CrossMark

Modeling cognitive load effects of conversation between a passenger and driver

Gabriel Tillman[1,4] · David Strayer[2] · Ami Eidels[1] · Andrew Heathcote[1,3]

Abstract Cognitive load from secondary tasks is a source of distraction causing injuries and fatalities on the roadway. The Detection Response Task (DRT) is an international standard for assessing cognitive load on drivers' attention that can be performed as a secondary task with little to no measurable effect on the primary driving task. We investigated whether decrements in DRT performance were related to the rate of information processing, levels of response caution, or the non-decision processing of drivers. We had pairs of participants take part in the DRT while performing a simulated driving task, manipulated cognitive load via the conversation between driver and passenger, and observed associated slowing in DRT response time. Fits of the single-bound diffusion model indicated that slowing was mediated by an increase in response caution. We propose the novel hypothesis that, rather than the DRT's sensitivity to cognitive load being a direct result of a loss of information processing capacity to other tasks, it is an indirect result of a general tendency to be more cautious when making responses in more demanding situations.

✉ Gabriel Tillman
  gabrieltillman.au@gmail.com

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

[2] Department of Psychology, University of Utah, Salt Lake City, UT, USA

[3] School of Medicine, University of Tasmania, Hobart, TAS, Australia

[4] Department of Psychology, Vanderbilt University, Nashville, TN, USA

Cognitive psychologists use the term capacity to refer to the human ability to cope with the cognitive load associated with increasing amounts of perceptual information (e.g., Eidels, Donkin, Brown, & Heathcote, 2010; Townsend & Eidels, 2011). Human capacity is often limited (Kahneman, 1973), yet many situations in modern life require simultaneous processing of information from multiple signals. Given the limited capacity for processing, it is important for researchers to understand the consequences of such limitations in safety critical activities, such as driving a car.

Cognitive load from secondary tasks, such as talking on a cell phone, is one of the main studied sources of distraction while driving (Strayer et al., 2013, 2015). Distraction while driving is a significant cause of injuries and fatalities for drivers and passengers on the roadway (Dingus et al., 2006; Ranney, Mazzae, Garrott, & Goodman, 2000; Sussman, Bishop, Madnick, & Walter, 1985; Wang, Knipling, & Goodman, 1996). Strayer and Johnston (2001) studied the effects of cell phone conversations on performance in a simulated driving task. They found that conversations with either a hand-held or a hands-free cell phone while driving resulted in a failure to detect traffic signals, as well as slower reactions when the traffic signals were detected (cf. Strayer, Drews, & Johnston, 2003). Surprisingly, no such decrements are observed when a similar conversation is held between the driver and a passenger in the car (Drews, Pasupathi, & Strayer, 2008). In fact, data on crash risk reveals lower accident rates when an adult passenger is in

the car than when the driver is alone (Rueda-Domingo et al., 2004; Vollrath, Meilinger, & Krüger, 2002).

The Detection Response Task (DRT) is an international standard for assessing cognitive load on drivers' attention (International Organization for Standardization, 2015) that can safely be deployed with no appreciable effect on driving performance (Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014). The DRT measures cognitive load by asking participants in a driving simulator to respond when they detect a small light in their peripheral vision. Increases in response times (RT) in the DRT measure the effect of increased cognitive load. Although the DRT is a valid measure of the effects of cognitive load during driving (Strayer et al., 2013, 2015), there is little research on what components of DRT processing are affected by increased cognitive load.

For instance, when using a hands-free cell phone, drivers are slower to respond in the DRT compared to when they are not using the device (Strayer et al., 2013). The increased RT is believed to result from a lower rate of information processing, perhaps because the DRT and cell phone share a limited pool of processing resources (Strayer, Watson, & Drews, 2011; Strayer et al., 2013). However, other causes are also possible. People could be more cautious in the DRT with increased cognitive load by setting a higher threshold for the amount of evidence needed to decide the light is present. Or people may require more time for processes other than the decision process (i.e., non-decision time), such as stimulus encoding or response production. We address the role of processing-rate, threshold, non-decision time, or some combination of these three, by fitting a cognitive model of the DRT task under conditions that vary in the load imposed by conversation. In the next section we outline the modeling framework applied to the DRT data. The data was collected from drivers performing a simulated driving task. Cognitive load was manipulated by having the driver converse with a passenger in person or with a person over a hands-free cell phone. These conditions were compared to a baseline where the driver took part in the simulator and DRT without any conversation.

Modeling the Detection Response Task

Sequential sampling models characterize responding as the result of a noisy process of accumulating evidence towards a response threshold. They have been extensively used to understand choice RT in terms of effects on evidence accumulation rate, response threshold, and non-decision time (e.g., Brown & Heathcote, 2008; Ratcliff & McKoon, 2008; Tillman, Benders, Brown, & van Ravenzwaaij, 2017; Tillman, Osth, van Ravenzwaaij, & Heathcote, 2017). Recently, sequential sampling models – and in particular the single-bound diffusion model (Heathcote, 2004;

Schwarz, 2001) – have been applied to simple RT data (i.e., data where participants make only one type of response) from a range of paradigms such as the psychomotor vigilance test and brightness detection tasks (Ratcliff & Van Dongen, 2011), simulated driving tasks (Ratcliff, 2015; Ratcliff & Strayer, 2014), go/no-go tasks (Heathcote, 2004; Schwarz, 2001), as well as pointing, picture naming and eye-movement tasks (Anders, Alario, & van Maanen, 2016). We collected simple RT data from the DRT ('press a key if you detect a light') and fit the single-bound diffusion model in order to investigate the causes underlying slowing due to increased cognitive load.

Figure 1 is a schematic of the single-bound diffusion model. The response threshold, 'a', quantifies the amount of evidence needed to make a response. On each trial, noisy evidence accumulates towards the response threshold at some rate – the drift rate. Within-trial (moment-to-moment) noise causes accumulation of evidence towards the threshold according to a Brownian motion. The Wald distribution (Wald, 1947) describes the first passage times for Brownian motion with positive drift rate toward a positive response threshold. When the threshold is crossed response production is triggered. The time it takes to reach the response threshold is the decision time. Non-decision time, T_{er}, is added to the decision time to make up the total observed RT, so simple RT is described by a shifted-Wald distribution, with a shift equal to the non-decision time.

Ratcliff & Van Dongen (2011; see also Ratcliff, 2015; Ratcliff & Strayer, 2014) fit an elaborated version of the single-bound diffusion model, where on each trial the drift rate is sampled from a normal distribution with mean v and standard deviation η. Including drift variability significantly improved their model fit to simple RT data. When the sampled drift rates are strictly positive, the resulting mixture of Wald distributions has an easily computed likelihood (see Equation 3 in Desmond & Yang, 2011). However, when the
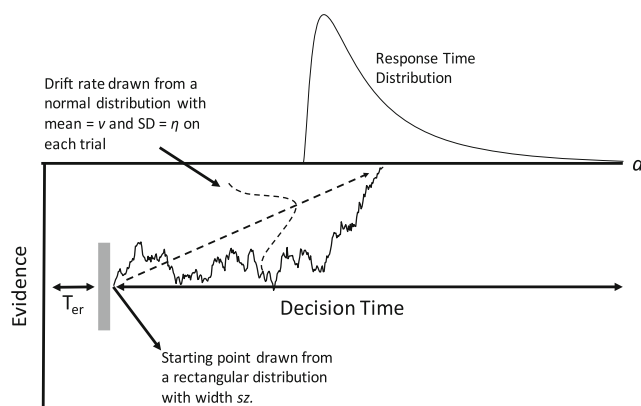

Fig. 1 The single-bound diffusion model and its parameter values: response boundary (a), mean drift rate (v), between-trial variability in drift rate (η), between trial variability in starting point (sz), and non-decision time (T_{er})

sampled drift rates can be negative the likelihood cannot be directly computed, and so, Ratcliff and Van Dongen resorted to simulation methods. They were interested in negative rates because they can result in the threshold never being crossed, and so can account for failures to respond, which were common in their application; simple RT data from sleep-deprived participants. The trial-to-trial rate variability gives the single-bound diffusion model more flexibility (Ratcliff & Van Dongen, 2011), yet also has a down side; it is only possible to identify two of the three parameters associated with evidence accumulation (i.e., the response threshold and the drift rate mean and standard deviation; see Ratcliff & Van Dongen, 2011).

We also fit a model that allowed the starting point of evidence accumulation to vary according to a uniform distribution with width sz.[1] Choice RT models commonly allow for start-point variability in order to account for fast errors (Ratcliff & Rouder, 1998), an issue that does not apply to simple RT. However, allowing for start-point variability may be as appropriate for simple RT tasks, due to the potential for "premature sampling" (Laming, 1968), which is the sampling of evidence before the onset of the stimulus. In choice RT, stimulus onset provides an appropriate signal to start sampling choice evidence (i.e., evidence that discriminates between different types of stimuli). In simple RT tasks, the presence of the stimulus is itself the evidence, and so, it is somewhat circular to assume that the onset of the stimulus can be used to trigger sampling of such evidence (i.e., stimulus detection cannot be used to trigger the detection process). In the context of the DRT task, where the only cue for a new stimulus is the inter-stimulus interval, continual or premature sampling during that interval seems likely.

In summary, we fit models without any between-trial variability, and with either trial-to-trial starting point variability or trial-to-trial rate variability. Because failures to respond were relatively rare in our data (less than 5% of all trials), we assumed that drift rate variability followed a normal distribution truncated below zero. The truncated normal assumption enabled the easy calculation of likelihoods, and consequently allowed us to use hierarchical Bayesian methods of estimation. This in turn allowed us to fit data sets with a relatively small number of observations per participant (200 per condition) based on the extra constraint afforded by hierarchical shrinkage effects (Shiffrin, Lee, Kim, & Wagenmakers, 2008), which improve estimation of each participant's parameters by constraining them using information about the parameters of all participants.

---

[1] Note that start point and threshold variability are mathematically indistinguishable in this model. We used the R code provided in association with Logan, Van Zandt, Verbruggen, & Wagenmakers (2014) to implement this model

The cognitive load effects of conversation

Established measures of cognitive capacity (Townsend & Eidels, 2011; Townsend & Nozawa, 1995) correspond with drift rates in choice RT tasks (Eidels et al., 2010). Increased cognitive load has been shown to have large effects on the tail of RT distributions in both choice (Shahar, Teodorescu, Usher, Pereg, & Meiran, 2014) and simple (Ratcliff & Strayer, 2014) RT tasks. Smaller drift rates and larger response thresholds are also known to lengthen the tail of RT distributions (Brown & Heathcote, 2008; Matzke & Wagenmakers, 2009; Ratcliff & McKoon, 2008; Usher & McClelland, 2001). Thus, it is tempting to conclude that increased cognitive load is related to and possibly even causes changes in drift rates and/or thresholds.

However, when researchers have used sequential sampling models to investigate cognitive load manipulations they have found that increased load affects a range of cognitive processes. Specifically, these studies found that increases in load either increase response thresholds (e.g., Heathcote, Loft, & Remington, 2015), trial-to-trial drift rate variability (McVay & Kane, 2012), non-decision times (Shahar et al., 2014), or decrease drift rates (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Sewell, Lilburn, & Smith, 2016). When the single-bound diffusion model with trial-to-trial rate variability was fit to data from a simulated driving task – where participants needed to press the brake to prevent a collision with a car in front – talking on a cell phone affected the drift rate and/or response threshold of drivers, but the effects could not be disentangled because of the aforementioned parameter identifiability issues (Ratcliff & Strayer, 2014).

To date, it is not clear how cognitive load imposed by passenger and cell phone conversation impacts the cognitive processes underpinning DRT performance. Our experiment investigated this issue by assigning pairs of participants to roles of a passenger or a driver in a high-fidelity driving simulator. Passengers were either seated next to the driver or in a separate room. In both cases they were instructed to converse casually with the driver, but to refrain from comments concerning the road. The latter stipulation aimed to remove a likely cause of the lack of DRT decrements noted by Drews et al. (2008) when the passenger was in the car; facilitation due to passenger-supplied warnings. However, other causes, such as timing of conversation to avoid conflict with safety critical events, may remain. Both driver and passenger were fitted with a DRT device (Strayer et al., 2013), as illustrated in Fig. 2. The driver was requested to drive as normal but also to respond quickly and accurately to the DRT signal when they detected the red light in their visual field.

Cognitive load was manipulated across three conditions for the driver: a baseline where they were driving alone with

Fig. 2 The DRT device used in the current study

no conversation, driving while talking with a passenger sitting next to them in the simulator, or driving while talking over a hands-free cell phone to a person in another room. We used a one-way Bayesian ANOVA (Morey, Rouder, & Jamil, 2014; Rouder, Morey, Speckman, & Province, 2012) to examine directly observed DRT performance. We hypothesized that drivers in the no conversation condition would respond more quickly to the DRT signal relative to driving while conversing with a passenger or over a cell phone. We also hypothesized that the decrements due to conversation could be larger with the hands-free cell phone relative to in-car conversation, but that this difference may be minimal due to our instruction to avoid comments about the driving task.

We then fit a set of single-bound diffusion models with different parameter settings instantiating different explanations of the effects of experimental manipulations in terms of drift rates, response thresholds, non-decision times, or any combination thereof. In addition, we compared model fits with and without between-trial drift rate variability or starting point variability. These competing explanations were compared based on the Watanabe-Akaike information criterion (WAIC) measure of out-of-sample prediction error (Gelman, Hwang, & Vehtari 2014; Watanabe, 2010). WAIC involves calculating a goodness-of-fit measure and subtracting a measure of model complexity from this value to approximate how well the model predicts future data. WAIC is similar to the deviance information criterion in this respect (DIC; Spiegelhalter, Best, Carlin, & van der Linde 2002), but is considered an improvement because it does not assume the posterior distribution is a normal distribution and it is calculated from each data point, which improves accuracy.

Method

Participants

Participants were undergraduates (40 total, 28 females) at the University of Utah. They had an average age of 23 years old and a driving experience ranging from 3 to 16 years with an average of 6.5 years. All participants had normal or corrected-to-normal visual acuity, a valid driver's license, and were fluent in English. All participants owned a cell phone and reported that they used it regularly while driving.

Stimuli and design

The DriveSafety^{TM} DS-600 simulator was used in this experiment. The DS-600 consists of a Ford Focus cab surrounded by three large screens encompassing a 270° view. The simulated vehicle is based on the vehicular dynamics of a compact passenger sedan with automatic transmission. The driving scenario was designed using DriveSafety HyperDrive Authoring Suite. A two-way, four-lane interstate highway scenario was designed for this experiment. The roadway has four straight sections (10 miles each) connected by two wide-radius curves (1 mile each).

Both the driver and passenger were fitted with the DRT. The light diode was positioned an average 15° to the left and 7.5° above the participant's left eye and was held in a fixed position on the head with a headband (see Fig. 2 again). RT to the DRT signal was recorded with millisecond accuracy via a button attached to participant's left thumb and encompassed the time between stimulus presentation and response. Participants were instructed to press the microswitch, attached to their left thumb, every time they detected the light presented by the their headset. The participants were steering the simulated vehicle, so they pressed the button against the steering wheel. The DRT protocol for each device followed the specifications outlined in International Organization for Standardization (2015).

There were a total of three within-subject conditions for each pair of participants, which were counterbalanced using a quasi Latin Square design. Each condition lasted approximately 14 minutes and the order of these conditions was counterbalanced across participants. For drivers these were: (1) single (driving only) task, where drivers drove the simulated car and responded to the red light, but were not engaged in any type of conversation; (2) Dual task 'passenger' – driving and conversing with a passenger seated next to the driver; (3) Dual task 'cell phone' – driving and conversing, through a cell phone, with another person seated in a separate room.

Procedure

The passenger and driver were peer students assigned randomly to the driver and passenger roles when they arrived at the laboratory. Participants were instructed to have a natural conversation as they would in real life; no restrictions about the topics covered in the conversation were provided to them, except they could not discuss the DRT. The DRT was used to assess the workload of the driver (and passenger) and it is not intended to mimic real-world aspects of driving (e.g., traffic lights).

Participants drove on a simulated multi-lane freeway with moderate traffic, which had approximately 1500 vehicles in each lane per hour. Participants were given a five-minute practice session to familiarize themselves with the driving simulator. In each of the conditions, except for condition 1, the drivers and passengers were asked to speak and listen in equal proportions (i.e., 50% speaking and 50% listening). In condition 3 the driver and passenger initiated a call via a hands-free Bluetooth earpiece. Participants were allowed to adjust the volume on the cell phone so that they could clearly hear the conversation.

The DRT task presented red lights every three to five seconds via the head mounted device, where the presentation times followed a rectangular distribution bounded between 3 and 5 seconds. This resulted in 200 trials per condition for each participant. The light signal remained illuminated until a response was made or 1000 msec had elapsed.

Mean RT analysis

All RTs below 250ms (0.02%) were discarded, as were all RTs slower than 1000ms (0.2%). There were total of 22 false alarms, approximately .001% of all trials. In the driver only condition, the miss rate was 0%; for the driver and passenger in car condition, the miss rate was 4.7%; for the driver and cell phone condition, the miss rate was 3.5%. Both misses and false alarms were excluded from analysis. Using the R programming language (R Development Core Team, 2016) and 'Bayesfactor' package (Morey et al., 2014), RTs were analyzed with a one-way JZS Bayesian ANOVA (Morey et al., 2014; Rouder et al., 2012), with a default setting for Cauchy priors and with subjects included as a random effect. The cognitive load manipulation was included as a main effect with levels – driver responding only (D), driver responding while talking with a passenger in the car (DP), driver responding while talking to the passenger on a cell phone (DC). The mean RTs (in seconds) in each condition were D = .466 (SD = .135), DP = .502 (SD = .140), and DC = .505 (SD = .140).

We tested the main effect model against a null model which suggests no main effect on RT, reporting Bayes factors ($BF10$), which quantify evidence in favor of the main effect model over the no main effect model as a ratio. For example, when $BF10 = 5$ the observed data make the main effect model 5 times more likely than the no main effect model. When $BF10 = 1/5 = .2$ the observed data increase the likelihood of the no main effect model by a factor of 5 relative to the main effect model. The Bayes factor ANOVA revealed that the cognitive load main effect model was preferred to the null model by a Bayes factor of 257. Thus the data provide strong evidence (Kass and Raftery, 1995) against the hypothesis of no main effect on RT.

We conducted post-hoc Bayesian paired-samples t-tests to examine which conditions differed from each other, with detailed results reported in Fig. 3. There was strong evidence that the driver's responses to the DRT signal were slower when conversing (DP vs. D and DC vs. D), indicating the DRT was sensitive to the additional cognitive load. There was positive evidence for no difference in the driver's RT as a function of the passenger's location, in the passenger seat or another room (DP vs. DC).
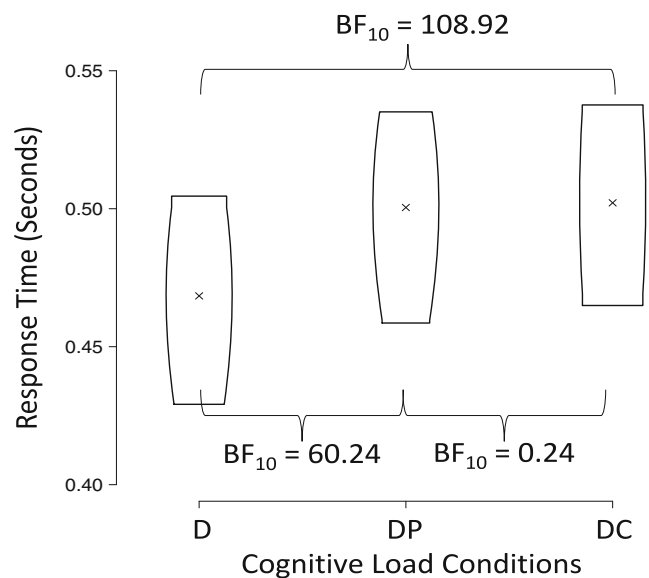

Fig. 3 Violin plots of predicted response time data from the one-way JZS Bayesian ANOVA. Violin plots include an ×, which marks the predicted median RT, and mirrored on either side are rotated kernel density plots of the 95% highest density interval of each posterior distribution. Superimposed are Bayes factors from post-hoc paired-samples t-tests. The three cognitive load conditions were driver responding only (D), driver responding with a passenger in the car (DP), and driver responding while talking to the passenger on a cell phone (DC)

Model-based analysis

Data, fitting code, and details about the hierarchical Bayesian model fitting routine are presented in the supplementary materials, which is available online at https://osf.io/d7eru. To select between competing models we measured how well each model could predict future data using WAIC. WAIC includes a goodness of fit value and a measure of the model's complexity. Model complexity is subtracted from the goodness-of-fit measure to approximate an unbiased estimate of the model's out-of-sample prediction error. When comparing models, the model with the lower WAIC value is better able to predict future data.

We fit 22 separate single bound diffusion models to the DRT data, 7 models with trial-to-trial rate variability (η) fixed at zero, 7 models where we estimated η, and 7 models where we estimated sz. For reference, we also fit a null model with no variability and no parameters free to vary across conditions. Table 1 provides a measure of the best fit for each model and a measure of each model's complexity, the effective number of parameters. The WAIC analysis indicates a clear preference for allowing both response threshold and non-decision time to vary over cognitive load conditions, and this was consistent regardless of the trial-to-trial variability assumption. Most importantly, models in which the mean drift rate explained the effect of the experimental factor were strongly rejected. We refer to the response threshold and non-decision time model with trial-to-trial starting point variability as the winning or $a + T_{er}$ model and we will focus further analysis on this model. Plots presented in the supplementary materials show that the $a + T_{er}$ model captured all response time distributions well. In addition, the model only predicted that 0.2% of trials were misses (i.e., greater than 1 second).

The winning $a + T_{er}$ model contained between-trial starting point variability. This suggests that accumulation of information that occurs without a signal present has non-negligible effects on the response time data. In particular, it appears likely the accumulation process does not always reset to a baseline after a response before the presentation of the next DRT signal. The starting point variability captures this, likely reflecting uncertainty caused by varying inter-trial-intervals of 3–5 seconds.

Table 2 shows median values of the group-level mean posterior distributions for response threshold, drift rate, and non-decision time from the winning $a + T_{er}$ model. We used Bayesian predictive p-values to statistically test for differences between the posterior distributions (Meng, 1994). We calculated the difference between subject level posterior distributions, or plausible values (Marsman, Maris, Bechger, & Glas, 2016), and then averaged the differences over subjects.

We calculated the probability (p-value) that the distributions of differences between parameters were equal to or less than 0. Similar to the traditional p-value, a low predictive p-value indicates a low probability of the model producing this or more extreme data. The response threshold increased from the D to the DP condition ($p < .001$) and from the D to the DC condition ($p < .001$), suggesting that drivers were more cautious when cognitive load increases. Thresholds were comparable between the DP and DC conditions ($p = .762$), indicating little evidence for a difference in caution between cell phone and in-car conversations. Non-decision time for drivers conversing with a passenger in person was only 6ms faster, on average over posterior estimates, compared to when they were talking over the phone or were driving alone. There was no strong statistical evidence that any non-decision time parameters were different from each other at the group-level (all $p > .215$). In Fig. 1

Table 1 WAIC results, number of effective parameters and deviance of the posterior mean for all models

Model	Drift variability		No between variability		Start point variability	
	WAIC	Effective parameters	WAIC	Effective parameters	WAIC	Effective parameters
$a \sim 1$ & $v \sim 1$ & $t_{er} \sim 1$	–	–	−11276	46	–	–
$a \sim F$ & $v \sim 1$ & $t_{er} \sim 1$	−11628	88	−11603	93	−11717	102
$a \sim 1$ & $v \sim F$ & $t_{er} \sim 1$	−11563	89	−11537	86	−11537	85
$a \sim 1$ & $v \sim 1$ & $t_{er} \sim F$	−11565	86	−11534	90	−11573	89
$a \sim F$ & $v \sim 1$ & $t_{er} \sim F$	−11652	116	−11635	116	**−11748**	122
$a \sim 1$ & $v \sim F$ & $t_{er} \sim F$	−11596	128	−11592	122	−11621	119
$a \sim F$ & $v \sim F$ & $t_{er} \sim 1$	−11624	124	−11615	123	−11715	131
$a \sim F$ & $v \sim F$ & $t_{er} \sim F$	−11625	138	−11623	137	−11729	141

Note. Bold WAIC value indicates the preferred model. The first row displays a null model with no parameters free to vary across conditions

$\sim$F indicates different parameter estimates were allowed for each level of the experimental factor

$\sim$1 that the same value was estimated for all levels

Table 2 Median values of the group-level mean posterior distributions from the winning $a + T_{er}$ model. 95% credible interval is presented in parenthesis

	D	DP	DC
a	1.243 (0.475, 1.998)	1.505 (0.935, 2.224)	1.478 (0.938, 2.171)
T_{er}	0.210 (0.193, 0.227)	0.204 (0.180, 0.227)	0.210 (0.187, 0.232)
v	5.052 (4.733, 5.371)	5.052 (4.733, 5.371)	5.052 (4.733, 5.371)
sz	0.389 (0.021, 0.921)	0.389 (0.021, 0.921)	0.389 (0.021, 0.921)

Note. T_{er} values are in seconds and the diffusion noise is fixed to 1

of the supplementary materials we show that despite their being no differences in non-decision time at the group level there are substantial individual effects. Therefore, model selection supports the inclusion of non-decision time effects because there are large differences between the individual drivers.

General discussion

In our experiment, pairs of participants, assigned as either driver or passenger, took part in a driving simulator and detection response task (DRT) simultaneously. In the DRT, participants were required to respond when a small light appeared in their peripheral vision. The drivers completed the DRT in three different conditions: by themselves in the simulator, talking with a passenger in the simulator, and talking to a passenger (who was outside of the simulator) on a cell phone. We recorded the response times (RT) in the DRT from both the driver and passenger.

RTs in the DRT are a validated measure of cognitive load (International Organization for Standardization, 2015), where slower RTs represent increased cognitive load. We found that drivers had slower RTs when they were conversing with a passenger in person or over the phone compared to when they were by themselves – suggesting that both types of conversation increased cognitive load. We modeled the DRT behavioral data with the single-bound diffusion model to determine if longer RTs, which reflect increased cognitive load, are due to differences in drift rates, response thresholds, or non-decision times.

We found that the cognitive load effect on DRT performance was due to an increase in the participant's response thresholds, rather than an effect of cognitive load on the time to encode stimuli and to produce responses or on the rate of evidence accumulation. In contrast to Ratcliff and Van Dongen (2011), but consistent with Anders et al. (2016), we did not find it necessary to allow for variability in the rate of evidence accumulation from trial-to-trial to provide a good account of our DRT data.

Our findings may at first seem surprising because they are not in line with the capacity sharing account of DRT and driving performance (e.g., Strayer et al., 2011, 2013). However, separate pools of capacity for DRT and driving is consistent with the finding that having to preform the DRT does not adversely impact driving (Strayer et al., 2014). Why then is the DRT a sensitive measure of cognitive load? We propose this could be due to the general tendency of people to be more cautious under increased cognitive load, but further work is required to better understand the processes underlying threshold adjustments, and why they occur.

One possibility is that the process is consciously mediated, with participants slowing in the DRT because they deliberately set higher threshold for the secondary DRT task when they perceive they are subject to a higher workload in the primary driving task. In other words, if the driver has a separate resource pool for the primary and secondary tasks then the DRT slowing is the result of an strategic increase in threshold. The strategy could be that the driver prioritizes responding to traffic over responding to the DRT signal when they perceive a higher workload. This possibility is consistent with the strong correlation found between DRT decrements and self-report measures of subjective workload (Strayer et al., 2013), such as the NASA Task Load Index (Hart and Staveland, 1988).

Alternately, threshold increases may occur to reduce the chance of response conflicts (i.e., one response preempting another), as suggested by *delay theory* (Loft & Remington, 2013) of dual task costs in prospective memory tasks (Heathcote et al., 2015). Prospective memory tasks involve a dual task paradigm that requires participants to respond to both a common "ongoing" task and an occasional prospective memory task. The ongoing activity can preempt the prospective memory response. Heathcote et al. found that participants raised their threshold in order to avoid preempting.

Many cognitive tasks require a decision between two or more alternatives and record both the choice and the time to make that choice. In such choice RT data, threshold and rate effects are relatively easy to disambiguate as they have opposite effects on accuracy and RT (i.e., a higher threshold increases accuracy and RT, whereas a higher drift rate increases accuracy but decreases RT). In simple RT, in contrast, these effects are differentiated only by relatively subtle effects on the distribution of RT. Although tests based on out-of-sample predictive accuracy clearly favored a threshold account of cognitive load effects, and the corresponding model produced clear and sensible effects on threshold estimates, it would be prudent in future work to seek converging evidence about our somewhat surprising findings.

One potential way forward is to examine the effects of speed vs. accuracy instructions, which are usually assumed

to selectively affect response thresholds (but see Rae, Heathcote, Donkin, Averell, & Brown, 2014). Another possibility is to compare cognitive load effects on the traditional (simple RT) DRT and a version requiring a choice response, which should allow for a stronger comparison of rate vs. threshold models. A choice version of the DRT (e.g., respond 'A' for a green light, 'B' for a red light) could offer a stronger test of cognitive load effects, although in practice it may be problematic if its greater complexity results in a detrimental impact on driving performance. The test could be stronger because, in choice data, drift rates and response thresholds have unique behavioral signatures in data that can be easily distinguished (Ratcliff & McKoon, 2008). Changes in the mean drift rate will cause small changes to the fastest responses, large changes to the slowest responses, and higher rates will lead to faster response times and higher accuracy. In contrast, response thresholds have relatively small effects on the fastest and slowest responses and higher thresholds will lead to slower response times and higher accuracy. By modeling these unique signatures in choice data, it may be possible to better distinguish rate and thresholds theories of cognitive load, and so provide a more robust test for our novel hypothesis about caution changes in response to cognitive load.

References

Anders, R., Alario, F.X., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological methods*, *21*(3), 309.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

Desmond, A., & Yang, Z. (2011). Score tests for inverse gaussian mixtures. *Applied Stochastic Models in Business and Industry*, *27*(6), 633–648.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S., Sudweeks, J., ..., & Gupta, S. (2006). *The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment* (Tech. Rep.). National Highway and Traffic Safety Administration.

Drews, F.A., Pasupathi, M., & Strayer, D.L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, *14*(4), 392–401.

Eidels, A., Donkin, C., Brown, S.D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*(6), 763–771.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Hart, S.G., & Staveland, L.E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Heathcote, A. (2004). Fitting Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package. *Behavior Research Methods*, *36*, 678–694.

Heathcote, A., Loft, S., & Remington, R.W. (2015). Slow down and remember to remember! a delay theory of prospective memory costs. *Psychological Review*, *122*(2), 376–410.

International Organization for Standardization (2015). *Road vehicles – transport information and control systems – detection-response task (drt) for assessing attentional effects of cognitive load in driving*. (ISO WD 17488).

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Laming, D.R.J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Loft, S., & Remington, R.W. (2013). Wait a second: Brief delays in responding reduce focality effects in event-based prospective memory. *The Quarterly Journal of Experimental Psychology*, *66*(7), 1432–1447.

Logan, G.D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, *121*(1), 66–95.

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2016). What can we learn from plausible values? *Psychometrika*, *81*(2), 274–289.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817.

McVay, J.C., & Kane, M.J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 525–549.

Meng, X.L. (1994). Posterior predictive *p*-values. *The Annals of Statistics*, *22*, 1142–1160.

Morey, R., Rouder, J., & Jamil, T. (2014). Bayesfactor: Computation of Bayes factors for common designs. *R package version 0.9, 8*.

R Development Core Team (2016). The r project for statistical computing [Computer software manual]. Vienna, Austria.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(5), 1226–1243.

Ranney, T.A., Mazzae, E., Garrott, R., & Goodman, M.J. (2000). Nhtsa driver distraction research: Past, present, and future. In *Driver distraction internet forum* (Vol. 2000).

Ratcliff, R. (2015). Modeling one-choice and two-choice driving tasks. *Attention, Perception, & Psychophysics*, *77*(6), 2134–2144.

Ratcliff, R., & McKoon, G.M. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic Bulletin & Review*, *21*(3), 577–589.

Ratcliff, R., & Van Dongen, H.P. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*(27), 11285–11290.

Rouder, J.N., Morey, R.D., Speckman, P.L., & Province, J.M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374.

Rueda-Domingo, T., Lardelli-Claret, P., de Dios Luna-del Castillo, J., Jiménez-Moleón, J.J., Garcıa-Martın, M., & Bueno-Cavanillas, A. (2004). The influence of passengers on the risk of the driver causing a car collision in spain: Analysis of collisions from 1990 to 1999. *Accident Analysis & Prevention*, *36*(3), 481–489.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W.W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414–429.

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods*, *33*(4), 457–469.

Sewell, D.K., Lilburn, S.D., & Smith, P.L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of experimental psychology: learning, memory, and cognition*, *42*(11), 1673.

Shahar, N., Teodorescu, A.R., Usher, M., Pereg, M., & Meiran, N. (2014). Selective influence of working memory load on exceptionally slow reaction times. *Journal of Experimental Psychology: General*, *143*(5), 1837–1860.

Shiffrin, R.M., Lee, M.D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *64*, 583–639.

Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile. *AAA Foundation for Traffic Safety*.

Strayer, D.L., Drews, F.A., & Johnston, W.A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*(1), 23–32.

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., Coleman, J., Ortiz, E., & Cooper, J.M. (2014). Measuring cognitive distraction in the automobile: Ii. assessing in-vehicle voice-based interactive technologies. *AAA Foundation for Traffic Safety*. Retrieved from https://www.aaafoundation.org/sites/default/files/MeasuringCognitiveDistractions.pdf

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.

Strayer, D.L., Watson, J.M., & Drews, F.A. (2011). Cognitive distraction while multitasking in the automobile. *Psychology of Learning and Motivation-Advances in Research and Theory*, *54*, 29.

Sussman, E.D., Bishop, H., Madnick, B., & Walter, R. (1985). Driver inattention and highway safety. *Transportation Research Record*, *1047*, 40–48.

Tillman, G., Benders, T., Brown, S.D., & van Ravenzwaaij, D. (2017). An evidence accumulation model of acoustic cue weighting in vowel perception. *Journal of Phonetics*, *61*, 1–12.

Tillman, G., Osth, A., van Ravenzwaaij, D., & Heathcote, A. (2017). A Diffusion Decision Model Analysis of Evidence Variability in the Lexical Decision Task. Advance online publication.

Townsend, J.T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin and Review*, *18*(4), 659–681.

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.

Usher, M., & McClelland, J.L. (2001). The time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Vollrath, M., Meilinger, T., & Krüger, H.-P. (2002). How the presence of passengers influences the risk of a collision with another vehicle. *Accident Analysis & Prevention*, *34*(5), 649–654.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Wang, J.-S., Knipling, R.R., & Goodman, M.J. (1996). The role of driver inattention in crashes: New statistics from the 1995 crashworthiness data system. In *40th annual proceedings of the association for the advancement of automotive medicine* (Vol. 377, p. 392).

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, *11*, 3571–3594.