility. *Trends in Neurosciences*, *42*(5), 323–336. (Funding Information: This work was supported by grants from the National Institutes of Health (1-R01-NS078311, 1-R01-NS096083, 1-F32-EY028846), the Office of Naval Research (N00014-15-1-2312), and the National Science Foundation (1723967 and CAREER award 1352632). Publisher Copyright: © 2019 Elsevier Ltd). https://doi.org/10.1016/j.tins.2019.02.003

Shelton, B., Nesbitt, K., Thorpe, A., & Eidels, A. (2020). Gauging the utility of ambient displays by measuring cognitive load. *Cognition, Technology & Work*, *23*, 459–480. https://doi.org/10.1007/s10111-020-00639-8

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. (PMID: 10694957). https://doi.org/10.1068/p281059

Stevens, C. J., Gibert, G., Leung, Y., & Zhang, Z. (2013). Evaluating a synthetic talking head using a dual task: Modality effects on speech understanding and cognitive load. *International Journal of Human-Computer Studies*, *71*(4), 440–454. https://doi.org/10.1016/j.ijhcs.2012.12.003

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

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. *Special Issue: Everyday Attention - Part I/L'attention Au Quotidian - Partie I*, *71*(2), 93–110. http://ovidsp.ovid.com/ovidweb.cgi?T=JSPAGE=referenceD=psyc13NEWS=NAN=2017-24966-002

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*, *57*(8), 1300–1324. https://doi.org/10.1177/0018720815575149

Thorpe, A., Innes, R., Townsend, J., Heath, R., Nesbitt, K., & Eidels, A. (2020). Assessing cross-modal interference in the detection response task. *Journal of Mathematical Psychology*, *98*, 102390. https://doi.org/10.1016/j.jmp.2020.102390

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* (pp. 44:1–44:8). ACM. http://doi.acm.org/10.1145/3290688.3290749

Tillman, G., Strayer, D. L., Eidels, A., & Heathcote, A. (2017, August 1). Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception & Psychophysics*, *79*(6), 1795–1803. https://doi.org/10.3758/s13414-017-1337-2

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. https://doi.org/10.3758/s13423-011-0106-9

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*(4), 321–359. https://doi.org/10.1006/jmps.1995.1033

van Beers, R. J., Haggard, P., & Wolpert, D. M. (2004). The role of execution noise in movement variability. *Journal of Neurophysiology*, *91*(2), 1050–1063. https://doi.org/10.1152/jn.00652.2003

van Zoest, W., & Kerzel, D. (2015). The effects of saliency on manual reach trajectories and reach target selection. *Vision Research*, *113*(Part B), 179–187. (The Efficiency of Vision and Action). http://www.sciencedirect.com/science/article/pii/S0042698914003216

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance viii* (pp. 239–257). Lawrence Erlbaum.

Wickens, C. D. (2002, 1). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

## About the Authors

**Alexander Thorpe** is a researcher at the University of Newcastle, Australia. His research addresses the cognitive impacts of information delivery systems such as user interfaces on their users. He uses knowledge gained from a background in cognitive psychology to apply cognitive models of user workload to real-world scenarios.

**Jason Friedman** is a senior lecturer at Tel Aviv University, where he is codirector of the Movement Sciences lab. He combines his computational background with a clinical approach, where he works on better understanding the basic processes of how we coordinate our movements and learn to produce new movements.

**Sylvia Evans** graduated with a bachelor's degree in Psychology (Honours) from the University of Newcastle, Australia in 2019. Her Honours research focused on the impact of increased cognitive load on computer users. Her work aimed to establish new ways of measuring workload through the application of mathematical models of workload.

**Keith Nesbitt** is a Senior Lecturer in the School of Electrical Engineering and Computing at the University of Newcastle, Australia. His main areas of scientific research include Human Interface Design and Information Visualisation with a particular focus on Perception and Cognition related to Computer Games and Virtual Reality.

**Ami Eidels** is an Associate Professor at the University of Newcastle, Australia. His research focuses on information processing models, and how various sources of information are processed and combined. He is co-editor of the Oxford handbook of computational and mathematical psychology, and guest editor for the Journal of Mathematical Psychology.

## Appendix A. Alternative Analysis

***Movement Times***

An alternative analysis of movement execution is the movement time, or the time between movement onset and the completion of the trial. Movement times were compared using two-way mixed ANOVA and equivalent Bayesian tests.

Figure 1 shows mean movement times across levels of primary task workload and stopping rule. A main effect of stopping rule was found, with participants in the exhaustive group ($M$ = 441 ms, $SD$ = 34 ms) completing movements faster than those in the minimum-time group ($M$ = 486 ms, $SD$ = 74 ms), $F(1, 58) = 8.75$, $p$ =.004. No main effect of primary task workload was found, with anecdotal evidence against including the factor, $BF_{\text{Inclusion}} = 0.48$. An analysis of simple effects found that workload affected movement times for the exhaustive group, with faster movement times under high workload than low workload, $F(1, 28) = 6.49$, $p = .017$.

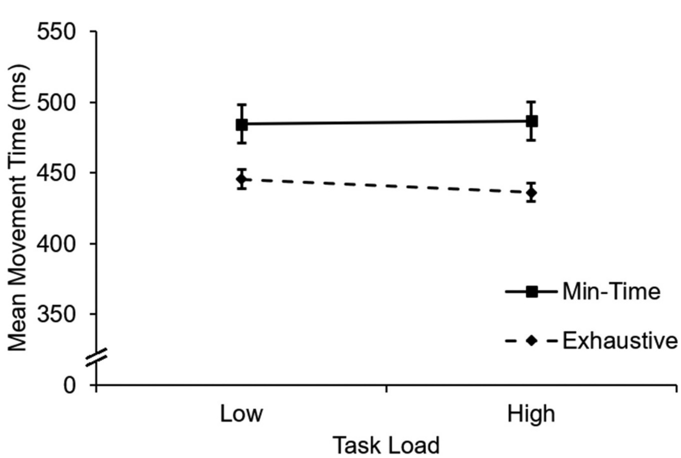


**Figure A1.** Mean movement time across levels of primary task workload and stopping rule. Error bars represent one standard error.

This pattern of results suggests the execution stage of the responses was not affected by primary task workload, which is in contrast with the findings of the trajectory analysis presented above. This does not imply one measure is incorrect, however. Slower onset times could facilitate faster movement execution, but higher primary task workload could still affect the smoothness of movements. To investigate this further, an analysis of trajectory curvature was carried out. This can be considered a complementary analysis to both the submovement analysis, as it is also derived from movement trajectories, and movement time analysis, as faster movement times have been found to be associated with smaller curvature (Bundt et al., 2018; Erb et al., 2016).

***Trajectory Curvature***

It is possible the difference in the findings of the movement time and submovement analyses were due to the scaling procedure, in which movement time was used as the scaling variable in the submovement analysis. This may have partialled out the shared variance in the two data sets. To further investigate whether the trajectory data exhibited different trends to movement time data, additional analyses of the curvature of mouse trajectories was carried out, using two dependent variables. *Scaled area* first compared the ideal trajectory toward a target, i.e., a straight line from the initial mouse position to the chosen target, to the actual movement trajectory. The area under this curvature was taken and scaled by the distance from the initial mouse position and the target, as longer distances would necessitate more movement and hence generate greater areas. The *maximum distance* between the ideal and ideal trajectories was also analysis, as a measure of the extent to which the actual movement deviated from the ideal. Analyses were carried out using two-way mixed ANOVA and equivalent Bayesian tests.

The left panel of Figure 2 shows mean scaled areas under the curve for each level of primary task workload and stopping rule. A main effect of primary task workload was found, with less curvature under low workload ($M = 3.49^{\circ 2}$, $SD = 1.00^{\circ 2}$) than high workload ($M = 3.71^{\circ 2}$, $SD = 1.00^{\circ 2}$), $F(1, 58) = 33.53$, $p < .001$. An analysis of simple effects found this effect was driven mainly by the exhaustive group, which exhibited less area under the curve under low workload ($M = 3.13^{\circ 2}$, $SD = 0.52^{\circ 2}$) than high workload ($M = 3.58^{\circ 2}$, $SD = 0.59^{\circ 2}$), $F(1, 28) = 54.30$, $p < .001$. No main effect of stopping rule was found, though Bayesian analysis found anecdotal evidence in favor of an effect, $BF_{\text{Inclusion}} = 1.25$, As is apparent from the non-parallel lines in Figure 2, a significant interaction effect between primary task workload and stopping rule was found, $F(1, 58) = 31.34$, $p < .001$.

The right panel of Figure 2 shows the mean maximum distance from ideal across each level of workload and stopping rule. A main effect of primary task workload was found, with lower maximum distance under low workload ($M = 1.22°$, $SD = 0.26°$) than high workload ($M = 1.32°$, $SD = 0.26°$), $F(1, 58) = 66.64$, $p < .001$. An analysis of simple effects again found this effect was driven by the exhaustive group, which showed significantly lower maximum distances under low workload ($M = 1.12°$, $SD = 0.16°$) than high workload ($M = 1.30°$, $SD = 0.22°$), $F(1, 58) = 74.44$, $p < .001$. No main effect of stopping rule was found, with anecdotal evidence against including this factor, $BF_{\text{Inclusion}} = 0.84$. As with scaled area, a significant interaction between primary task workload and stopping rule was found, $F(1, 58) = 37.04$, $p < .001$.

These results indicate trajectory smoothness was affected by primary task workload in the exhaustive group but not the minimum-time group, in agreement with the scaled submovement analysis and with previous studies that found trajectories with more submovements can be more curved (Flash & Henis, 1991). However, this finding is unexpected in light of the findings of the movement time analysis, and the previous findings that movement times should be positively correlated with trajectory curvature. The pattern of results presented above and in the main body of the current study suggest the movement time and trajectory

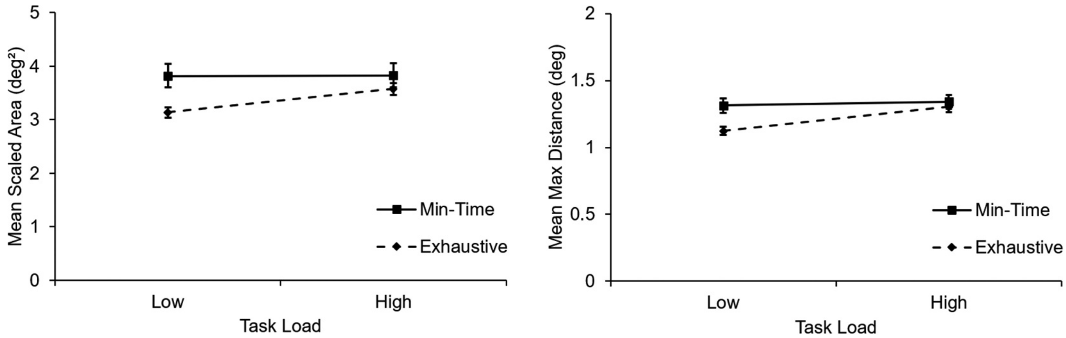


**Figure A2.** Mean scaled areas under the curve (left) and maximum distances from ideal (right) across levels of primary task workload and stopping rule. Error bars represent one standard error.

analyses reflect different behaviors, and may in fact be measuring different underlying cognitive processes or at least be affected by confounding factors to differing extents. However, further research would be required to address this discrepancy, as it is beyond the scope of the current study.