The Residual Costs of Multitasking: Causing Trouble down the Road

Jonna Turrill, James R. Coleman, Rachel J. Hopman, Joel M. Cooper, & David L. Strayer
University of Utah

Objective: The cognitive workload of three Smartphone Digital Assistants (SDA) was manipulated in an on-off manner while participants drove an instrumented vehicle in order to measure the costs associated with intermittent dual tasking. **Background**: Previous research has shown costs in productivity when switching between two discrete tasks; however, similar costs have not yet been examined using intermittent, continuous dual tasks. **Methods**: Participating drivers completed 5 conditions: baseline driving, 3 SDA conditions, and a cognitively demanding math-memory operation span (OSPAN) task, each while responding to Detection Response Task (DRT) stimuli. Within the SDA conditions, on- and off-task DRT performance was compared to baseline driving and to the OSPAN task performance. **Results**: The on-task periods of the SDA conditions resulted in similar RTs as the OSPAN condition, while the off-task periods did not immediately return to baseline driving performance. Post hoc analyses of the on-off transitions within the SDA conditions revealed a gradual return to baseline driving performance over the course of 18 seconds. **Conclusions**: The delays in returning to baseline driving performance after completing a secondary task raise concerns about the usage of in-vehicle devices while driving as the effects of the delays last beyond the cessation of the SDA interaction.

INTRODUCTION

Individuals believe themselves to be more productive when they multitask. However, research suggests that people experience productivity time costs every time they switch focus to a different task compared to completing one task at a time (Cardosa-Leite, Green, & Bavelier, 2015; Sanbonmatsu, Strayer, Medeiros-Ward, & Watson, 2013). While in many settings multitasking is merely wasting time, in other contexts, this behavior can have important safety repercussions (Janssen, Gould, Li, Brumby, & Cox, 2015).

Previous studies have examined the performance decrements associated with drivers' engaging in secondary, non-driving related tasks (e.g., Harbluk, Noy, Trbovich, & Eizenman, 2007; Rantanen & Goldberg, 1999; Recarte & Nunes, 2003; Reimer, 2009; Strayer & Drews, 2007; Strayer et al., 2015; Strayer, Watson, & Drews, 2011). Specifically, Strayer and colleagues (Strayer et al., 2015; Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014) created a benchmark rating scale for common cognitive in-vehicle secondary tasks.

Using Detection Response Task (DRT) reaction time (RT) measurements, Strayer et al. (2014, 2015) assessed the workload associated with these common tasks. Because participants executed each of these studies' secondary tasks continuously, averaging RTs within each condition provided an accurate measurement of a task's cognitive workload. However, the averaging approach becomes less appropriate when applied to secondary tasks performed in an on- and off-task manner, such as seen when drivers use In-Vehicle Infotainment Systems (IVIS) or interact with Smartphone Digital Assistants (SDA).

Previous work in task switching using discrete laboratory tasks has shown that the residual switch cost dissipated after a single performance of the switched-to task and performance returned to near-optimal levels by the second task trial (i.e., a repeat trial) in a step-like function (Monsell, Sumner, & Waters, 2003; Rogers & Monsell, 1995). By adding markers to DRT data when a driver was on- or off-task with an in-vehicle device, changes in performance could be noted. If intermittent continuous dual tasks demonstrated residual switch costs similar to those found in discrete task switching, then RT performance would be expected to display a step-like function in workload as the secondary task was introduced. That is, RT would reflect a temporary increase in workload as the in-vehicle device task was performed that should then return to single-task driving workload once the interaction ended.

METHODS

Participants

Sixty-five (30 females) total participants ranging in age from 21 to 68 years ($\bar{x}$ = 41.3) participated in this study. Thirty-one participants completed the first phase, and an additional 34 participants completed the second study phase. All reported normal neurological functioning, normal or corrected-to-normal visual acuity, normal color vision (Ishihara, 1993), had a valid driver's license, and were fluent in English. Years of driving experience ranged from 4 to 52 years ($\bar{x}$ = 25.5). The University of Utah conducted a motor vehicle record report, required drivers to complete a defensive driving online training, and pass a certification test. All participants owned a smartphone and 66% reported using their phone regularly while driving.

Equipment

The study utilized three popular SDA engineered by Apple, Google, and Microsoft: Apple iPhone 6 with iOS 8.2 (Build 12D508) providing the Siri digital assistant, Google Nexus with Android 5.0.1 (Build LRby22C) providing the Google Now digital assistant, and Nokia Lumia 635 with Windows 8.1 (O.S. Version 8.10.12400.899) providing the

Cortana digital assistant, respectively. Identical music libraries and contacts were loaded onto each smartphone. Each SDA was capable of voice-initiated contact calling, smartphone number dialing, music selection, and voice-dictated text messaging.

Apple "EarPods with Remote and Mic" was attached to each smartphone with the right earpiece placed in the participant's ear. Participants pressed a small button attached to the headphones' cord to activate / deactivate each SDA. The identical headphones ensured that differences between smartphones were not related to audio quality, microphone sensitivity, or other aspects of the physical interface. The smartphones were securely attached to the center console, to the right of the steering wheel, with a universal suction mount.

A 2015 Chevy Malibu and Chrysler 200c with automatic transmissions were used. Participants became familiar with the handling of the vehicle by driving for approximately 20 minutes prior to the experiment.

Participants wore a head-mounted Detection Response Task (DRT) device (mfg. Precision Driving Research), which consists of a soft headband with an LED light positioned on a flexible arm. Participants responded to the red LED light by depressing a microswitch attached to their left thumb. The LED light flashed aperiodically between 3-5 seconds, remained on for 1000 msec unless the participant responded, and was placed in the periphery of the left eye, 15° to the left and 7.5° above the participant's pupil (ISO WD 17488, 2015).

Procedure and Tasks

A researcher was assigned to ride with each participant for the duration of the study to precisely administer the research procedure. Additionally, the researchers ensured the safety of the driver, provided in-car training, and delivered task cues to participants.

For each condition, participants drove an 8 minute, 4.3 km loop in a residential neighborhood with 7 all-way stop signs, 1 two-way stop, and 2 stop lights.

Throughout each of the five conditions, participants responded to the DRT stimuli. The DRT software recorded participants' RT data for every light-stimulus onset. On- and off-task behavior could be compared by using the RT markers indicating when the participant was on-task within an SDA interaction.

Participants completed five conditions in two between-subject phases. Participants in the first phase drove without a secondary in-vehicle task (1. single task) throughout the driving loop. For each smartphone condition (conditions 2-4), participants completed 6 tasks during the driving loop, consisting of calling contacts twice, dialing familiar numbers twice, and selecting music twice. Lastly, participants drove while completing an auditory math-memory operation span task (5. OSPAN) to induce a high workload baseline (Watson & Strayer, 2010).

Participants in the second phase of the study also completed the five conditions, but instead of voice-initiated calling, dialing, or music selection with each SDA, the participants dictated one sentence text messages in response to prompts (e.g., *"Tell Amy Smith you're running late. Ask her to start dinner."*). Participants practiced dictating correctly the text recipient and key content of the text message, although they were free to select their own phrasing to convey the message meaning.

In both phases of the study during SDA conditions, the researcher cued participants with 6 prompts at specific intersection locations (e.g., "Dial your own number"). Upon hearing the prompt, participants initiated the SDA interaction when ready. Participants pressed the remote button on the earbud to activate the smartphone SDA and gave verbal commands to the SDA to complete the desired task (e.g., "Call Amy Smith"). Participants then pressed the remote button to end the interaction. Thus, participants drove while interacting with the SDA system in an intermittent fashion: Once participants completed the interaction, they would be "just driving" until they reached the next cued intersection.

Regardless of the specific interactions with the SDA, participants were either "on" or "off" task within SDA conditions while driving.

RESULTS

Following ISO standards, a DRT response was considered a Hit if the RT was between 100 – 2500 msec and a Miss if after 2500 msec (ISO DIS 17488, 2015). Hit Rate, or accuracy, was calculated by dividing the number of Hits by the total number of presented stimuli.

By collapsing between the two phases and across the three smartphone SDA conditions, the off- and on-task performance measures were more accurately compared. SDA-0 reflects performance measures from the off-task portions of the SDA condition while SDA-1 reflects on-task performance.

DRT RT and Hit Rate were analyzed using a repeated measures ANOVA with four levels of workload to match Single-Task, SDA-0, SDA-1, and OSPAN. There was a main effect of workload for RT, $F(3, 192) = 178.89$, $p < .001$, partial $\eta^2 = .74$. Similarly, there was a main effect of workload for Hit Rate, $F(3, 192) = 33.08$, $p < .001$, partial $\eta^2 = .34$.

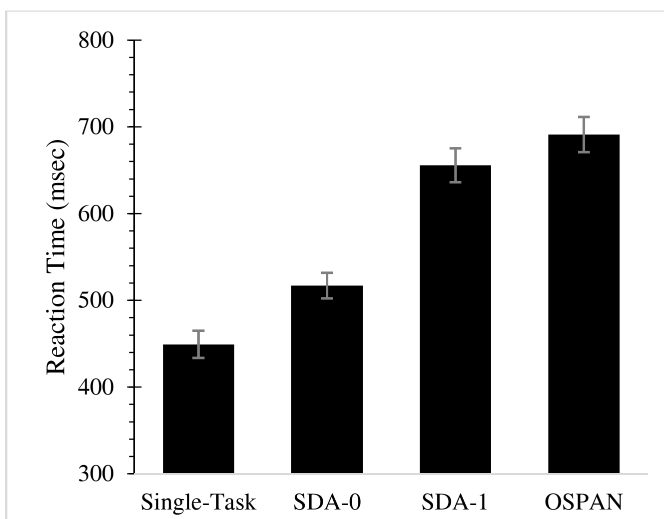

Figure 1. Mean DRT RT (in msec) for Single-Task, off-task SDA, on-task SDA, and OSPAN performance. Error bars reflect standard error of the mean.

Figures 1 and 2 display the RT and the Hit Rates across each condition. Participants engaged in SDA interactions (on-task) experienced increased RTs similar to the OSPAN condition. Additionally, once participants finished the voice-controlled SDA interaction, their RT and accuracy did not immediately return to single-task baseline performance even though, at that point, they were just driving. These increased RTs could be seen as the residual cost of task switching (see Figure 3).

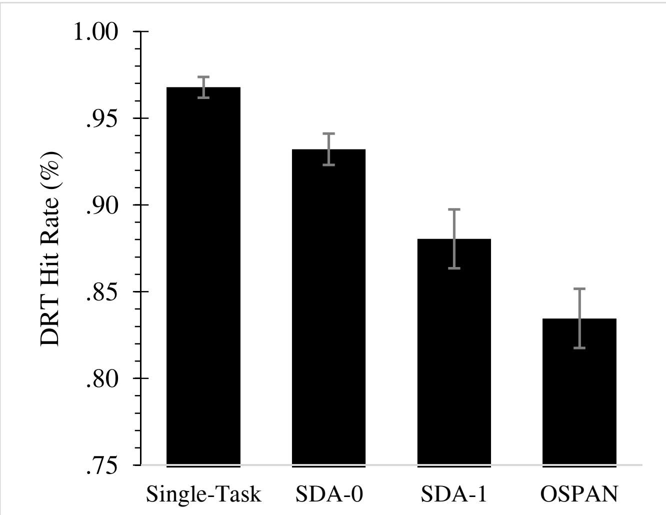

Figure 2. Mean DRT Hit Rate for Single-Task, off-task SDA, on-task SDA, and OSPAN performance. Error bars reflect standard error of the mean.

Post-hoc analyses were performed to examine the transitions immediately following the cessation of the SDA interactions. For these analyses, the off-task RTs were sorted into 3 second bins up to 24 seconds off-task. These bins began at the task-offset timestamp. For example, a DRT stimulus onset occurring 8 seconds after cessation of the SDA interaction would be sorted into the third bin (which reflects the average RT of DRT stimuli between 7 and 9 seconds).

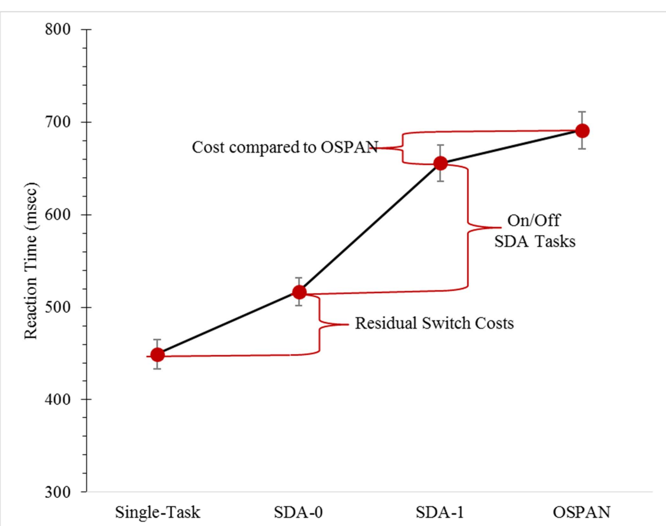

Figure 3. Residual switch costs as compared to Single-Task, OSPAN, and between on- and off-task SDA performance.

In this manner, changes in the average RT over time could be visualized. As seen in Figure 4, RT gradually returned to baseline driving performance over the course of an 18 second interval.

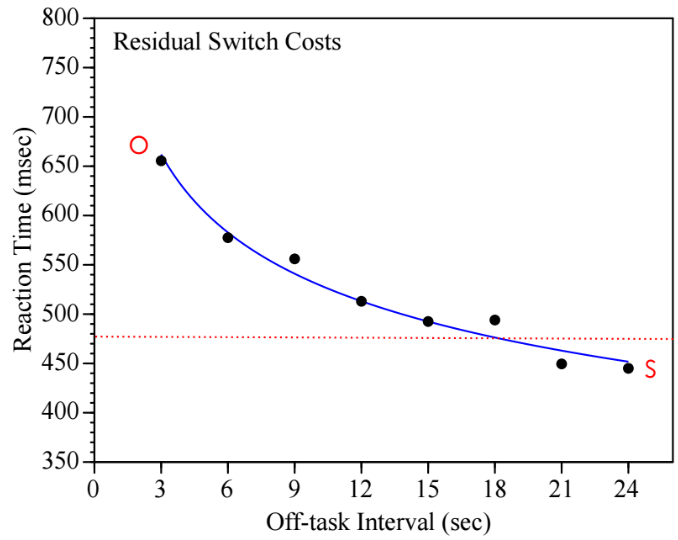

Figure 4. Residual switch cost curve in transitioning from on- to off- dual task performance. The red "O" indicates average OSPAN RT; the red "S" indicates the average single-task RT. Off-task performance is distributed into 3 second interval bins relative to the cessation of the SDA interaction. The blue line represents the best fitting power function relating the transition from on-task to single-task levels of performance. The dotted red line represents the critical *t*-value for significant differences from the single-task condition. From the figure, residual switch costs are significantly different from the single-task baseline up to 18 seconds after the SDA interaction had terminated.

DISCUSSION

Our study demonstrated that SDA interactions performed in an on-off manner (e.g., driving, placing a call, and then ending the call) demonstrated residual costs similar to those found in attention switching literature from laboratory-based studies. For example, our study revealed average residual switch costs of 68 msec between single-task and SDA-0, and an average of 206 msec between SDA-0 and SDA-1 (see Figure 3). Similarly, Rogers and Monsell (1995) found switch costs ranging from 50 msec up to 300 msec with a variety of discrete tasks (as cited in Pashler, Johnston, & Ruthruff, 2001). Participants experienced high levels of cognitive workload when engaged with an SDA which cautions engagement with such secondary, non-driving related tasks. Moreover, the participants remained impaired when that interaction had ceased.

Upon dual-task cessation, the participants' performance did not immediately return to baseline levels. Thus, just because a driver has terminated a call or completed a voice-dictated text does not mean that they are no longer impaired. By examining the temporal changes occurring immediately upon off-task SDA interactions, the duration of the residual switch cost impairment was greater than those discovered within traditional laboratory based settings.

Increased performance demands of SDA interactions take longer than expected to dissipate, as shown by the residual switch cost curve in Figure 4. The residual curve reflects the additive effects of single-task performance and the lingering costs associated with the voice-based SDA interactions. The actual time to complete the SDA interaction was approximately 30 seconds, which makes the 18 seconds for the residual cost to subside a sizeable effect by comparison.

Understanding the causal foundations for the residual switch cost curves can inform the settings in which they might occur, most significant in safety-relevant situations. The curve could reflect two different processes, or a combination of both. First, the curve may be the result of disengaging from the cognitive processing associated with the SDA task. Known as Task Set Inertia (TSI; Evans, Herron, & Wilding, 2015), this proactive interference from the goals of the previous task is believed to decay slowly over time causing interference with the efficient performance of the other task (Allport et al., 1994, p. 421; Altmann & Gray, 2008). The presence of the residual switch cost curves from our driving study suggests a role for TSI effects in complex intermittent dual tasks.

Alternatively, Situation Awareness Recovery (SAR) processes could be initiating. Situation Awareness (SA) is the cognizance of one's environment over time for action (Endsley, 2000). When applied to driving, SA can be broken down into a series of cognitive processes (see SPIDER framework, Strayer & Fisher, 2015). When SA is lost, drivers must *recover* situation awareness of their surroundings in order to respond to changes. Gartenberg and colleagues (2014) measured the behavioral effects when participants had to recover SA of a complex, dynamic environment after being interrupted. Their research identified distinct changes in eye movement behaviors in addition to delays in making primary task decisions.

In summary, the residual switch cost curve indicates that cessation of verbal interactions with in-vehicle technologies does not assure that the driver is unimpaired, whether because of the lingering effects of the SDA task or because of the recovery processes that must be engaged to repair SA of the driving environment. Future research can explore these two proposed causal explanations in detail.

ACKNOWLEDGEMENTS

This work was supported by the AAA Foundation for Traffic Safety (AAA FTS).

REFERENCES

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà, M. Moscovitch, C. Umiltà, M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA, US: The MIT Press.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. Psychological Review, 115(3), 602-639. doi:10.1037/0033-295X.115.3.602

Cardoso-Leite, P., Green, C. S., & Bavelier, D. (2015). On the impact of new technologies on multitasking. *Developmental Review*, *35*98-112. doi:10.1016/j.dr.2014.12.001

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), Situation awareness analysis and measurement (pp. 3–28). Mahwah, NJ: Lawrence Erlbaum.

Evans, L. H., Herron, J. E., & Wilding, E. L. (2015). Direct Real-Time Neural Evidence for Task-Set Inertia. *Psychological Science (Sage Publications Inc.)*, *26*(3), 284-290. doi:10.1177/0956797614561799

Gartenberg, D., Breslow, L., McCurry, J. M., & Trafton, J. G. (2014). Situation Awareness Recovery. *Human Factors*, *56*(4), 710-727. doi:10.1177/0018720813506223

Harbluk, J. L., Noy, Y., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, *39*, 372-379. doi:10.1016/j.aap.2006.08.013

Ishihara, S. (1993). Ishihara's test for color-blindness. Tokyo: Kanehara.

ISO DIS 17488 (2015). Road Vehicles -Transport information and control systems - Detection Response Task (DRT) for assessing selective attention in driving. Draft International Standard, ISO TC 22/SC39/WG8.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327-342. doi:10.3758/BF03194391

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*629-651. doi:10.1146/annurev.psych.52.1.629

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*, 119-137. doi:10.1037/1076-898X.9.2.119

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 13–19.

Rogers, R. D., & Monsell, S. D. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology. General*, *124*(2), 207.

Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who Multi-Tasks and Why? Multi-Tasking Ability, Perceived Multi-Tasking Ability, Impulsivity, and Sensation Seeking. Plos ONE, 8(1), 1-8. doi:10.1371/journal.pone.0054402

Strayer, D. L., & Fisher, D. L. (2015). SPIDER: A framework for understanding driver distraction. *Human Factors*, Vol. XX, No. X, Month XXXX, pp. 1 –8. DOI: 10.1177/0018720815619074

Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). Measuring cognitive distraction in the automobile II: Assessing in-vehicle voice-based interactive technologies. *AAA Foundation for Traffic Safety.*

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. doi:10.1177/0018720815575149

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In B. H. Ross, B. H. Ross (Eds.), *The psychology of learning and motivation: Advances in research and theory (Vol 54)* (pp. 29-58). San Diego, CA, US: Elsevier Academic Press.