Metacognition of Multi-Tasking: How Well Do We Predict the Costs of Divided Attention?

Jason R. Finley,
Washington University in St. Louis

Aaron S. Benjamin, and
University of Illinois at Urbana-Champaign

Jason S. McCarley
Flinders University

Abstract

Risky multi-tasking, such as texting while driving, may occur because people misestimate the costs of divided attention. In two experiments, participants performed a computerized visual-manual tracking task in which they attempted to keep a mouse cursor within a small target that moved erratically around a circular track. They then separately performed an auditory *n*-back task. After practicing both tasks separately, participants received feedback on their single-task tracking performance and predicted their dual-task tracking performance before finally performing the two tasks simultaneously. Most participants correctly predicted reductions in tracking performance under dual-task conditions, with a majority overestimating the costs of dual-tasking. However, the between-subjects correlation between predicted and actual performance decrements was near zero. This combination of results suggests that people do anticipate costs of multi-tasking, but have little metacognitive insight on the extent to which they are personally vulnerable to the risks of divided attention, relative to other people.

Keywords

divided attention; dual task; multi-tasking; metacognition; tracking; prediction

Modern life and technology place increasing demands on human attention, including the frequent demand to perform multiple tasks at once. Dividing attention between two tasks, or time-sharing, generally impairs performance on one or both tasks (Pashler, 1994; Wickens, 1980). This can have serious consequences. For example, U.S. police reports implicated distraction as a contributing factor in 20% of injury-causing car crashes in 2009 (307,000 of 1,517,000; NHSTA, 2010). One particular driver distraction of increasing concern is cell phone use. A large observational study found 5% of U.S. drivers using handheld cell phones

in 2010 (NHTSA, 2011), despite the fact that holding a phone conversation, whether handheld or hands-free, has been shown to impair driving performance, particularly by slowing reaction times to events such as signal changes and braking cars (Basacik, Reed, & Robbins, 2011; Horrey & Wickens, 2006; Strayer & Drews, 2007; Strayer, Drews, & Johnston, 2003). The dangers of distracted driving have prompted the U.S. government to create a website dedicated to the topic (http://distraction.gov).

But how much are people aware of such dual-task costs? That is, to what extent do we have accurate *metacognition* about decrements in performance under divided attention? The decision to engage in multi-tasking behavior, such as using a cell phone while driving, is often under volitional control. People may therefore be more likely to engage in such risky behavior if they underestimate its costs. Although many studies have addressed metacognitive monitoring and control in the context of human learning and memory (cf. Benjamin, 2008; Finley, Tullis, & Benjamin, 2010), little is yet known about metacognition in multi-tasking. Some relevant work on metacognition about visual attention has suggested that people tend to overestimate their ability to detect changes (Levin, Momen, Drivdahl, & Simons, 2000) and their ability to simultaneously allocate attention to multiple locations in natural scenes (Kawahara, 2010). Studies concerning the simultaneous use of several media sources (media multi-tasking) have shown that people who self-report the most multi-tasking are often those least able to filter out irrelevant information in laboratory task switching and *n*-back tasks (Ophir, Nass, & Wagner, 2009) and that people tend to overestimate their general ability to multi-task, relative to others (Sanbonmatsu, Strayer, Medeiros-Ward, & Watson, 2013). Several studies have used closed-circuit driving tasks to investigate peoples' post-task estimates of decrements in their driving performance due to simultaneous secondary tasks such as a guessing game, digit recognition, or mental arithmetic (Horrey, Lesch, & Garabet, 2008, 2009; Lesch & Hancock, 2004). They have generally found that participants indeed recognized that their driving suffered, but that their estimates of decrement did not correspond well to their actual decrements. That is, participants whose performance was impaired the most did not generally give the biggest decrement estimates, and vice versa. Horrey et al. (2008) found approximately equal numbers of participants under-estimating and over-estimating the driving performance costs of divided attention. It is worth noting that there are considerable individual differences in the impact of dual-tasking on driving performance, as evidenced by Watson and Strayer (2010), who even found that some "supertaskers" were not impaired at all. But these individual differences have not been well reflected by participants' own performance estimates.

Although such studies provide valuable data with high ecological validity, the conclusions we can draw from them about metacognition are limited due to the fact that estimates of dual-task costs were made *after* performing the tasks. Participants' estimates could have simply been based on their memories of how well they performed. For people to make strategic decisions about whether and when to engage in multi-task behavior, they must be able to accurately *predict* what the performance costs will be. Thus, the purpose of the present study was to investigate the extent to which people can accurately predict the costs of divided attention, in a controlled laboratory setting. The primary task was a visual-manual

pursuit tracking task (chosen to be roughly analogous to the demands of vehicle control), and the secondary task was an auditory *n*-back task in which the value of n varied from 1 to 3 across blocks. An auditory secondary task was chosen to roughly mimic the demands of engaging in a conversations while driving, and to ensure that any dual-task decrements in tracking performance could not be attributed to the need to visually scan between multiple stimuli. Participants practiced both types of tasks individually, saw feedback on their performance, and then predicted what their tracking performance would be when the two tasks were combined. They then performed the two tasks together.

Experiment 1

The purpose of Experiment 1 was to evaluate the accuracy of participants' predictions of dual-task performance. Participants made predictions about their tracking performance under conditions in which they were asked to simultaneously engage in a memory task. *N*-back tasks with different values of *n* (1, 2, and 3) were used to vary the difficulty of the secondary task (Jaeggi, Buschkuehl, Perrig, & Meier, 2010). This allowed us to assess the effects of increasing memory demand on predicted versus actual performance, and to compare participants' overall predicted change in performance (from single- to dual-task) to their actual change in performance. Furthermore, we sought to assess the between-subjects calibration of the magnitudes of predicted dual-task costs with the magnitudes of actual dual-task costs.

Method

Participants—Participants were 69 right-handed undergraduates (41 female) who received partial course credit. Their mean age was 19.1 years (SD = 1.7), and 46 reported that English was their first language. Data were excluded from one additional participant who did not follow instructions. Data were additionally collected from 9 left-handed participants, and are not reported here.

Design and procedure—The experiment used a 2 x 3 within-subjects design, where the independent variables were task concurrence (single- vs. dual-task) and *n*-back level (1-, 2-, 3-back). The primary dependent measures were tracking performance, *n*-back performance, and dual-task tracking predictions. We will first outline the overall procedure, and will then describe the *n*-back and tracking tasks in detail.

The overall procedure was as follows. Participants first completed a single-task phase consisting of six individual task blocks in the following order: tracking, 1-back, 2-back, 3-back, tracking, and tracking. After completing all of the single-task blocks, participants were instructed that they would next be completing three dual-task blocks in which the two types of single task were combined as follows: tracking + 1-back, tracking + 2-back, and tracking + 3-back. They were furthermore told that the tracking task would have the highest priority. We chose to prioritize tracking by analogy to a driving situation: it is more important to keep the vehicle within a lane than it is to maintain a cell phone conversation (though we did not inform participants of this analogy). Participants were shown their scores (percent of time on target) from the three single-task tracking blocks, and on the same screen were asked to predict (0–100%) what their tracking performance would be when done at the same

time as the 1-back, 2-back, and 3-back tasks. Thus participants made three predictions, one for each value of *n*. Note that participants' single-task *n*-back performance was not shown on the prediction screen. Finally, participants completed the three dual task blocks, in an order that was counterbalanced between subjects. After completing all three dual task blocks, participants were asked to describe any strategies they had used during those blocks, either for the tracking component or the *n*-back component.

Tasks: Participants engaged in the tasks individually on computers running Windows XP and programmed with REALBasic. Visual stimuli were presented on the computer screen and auditory stimuli were presented via headphones. Participants responded using a standard keyboard and mouse. Computer screens were 17 inches diagonal (43.18 cm) with resolution 1024 x 768 pixels (px). All task parameters were constant as described below and did not change in response to performance. The parameter values were chosen based on pilot data in order to obtain intermediate levels of mean performance.

N-back task: At the start of each *n*-back task block, participants were informed that they would hear a series of numbers and would have to indicate whether or not each number was the same as the number they heard *n* (1, 2, or 3) places ago. An appropriate example was described in each case. Participants were asked to respond as quickly and accurately as possible. They then listened to a series of single-digit numbers (1–9) spoken in a synthesized voice at a rate of one number every 2.4 s, for a total duration of 60 s. The number sequence was generated randomly for each participant and each *n*-back task, with the constraint that the correct answer was yes for 50% of the numbers (rounded up when applicable). For each number after the first *n* digits, participants responded using their left hand on the keyboard, pressing the c key for yes/same and the z key for no/different. A reminder of the meaning of the two response keys remained on the screen during the task. Participants were given ongoing feedback as follows. When participants gave a correct response, a green check mark was shown on the screen until the next number was spoken. When participants gave an incorrect response, a red x mark was shown instead. No such feedback was shown in cases where participants did not respond for a number. At the end of a task block, participants were given their score as the percent of numbers (after the first *n*) on which they responded correctly.

Visual-manual tracking task: This task was modeled after classic rotary pursuit tasks (Adams, 1961). At the start of each tracking task, participants were instructed that their task would be to keep the tip of the cursor arrow inside a small circular target as it moved around a blue circular track. The target circle was 32 px in diameter (≈ 1 cm), and the track circle was 300 px in diameter (≈ 10 cm) and 4 px thick. Participants controlled the position of the cursor arrow with their right hand on the mouse (a zero-order control, Wickens & Hollands, 2000, pp. 417–418). The task began when participants positioned the tip of the cursor arrow inside the target and pressed the space bar. For a duration of 60 s, the target circle moved around the track at a rate of 1 revolution per 4.5 s, changing directions between clockwise and counterclockwise a total of 20 times at intervals randomly determined for each participant and each task block. Every 15 ms the program recorded whether the tip of the cursor arrow was within the target, and updated the position of the target. The target was

solid black when the tip of the cursor arrow was outside of it, and turned white with a black outline when the tip of the cursor arrow was inside of it. At the end of the task block, participants were given their score as the percent of time that they had been on target.

Dual task: The dual task was simply the concurrent combination of the two single tasks, with the following changes. The instructions asked participants to do their best on both tasks, but emphasized that tracking was more important. After reading the dual-task instructions and just before beginning the dual-task, participants had to confirm which *n*-back task they were about to attempt (i.e., 1, 2, or 3). This was done to ensure that they had carefully read the dual-task instructions. Additionally, participants were not given ongoing feedback on their *n*-back performance (i.e., no green check marks or red x marks), nor were they given their *n*-back or tracking scores at the end of a task block.

Results and Discussion

An alpha level of .05 was used for all tests of statistical significance unless otherwise noted. Effect sizes for comparisons of means are reported as Cohen's *d* calculated using the pooled standard deviation of the groups being compared (Olejnik & Algina, 2000, Box 1 Option B). Standard deviations reported are uncorrected for bias (i.e., calculated using *N*, not *N*-1). An italicized lowercase *n* denotes the value used in an *n*-back task (i.e., 1, 2, or 3). Each participant's dual-task tracking performance *decrements*, both actual and predicted, were calculated with respect to his or her performance in the final single-task tracking block.

Figure 1 provides an overview of single- and dual-task performance, as well as dual-task predictions. A Supplemental Table provides the corresponding means and standard deviations.

Single-task performance—As expected, single-task *n*-back performance decreased as *n* increased, confirmed by the mean slope of separate simple linear regressions for each participant, $M_b = -.06$, $SD_b = .11$, $t(68) = 4.87$, $p < .001$. Single-task tracking performance increased from the first to the second block, $M = .06$, $SD = .08$, $t(68) = 6.80$, $p < .001$, $d = 0.58$, and decreased slightly from the second to the third block, $M = -.02$, $SD = .06$, $t(68) = 3.09$, $p = .003$, $d = 0.21$. The latter result is important because it suggests that single-task tracking performance had reached asymptote by the third block. Thus it was unlikely that there were further increases in performance due to practice, which could have offset any dual-task decrement.

Dual-task performance—Averaged over *n*, dual-task *n*-back performance did not reliably differ from single-task *n*-back performance, $M = .03$, $SD = .14$, $t(68) = 1.74$, $p = .087$, $d = 0.17$. First, this demonstrates that participants were still putting effort into the *n*-back task (i.e., *n*-back performance did not drop to floor). Second, the marginal increase in performance (from .75 to .78) suggests that there were practice effects for *n*-back. That is, single-task *n*-back performance probably had not reached asymptote as single-task tracking performance had.[1] As *n* increased, dual-task *n*-back performance reliably decreased, $M_b = -.07$, $SD_b = .08$, $t(68) = 7.21$, $p < .001$, just as it had in the single-task blocks.

Averaged over n, dual-task tracking performance was reliably lower than single-task tracking performance had been on the third single-task block, $M = -.05$, $SD = .05$, $t(68) = 7.21$, $p < .001$, $d = 0.45$. As n of the concurrent n-back task increased, dual-task tracking performance reliably decreased, $M_b = -.02$, $SD_b = .02$, $t(68) = 5.24$, $p < .001$. Thus, there was indeed an overall dual-task decrement in tracking performance, and this decrement became larger with increasing difficulty of the concurrent n-back task. There were 55 participants whose mean tracking performance numerically decreased under divided attention ($M_{decrement} = 6\%$, $SD = 4\%$), and only 14 whose mean tracking performance numerically increased under divided attention ($M_{increment} = 2\%$, $SD = 2\%$).

Metacognition—Prediction data were converted from percentages to proportions for analysis. Although participants did not directly predict dual-task decrements in tracking performance, they made their predictions of raw dual-task tracking performance in the presence of their single-task tracking performance scores, and thus we will use "predicted decrement" as a term of convenience when referring to the difference between final single-task tracking performance and predicted dual-task tracking performance.

Overall pattern: Averaged over n, participants' predictions of dual-task tracking performance were reliably lower than their most recent single-task tracking performance had been, $M = -.19$, $SD = .17$, $t(68) = 9.02$, $p < .001$, $d = 1.24$. As n of the concurrent n-back task increased, predictions of dual-task tracking performance reliably decreased, $M_b = -.08$, $SD_b = .07$, $t(68) = 10.09$, $p < .001$. Thus, participants correctly predicted a dual-task decrement in tracking performance, and furthermore predicted a larger decrement with increasing difficulty of the concurrent n-back task. These predictions concurred with the pattern of actual performance.

As apparent in Figure 1, the slope of participants' predictions was more similar to the slope of their single-task n-back performance than to the slope of their actual dual-task tracking performance, $t(68) = 3.66$, $p < .001$, $d = 0.44$. This is consistent with participants basing their predictions in part on a heuristic that difficulty increases with n (as supported by their memory of their single-task n-back performance) and that greater difficulty should translate into greater interference with the tracking task.

Absolute accuracy (calibration): To evaluate the absolute accuracy (i.e., *calibration*) of participants' metacognition, we calculated the mean of signed difference scores between predicted and actual decrements in performance (both measures averaged over n for each participant). The predicted decrement (19%) was much larger than the actual decrement (5%), $M_{diff} = .14$, $SD_{diff} = .18$, $t(68) = 6.45$, $d = 1.11$. Furthermore, the predicted slope of the decrement across levels of n was greater than the actual slope, $M_{diff} = .07$, $SD_{diff} = .07$, $t(68) = 7.90$, $d = 1.31$. Out of all 69 participants, 49 over-estimated the dual-task cost by more

---

[1]In fact, there was a fairly even split between participants whose mean n-back performance numerically increased (37 participants in Experiment 1, and 22 in Experiment 2) versus decreased (31 participants in Experiment 1, and 26 in Experiment 2). Although this may be a limitation for traditional dual-task analyses, it does not pose a problem for our metacognitive analyses in either experiment, because the correlation between participants' mean change in n-back performance (from single- to dual-task) and their mean difference between predicted and actual dual-task tracking performance was essentially zero: −.016 in Experiment 1 and .005 in Experiment 2.

than 5%, 9 under-estimated the cost by more than 5%, and 11 produced estimates within 5% of the actual cost. Thus, most participants *overestimated* the cost of divided attention and overestimated how much that cost would increase as the secondary task became more difficult. Fortunately, this type of miscalibration should bias most people toward being more cautious. But those few who underestimate the cost of dual-tasking may be more likely to engage in it.

Relative accuracy (group-level resolution): To evaluate the relative accuracy (i.e., *resolution*) of participants' metacognition, we calculated a between-subjects correlation. Note that in much of the literature on metamemory and confidence, resolution correlations are calculated within-subjects to evaluate individual participants' ability to discriminate between, for example, low and high probability events (cf. Dunlosky & Metcalfe, 2009, pp. 49–57; Lichtenstein & Fischhoff, 1977). Here, however, we were interested in the general tendency for high performers to give higher predictions and low performers to give lower predictions. To avoid confusion, we will refer to this tendency as group-level resolution.

Across participants and averaged over *n*, the Pearson correlation between predicted dual-task tracking performance and actual dual-task tracking performance was .37, $t(67) = 3.30$, $p = .002$, 95% CI [.15, .56]. Figure 2 illustrates the relationship. Both measures had high reliability, Cronbach's $\alpha = .948$ and .958, respectively. This result suggests a somewhat good group-level resolution for raw tracking performance in a dual-task situation.

The above traditional correlation analysis required combining each participant's three predictions and three actual tracking scores using simple arithmetic means, which assume equal weights for each value. A more sophisticated alternative is to use a multivariate analysis: the canonical correlation analysis. Canonical correlation analysis allows us to assess the relationship between two *sets* of variables, without the need for averaging (Sherry & Henson, 2005). Because there are multiple possible linear combinations of the variables in each set, a canonical correlation analysis produces multiple solutions, the first of which is the strongest possible, given the data. We performed such an analysis assessing the relationship between the three dual-task tracking predictions and the three dual-task tracking performance measurements. The first canonical correlation was .44 (i.e., 20% overlapping variance), and the remaining two were effectively zero. With all three canonical correlations included, $\chi^2(9) = 16.93$, $p = .050$, and with the first removed, $\chi^2(4) = 2.94$, $p = .568$, indicating that the first canonical correlation indeed reliably differed from zero. With a cutoff correlation of .3, all three predictions loaded on the prediction variate, and all three performance measurements loaded on the performance variate. Thus, we confirmed the positive group-level resolution suggested by the standard correlation.

However, across participants, the correlation between predicted dual-task *decrement* and actual dual-task decrement was −.008, $t(67) = 0.07$, $p = .947$, 95% CI [−.26, .23]. Figure 3 illustrates the relationship. Both measures had high reliability, Cronbach's $\alpha = .936$ and .832, respectively; applying Spearman's correction for attenuation yielded an adjusted correlation of −.009. As before, we additionally performed a canonical correlation analysis, this time using the sets of predicted and actual decrements in tracking performance. The first

canonical correlation was .26 (i.e., 7% overlapping variance), but neither it nor any of the others were statistically significant, $\chi^2(9) = 7.03$, $p = .634$.

To summarize, although participants who ended up generally performing worse indeed gave lower predictions, those whose performance suffered the most under divided attention did not correspondingly predict larger decrements. The latter finding is further demonstrated by the fact that the correlation between predicted and actual dual-task tracking slopes over *n*-back values was .10, $t(67) = 0.81$, $p = .423$, 95% CI [−.14, .33]. These results suggest that participants had almost no insight into their relative level of susceptibility to the performance costs of divided attention.

Experiment 2

Experiment 1 showed that participants' tracking performance decreased under dual-task conditions, and did so to a greater degree as the difficulty of the concurrent *n*-back task increased. Furthermore, participants' predicted patterns of dual-task tracking performance reflected the patterns of their actual performance but in an exaggerated way (both in level and slope). Finally, although participants' predictions showed reasonably good relative accuracy on their raw performance, they showed practically no relative accuracy on their change in performance due to divided attention (i.e., dual task costs).

The instructions in Experiment 1, both at the time of predictions and at the time of the dual tasks themselves, emphasized that tracking performance was more important than *n*-back performance. One possible consequence of this is that some participants may have shirked the *n*-back task to the degree that their dual-task tracking performance was barely influenced, and thus their predictions were farther off than they might have been otherwise. Although the fact that *n*-back performance did not decrease from single- to dual-task argues against this possibility, we nevertheless wanted to rule it out with Experiment 2 by making the instructions emphasize performance on the *n*-back task. Furthermore, Experiment 2 provided a chance to replicate the basic findings of Experiment 1.

Method

Participants—Participants were 48 right-handed undergraduates (31 female) who received partial course credit. Their mean age was 18.9 years ($SD = 1.0$), and 36 reported that English was their first language. Data were excluded from two additional participants who did not follow instructions. Data were additionally collected from 8 left-handed participants, and are not reported here.

Design and procedure—The design and procedure were identical to those used in Experiment 1, with the exception that the instructions, both at the time that dual-task tracking performance predictions were made and at the time that the dual tasks were performed, told participants to prioritize performance on the *n*-back tasks.

Results and Discussion

Figure 4 provides an overview of single- and dual-task performance, as well as dual-task predictions. A Supplemental Table provides the corresponding means and standard deviations. The overall pattern of results was the same as in Experiment 1.

Single-task performance—Single-task n-back performance again decreased as n increased, confirmed by the mean slope of separate simple linear regressions for each participant, $M_b = -.06$, $SD_b = .12$, $t(47) = 3.54$, $p = .001$. Single-task tracking performance increased from the first to the second block, $M = .07$, $SD = .09$, $t(47) = 5.66$, $p < .001$, $d = 0.72$, and did not reliably change from the second to the third block, $M = .01$, $SD = .06$, $t(47) = 0.98$, $p = .353$, $d = 0.09$.

Dual-task performance—Averaged over n, dual-task n-back performance did not reliably differ from single-task n-back performance, $M = .02$, $SD = .11$, $t(47) = 1.55$, $p = .128$, $d = 0.14$. As n increased, dual-task n-back performance reliably decreased, $M_b = -.09$, $SD_b = .09$, $t(47) = 7.26$, $p < .001$, just as it had in the single-task blocks, and just as was observed in Experiment 1.

Averaged over n, dual-task tracking performance was reliably lower than single-task tracking performance had been on the third single-task block, $M = -.06$, $SD = .06$, $t(47) = 7.60$, $p < .001$, $d = 0.67$. As n of concurrent n-back increased, dual-task tracking performance reliably decreased, $M_b = -.02$, $SD_b = .04$, $t(47) = 3.25$, $p = .002$. These performance data replicate those found in Experiment 1. There were 43 participants whose mean tracking performance numerically decreased under divided attention ($M_{decrement} = 8\%$, $SD = 5\%$), and only five whose mean tracking performance numerically increased under divided attention ($M_{increment} = 4\%$, $SD = 3\%$).

Metacognition

Overall pattern: Averaged over n, participants' predicted dual-task tracking performance was reliably lower than their most recent single-task tracking performance had been, $M = -.20$, $SD = .15$, $t(47) = 9.12$, $p < .001$, $d = 1.53$. As n of concurrent n-back increased, predicted dual-task tracking performance reliably decreased, $M_b = -.08$, $SD_b = .06$, $t(47) = 9.16$, $p < .001$. Thus, participants were again successful in predicting a drop in tracking performance under divided attention and a negative slope as the n-back task became harder. Also consistent with the results of Experiment 1, the slope of participants' predictions was more similar to the slope of their single-task n-back performance than to the slope of their actual dual-task tracking performance, $t(47) = 2.44$, $p = .019$, $d = 0.45$, again suggesting partial reliance on memory for n-back performance in the prediction of dual-task tracking performance.

Absolute accuracy (calibration): However, as in Experiment 1, the predicted decrement (20%) was much larger than the actual decrement (6%), $M_{diff} = .13$, $SD_{diff} = .15$, $t(47) = 6.07$, $d = 1.18$, and the predicted slope of the decrement was greater than the actual slope, $M_{diff} = .06$, $SD_{diff} = .07$, $t(47) = 6.09$, $d = 1.30$. Out of all 48 participants, 35 over-estimated the dual-task cost by more than 5%, 5 under-estimated the cost by more 5%, and 8 produced

estimates within 5% of the actual cost. Thus, participants again overestimated the cost of divided attention and overestimated how much that cost would increase as the *n*-back task became more difficult. This replication of the results found in Experiment 1 indicates that the instructions emphasizing one task over another did not distort participants' true calibration.

Relative accuracy (group-level resolution): Across participants and averaged over *n*, the correlation between predicted dual-task performance and actual dual-task performance was .37, $t(46) = 2.70$, $p = .010$, 95% CI [.10, .60]. Figure 5 illustrates the relationship. Both measures had high reliability, Cronbach's α = .946 and .929, respectively. We again performed a canonical correlation analysis to assess the overall relationship between the three predictions and the three performance measurements. The first canonical correlation was .62 (i.e., 38% overlapping variance), and the remaining two were effectively zero. With all three canonical correlations included, $\chi^2(9) = 23.80$, $p = .005$, and with the first removed, $\chi^2(4) = 3.16$, $p = .532$, indicating that the first canonical correlation reliably differed from zero. With a cutoff correlation of .3, all three predictions loaded on the prediction variate, and all three performance measurements loaded on the performance variate. Thus, participants again showed fairly good group-level resolution in predicting their raw performance.

However, across participants, the correlation between predicted dual-task decrement and actual dual-task decrement was .17, $t(46) = 1.15$, $p = .256$, 95% CI [−.12, .43]. Figure 6 illustrates the relationship. Both measures had high reliability, Cronbach's α = .941 and .802, respectively; applying Spearman's correction for attenuation yielded an adjusted correlation of .19. We again performed a canonical correlation analysis using the predicted and actual decrements in performance. The first canonical correlation was .41 (i.e., 17% overlapping variance), but neither it nor any of the others were statistically significant, $\chi^2(9) = 10.11$, $p = .341$. As in Experiment 1, the correlation between predicted and actual dual-task tracking slopes over *n*-back values did not reliably differ from zero, −.07, $t(46) = 0.50$, $p = .618$, 95% CI [−.35, .21]. Thus, participants again showed poor group-level resolution in predicting the costs of divided attention.

General Discussion

The present experiments employed a laboratory visual-manual tracking task for which performance generally suffers with the addition of a concurrent auditory *n*-back task, and does so to a greater extent as the difficulty of the *n*-back task increases. The goal of the study was to investigate participants' metacognition about the costs of divided attention by measuring their predictions of performance under dual-task conditions. Participants correctly predicted the overall pattern of their tracking performance, but tended to overestimate their dual-task decrement and the slope of that decrement as the *n*-back task became harder. Moreover, across participants there was practically no correlation between predicted and actual decrements. That is, those participants whose performance suffered the most under divided attention did not generally predict the largest decrement.

These results are consistent with results from driving studies in which participants gave post-task estimates (postdictions) of their performance. For example, Kidd and Horrey (2010) found that participants overestimated the decrement in their driving performance that had occurred due to a concurrent auditory arithmetic task. Horrey, Lesch, and Garabet (2008) found that participants' postdictions of driving performance decrements (due to concurrent paced auditory serial addition task) were unrelated to their actual performance decrements, indicating poor group-level resolution. The similarity between such previous findings and the current findings is important because it illustrates the same pattern for both predictions and postdictions, and with both an unfamiliar primary task and a familiar primary task. Note however that in most cases the secondary task has been an unfamiliar one; it may be that drivers in fact underestimate the dual-task costs for extremely familiar tasks such as having a cell phone conversation, and future research should investigate this. Finally, it is also worth considering that demand characteristics could influence participants' estimates of dual-task performance, such that participants adjust their reports toward what they think the researchers want to hear; future research should seek to rule out such a possibility.

If people truly tend to overestimate the costs of attempting two tasks at once, how can we reconcile that with the rate of car crashes involving distracted driving (17% of all crashes, 20% of injury-causing ones, and 16% of fatal ones; NHSTA, 2010)? It may be that most cases of distracted driving are due to the minority of people who in fact underestimate the cost of divided attention. Or, it could be that drivers rely on *relative* judgments to guide multi-tasking behavior (e.g., "I'm better than the average driver at handling distraction, so I'll answer this phone call."); in that case, the poor group-level metacognitive resolution that we and others have documented may contribute to risky behavior. Another possibility is that drivers often simply fail to consider performance costs. In fact, the ongoing demand of the primary task itself may reduce the cognitive resources necessary to make such a metacognitive judgment, or to make a strategic plan based on the judgment. Additional stress or distraction may further exacerbate this problem.

Finally, perhaps drivers engage in multi-tasking behavior *despite* their metacognitive judgments about performance costs. For example, Atchley and colleagues have found that young drivers' perceptions of the general riskiness of cell phone use while driving are only a weak predictor of self-reported distracted driving behavior (Atchley, Atwood, & Boulton, 2011), and that perceived risk could be outweighed by the perceived importance of a particular communication (Nelson, Atchley, & Little, 2009). There may be a variety of reasons for such an apparent disconnect between judgment and behavior. First, drivers may inappropriately judge an estimated reduction in performance to be acceptable if they lack accurate understanding of how that reduction translates into greater probability of an adverse outcome, such as a crash. Second, drivers' behaviors may be influenced by personality factors such as optimism, risk aversion, and sensation-seeking (Schwebel, Severson, Ball, & Rizzo, 2006). Finally, drivers' poor choices may furthermore be compounded by the generally poor group-level metacognitive resolution found in the current study and in prior studies. Drivers whose risk would increase the most may underestimate their increase in risk, and thus be more likely to judge that risk as acceptable.

There is some evidence that drivers indeed fail to manage distraction in the same kinds of situations where they have overestimated the costs of divided attention. Horrey and Lesch (2009) gave participants control over when to initiate secondary tasks during a driving task. They found that participants did *not* strategically delay initiation of secondary tasks until driving conditions were less demanding. These results suggest a failure of judicious use of metacognition about divided attention, a topic which continues to gain importance in a distraction-laden world.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Adams JA. Human tracking behavior. Psychological Bulletin. 1961; 58(1):55–79.10.1037/h0041559 [PubMed: 13681394]

Atchley P, Atwood S, Boulton A. The choice to text and drive in younger drivers: Behavior may shape attitude. Accident Analysis and Prevention. 2011; 43:134–142.10.1016/j.aap.2010.08.003 [PubMed: 21094307]

Basacik, D.; Reed, N.; Robbins, R. Smartphone use while driving: a simulator study. Published Project Report PPR592. 2011. Retrieved from Institute of Advanced Motorists website: http://www.iam.org.uk/images/stories/policy-research/PPR592_secure.pdf

Benjamin, AS. Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In: Benjamin, AS.; Ross, BH., editors. The Psychology of Learning and Motivation: Skill and Strategy in Memory Use. Vol. 48. London: Academic Press; 2008. p. 175-223.

Dunlosky, J.; Metcalfe, J. Metacognition. Thousand Oaks, CA: Sage Publications, Inc; 2009.

Finley, JR.; Tullis, JG.; Benjamin, AS. Metacognitive control of learning and remembering. In: Khine, MS.; Saleh, I., editors. New science of learning: cognition, computers and collaboration in education. Springer; 2010. p. 109-131.

Horrey WJ, Lesch MF. Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. Accident Analysis & Prevention. 2009; 41(1):115–122.10.1016/j.aap.2008.10.008 [PubMed: 19114145]

Horrey WJ, Lesch MF, Garabet A. Assessing the awareness of performance decrements in distracted drivers. Accident Analysis & Prevention. 2008; 40(2):675–682.10.1016/j.aap.2007.09.004 [PubMed: 18329420]

Horrey WJ, Lesch MF, Garabet A. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task situations. Journal of Safety Research. 2009; 40(1):7–12.10.1016/j.jsr.2008.10.011 [PubMed: 19285580]

Horrey WJ, Wickens CD. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Human Factors. 2006; 48(1):196–205.10.1518/001872006776412135 [PubMed: 16696268]

Jaeggi SM, Buschkuehl M, Perrig WJ, Meier B. The concurrent validity of the *N*-back task as a working memory measure. Memory. 2010; 18(4):394–412.10.1080/09658211003702171 [PubMed: 20408039]

Kawahara J. Measuring the spatial distribution of the metaattentional spotlight. Consciousness and Cognition. 2010; 19:107–124.10.1016/j.concog.2009.10.004 [PubMed: 20006524]

Kidd DG, Horrey WJ. Do drivers' perceptions of distracted driving become more accurate over time? Professional Safety. 2010; 2010(January):40–45.

Lesch MF, Hancock PA. Driving performance during concurrent cellphone use: Are drivers aware of their performance decrements? Accident Analysis & Prevention. 2004; 36(3):471–480.10.1016/S0001-4575(03)00042-3 [PubMed: 15003592]

Levin DT, Momen N, Drivdahl SB, Simons DJ. Change blindness blindness: The metacognitive error of overestimating change-detection ability. Visual Cognition. 2000; 7:397–412.10.1080/135062800394865

Lichtenstein S, Fischhoff B. Do those who know more also know more about how much they know? Organizational Behavior and Human Performance. 1977; 20:159–183.10.1016/0030-5073(77)90001-0

National Highway Traffic Safety Administration [NHTSA]. Distracted driving 2009. 2010. (Report No. DOT HS 811 379). Retrieved from http://www.distraction.gov/research/PDF-Files/Distracted-Driving-2009.pdf

National Highway Traffic Safety Administration [NHTSA]. Driver electronic device use in 2010. 2011. (Report No. DOT HS 811 517). Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/811517.pdf

Nelson E, Atchley P, Little TD. The effects of perception of risk and importance of answering and initiating a cellular phone call while driving. Accident Analysis and Prevention. 2009; 41:438–444.10.1016/j.aap.2009.01.006 [PubMed: 19393790]

Olejnik S, Algina J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. Contemporary Educational Psychology. 2000; 25:241–286.10.1006/ceps.2000.1040 [PubMed: 10873373]

Ophir E, Nass C, Wagner AD. Cognitive control in media multitaskers. Proceedings of the National Academy of Sciences. 2009; 106(37):15583–15587.10.1073/pnas.0903620106

Pashler H. Dual-task interference in simple tasks: Data and theory. Psychological Bulletin. 1994; 116:220–244.10.1037/0033-2909.116.2.220 [PubMed: 7972591]

Sanbonmatsu DM, Strayer DL, Medeiros-Ward N, Watson JM. Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. PLoS ONE. 2013; 8(1):e54402.10.1371/journal.pone.0054402 [PubMed: 23372720]

Schwebel DC, Severson J, Ball KK, Rizzo M. Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. Accident Analysis and Prevention. 2006; 38(4):801–810.10.1016/j.aap.2006.02.004 [PubMed: 16527223]

Sherry A, Henson RK. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. Journal of Personality Assessment. 2005; 84(1):37–48.10.1207/s15327752jpa8401_09 [PubMed: 15639766]

Strayer DL, Drews FA. Cell-phone-induced driver distraction. Current Directions In Psychological Science. 2007; 16:128–131.10.1111/j.1467-8721.2007.00489.x

Strayer DL, Drews FA, Johnston WA. Cell phone-induced failures of visual attention during simulated driving. Journal of Experimental Psychology: Applied. 2003; 9:23–52.10.1037/1076-898X.9.1.23 [PubMed: 12710835]

Watson JM, Strayer DL. Supertaskers: Profiles in extraordinary multitasking ability. Psychonomic Bulletin & Review. 2010; 17(4):479–485.10.3758/PBR.17.4.479 [PubMed: 20702865]

Wickens, CD. The structure of attentional resources. In: Nickerson, R., editor. Attention and Performance VIII. Hillsdale, NJ: Erlbaum; 1980. p. 239-257.

Wickens, CD.; Hollands, JG. Engineering Psychology and Human Performance. 3. Upper Saddle River, New Jersey: Prentice-Hall Inc; 2000.
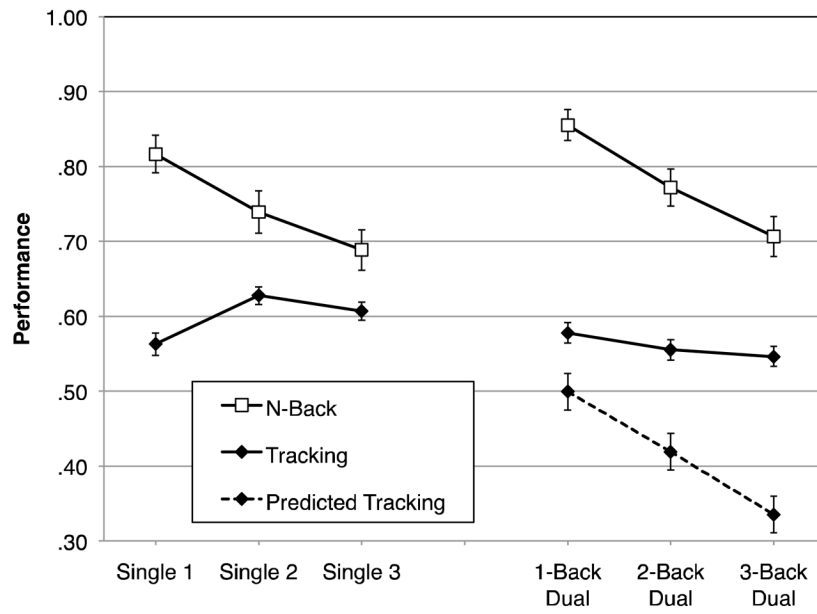
Figure 1.
Mean single- and dual-task *n*-back performance (proportion correct) as a function of n, mean single-task tracking performance (proportion time on target) as a function of task number, and mean dual-task tracking performance and predictions (proportion time on target) as a function of *n* in concurrent *n*-back, in Experiment 1. Error bars represent standard error per cell.
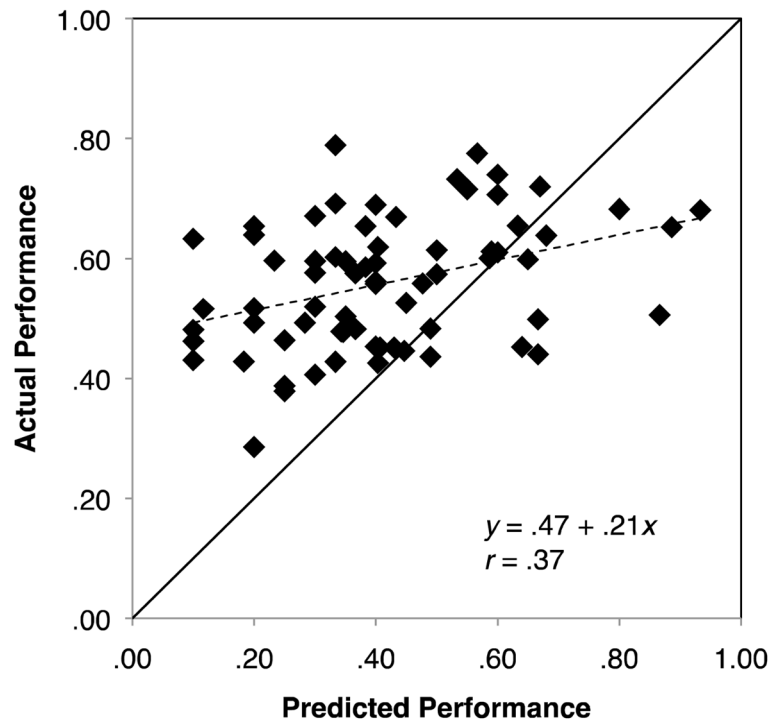
Figure 2.
Actual dual-task tracking performance as a function of predicted dual-task tracking performance, in Experiment 1. Each dot represents one participant. Values are averaged over *n* of concurrent *n*-back task for each participant. The dark diagonal line represents perfect calibration, and the dotted line is the regression line.
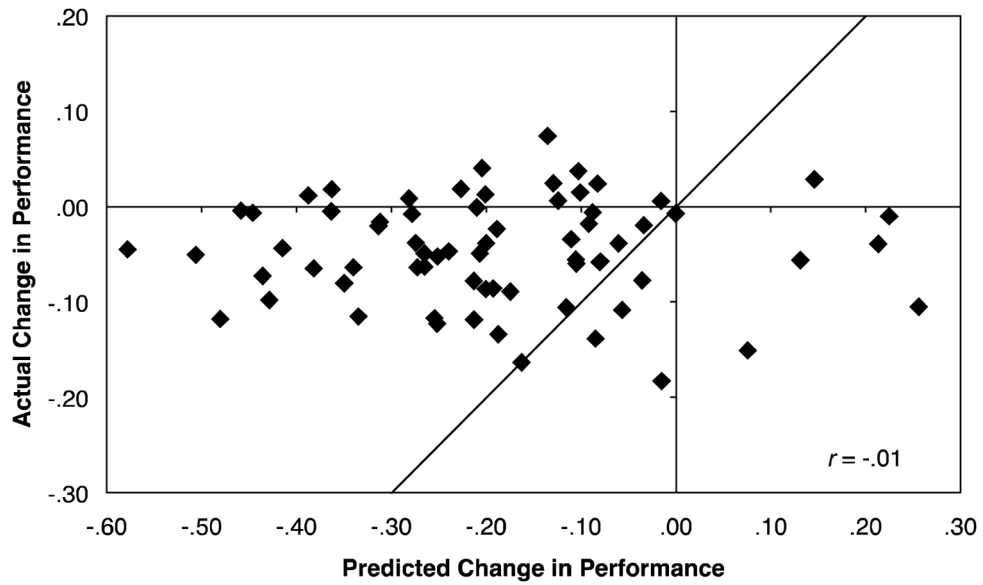
Figure 3.
Actual change in tracking performance as a function of predicted change in tracking performance, in Experiment 1. Change was measured from performance on the last single-task tracking block to performance averaged across dual-task blocks. Each dot represents one participant. The dark diagonal line represents perfect calibration.
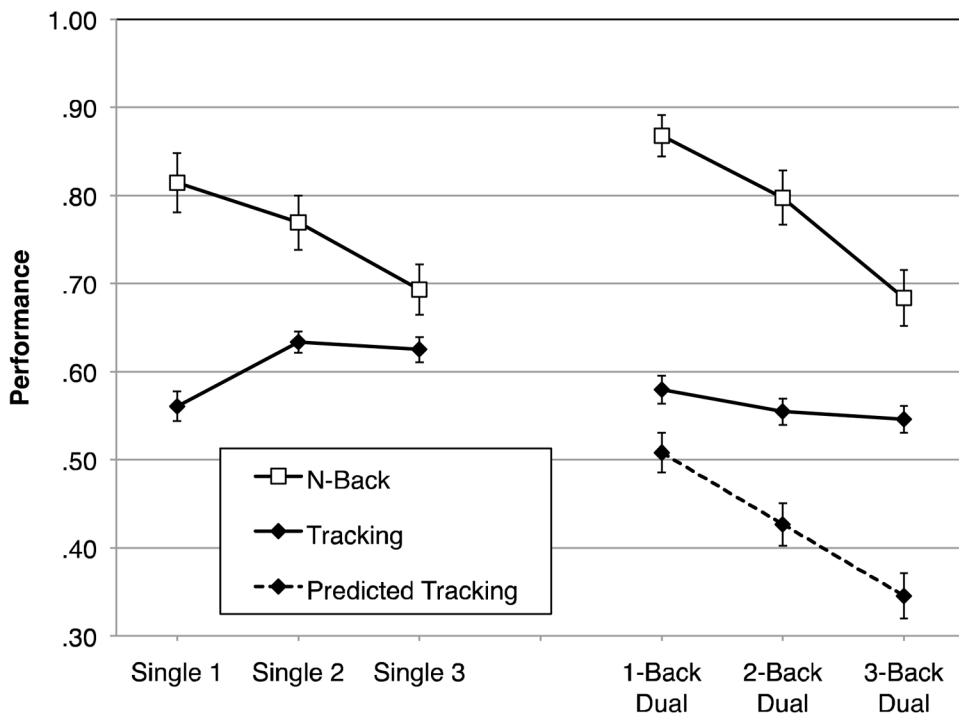
Figure 4.
Mean single- and dual-task *n*-back performance (proportion correct) as a function of n, mean single-task tracking performance (proportion time on target) as a function of task number, and mean dual-task tracking performance and predictions (proportion time on target) as a function of *n* in concurrent *n*-back, in Experiment 2. Error bars represent standard error per cell.
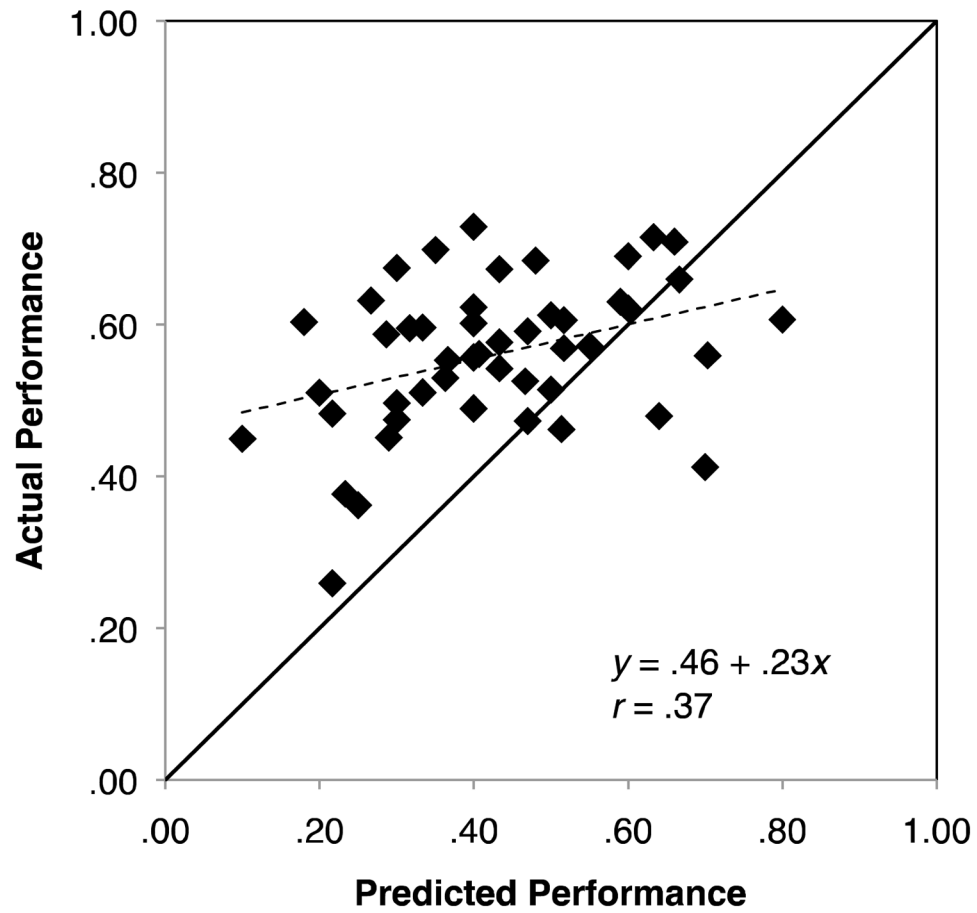
$$y = .46 + .23x$$
$$r = .37$$

Figure 5.
Actual dual-task tracking performance as a function of predicted dual-task tracking performance, in Experiment 2. Each dot represents one participant. Values are averaged over *n* of concurrent *n*-back task for each participant. The dark diagonal line represents perfect calibration, and the dotted line is the regression line.
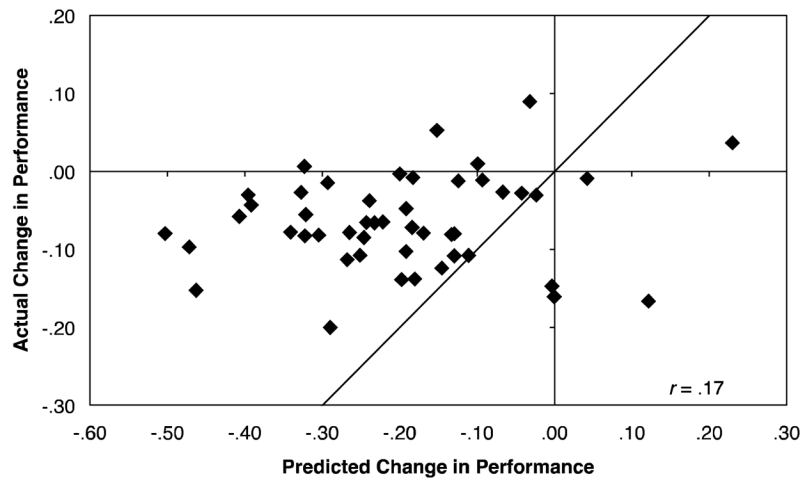
Figure 6.
Actual change in tracking performance as a function of predicted change in tracking performance, in Experiment 2. Change was measured from performance on the last single-task tracking block to performance averaged across dual-task blocks. Each dot represents one participant. The dark diagonal line represents perfect calibration.