Title
Multitasking in Working Memory

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Heathcote, Andrew
Eidels, Ami
Houpt, Joseph
et al.

Publication Date
2014

Peer reviewed

Multi-tasking in Working Memory

Andrew Heathcote (andrew.heathcote@newcastle.edu.au)
Ami Eidels (ami.eidels@newcastle.edu.au)
School of Psychology, University Avenue, Callaghan, NSW, 2308, Australia
Joseph Houpt (joseph.houpt@wright.edu)
Department of Psychology, Wright State University, Dayton, OH 45435-0001, USA
James Colman (james.coleman@psych.utah.edu)
Jason Watson (Jason.watson@psych.utah.edu)
David Strayer (david.strayer@utah.edu)
Department of Psychology, University of Utah, Salt Lake City, UT 84112, USA

Abstract

We developed a novel and game like dual 2-back computerized task, Gatekeeper, which we deployed online with 245 male and female participants ranging in age from 13 to 83 years. Gatekeeper requires participants to remember only 4 items, so does not target memory capacity, but rather measures multitasking ability and interference control in working memory. Participants were faster and more accurate with two-targets than one-target, and Bayesian analysis supported a null effect of gender on accuracy, but accuracy did decrease with age. These results are consistent with the ability to divide attention and control proactive interference being equal for males and females but showing an age-related decline.

Keywords: Multitasking; Working Memory; n-back task, Aging, Gender differences.

The Gatekeeper Task

Having to perform more than one task at a time – multi-tasking – is increasingly common in modern life (Ophir, Nass, & Wagner, 2009; Wallis, 2006). Multitasking almost always degrades performance relative to single-task settings (Wickens, 1980, but see Watson & Strayer, 2010, for an exception). This has important real world consequences, such as when talking on a cell phone while driving (Strayer, & Drews, 2007; Strayer, Drews, & Johnston, 2003; Strayer, & Johnston, 2001), and is exacerbated by the fact that the people who are most capable of multi-tasking are not those who are most likely to engage in it (Sanbonmatsu, Strayer, Medeiros-Ward & Watson, 2013).

Multitasking is also an important component of working-memory measures, such as in complex span tasks, which require performance on a secondary task during memorization. For example, in the Operation Span task (Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005), participants intersperse memorizing items with deciding whether simple addition equalities are true or false. The short-term memory and multi-tasking capacities measured by complex-span tasks can be quite strongly correlated with fluid intelligence measures (Conway, Kane & Engle, 2003), particularly when fluid intelligence is measured under time pressure (Chuderski, 2013).

In this paper we develop a task that measures "memory multitasking", the capacity to perform multiple memory operations. Our aim was to provide a measure targeting attention-demanding processes, such as resolving proactive interference and establishing and updating bindings between stimuli and temporal contexts (Oberauer, 2005). In contrast to operation-span tasks, however, the task was designed to minimise the impact of memory-capacity limitations.

Our task is a variant of the dual n-back task (Jaeggi and colleagues, 2003, 2007, 2008, 2010a), which we called the "Gatekeeper" task. Gatekeeper is a 2-back task, requiring participants to make a decision about a pair of stimuli on the current trial based on their memory about the pair of stimuli that occurred two trials previously. Hence, it requires only four items to be held in memory at any time, minimizing the impact of storage capacity limits (Cowan, 2001; Morey & Cowan, 2004) that reduce performance in higher-order n-back tasks. At the same time, it also avoids easy and fast familiarity-based strategies available in a 1-back task (McElree, 2001).

Participants in the Gatekeeper task were told they were in training to become a doorperson at an exclusive nightclub, and that their task was to allow in only cool patrons. As shown in Figure 1, the task stimuli were both visual (an image of three doors) and auditory (a spoken letter). A potential patron tries to gain access through one of the three doors, which is indicated by that door being colored red, and by saying one of three password letters, "P", "Y" or "O". The small stimulus sets make the Gatekeeper task difficult because of high levels of proactive interference (Keppel & Underwood, 1962) caused by the rapidly varying mapping of the current stimuli to target (i.e., 2-back) and non-target roles (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). As a consequence – and in contrast to other n-back tasks (see Gray, Chabris & Braver, 2003) – all trial types in the Gatekeeper task place a high demands on cognitive control mechanisms required to resolve proactive interference.

Participants were told that no potential patron would ever be so uncool as to use the same door or password as on the previous trial (i.e., stimuli on the current trial are never the same as on the previous trial). However, many patrons slip up by using the door and/or password used two trials back, and the gatekeeper's job is to block their access by pressing a designated key on the keyboard (or allow entry by pressing a different key). We recorded both the accuracy of responses and response time (RT). Decision speed was emphasised by telling participants that only Gatekeepers who can decide both quickly and accurately make the grade and will be employed by the nightclub.

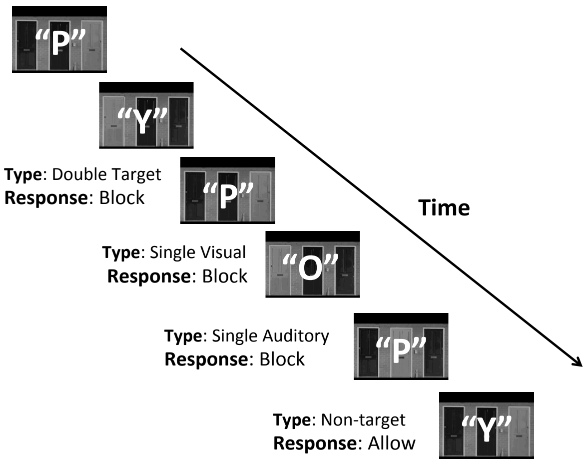

Figure 1. Example of the first 6 trials in a dual Gatekeeper block. White letters indicate auditory stimuli (passwords), and visual targets are the light-grey doors. Visual stimuli were presented in colour, with light-grey regions in red and dark regions in black. No response was required for the first two trials. For each trial thereafter the trial type and correct response are indicated. For the auditory case the correct response sequence would be *Block-Allow-Block-Allow*. For the visual case the correct response sequence would be *Block-Block-Allow-Allow*. *Allow* responses require neither the auditory nor the visual stimulus to match 2 trials back.

Responding in Gatekeeper differs from that required in most dual n-back tasks where separate responses are made to stimuli in each modality (e.g., Jaeggi et al., 2003, 2007, 2008, 2010a, 2010b). Because only a single response is made in Gatekeeper, single target trials – where one stimulus is a target (i.e., it occurred 2-back) and one is not (i.e., it occurred 3-back) – have added interference due to the conflicting individual stimulus-to-response associations.

Here we report the results of an experiment examining a purely dual-task version of Gatekeeper as a measure of mnemonic multitasking ability. Results from this experiment were used to investigate the reliability of performance measures. Previous work (see Jaeggi et al., 2010b for a summary) suggests that measures from the single and dual n-back task can suffer from low reliability. We use a relatively large number of trials and examine the way reliability changes for smaller subsets of trials. Traditional n-back tasks with large stimulus sets have widely varying levels of interference and a response is required only on target trials. Because proactive interference is at a high and constant level and because a response is required on every trial, we predicted greater reliability for Gatekeeper than is typically found in n-back tasks.

We also examined how Gatekeeper performance varies as a function of individual-difference variables commonly thought to affect multitasking, namely age and gender. Poorer multitasking performance in older adults has been attributed to their reduced attentional-control resources, as indicated, for example, by reduced performance in complex span tasks (Watson, Lambert, Miller & Strayer, 2011). Males have also been claimed to be poorer multitaskers than females (e.g., Mäntylä, 2013), although this claim is controversial (Mäntylä & Todorov, 2013; Strayer, Medeiros-Ward, & Watson, 2013).

Experiment

The Gatekeeper task was made available through a link associated with the online version of Strayer and Watson (2012). Strayer and Watson discussed multitasking, and in particular Watson and Strayer's (2010) findings about individual with extraordinary multitasking ability. Readers of Strayer and Watson were invited to attempt the Gatekeeper task in order to test their multitasking ability. The relatively large sample we obtained had good gender balance and a large range of ages, allowing us to look at the effects of these factors on performance, while acknowledging the likely impact of subject-selection effects (e.g., it is likely that only individuals who thought that they had good multitasking ability would attempt the task). We also examined differences among the four within-subject conditions (no targets, visual target only, auditory target only and double target). Because each condition occurred equally often "allow entry" responses were only appropriate on 25% of trials, so we expected to see a bias against them.

We also expected double-target responses to be faster and more accurate than single-target responses because of what is known as "statistical facilitation" (Raab, 1962). Because a correct response in the double target condition can be based on either the visual or auditory modality alone, faster responses can occur because participants can take advantage of chance fluctuations in speed in either modality. Facilitation of the double-target condition occurs both if modalities are processed in parallel or if they are processed serially, as long as the slowest modality is not always processed first. Similarly, we would expect higher accuracy in the double-target condition because a failure to detect a target in one modality can be compensated for by a correct detection in the other modality.

We used Bayesian methods implemented in the BayesFactor package for tests of correlations, t-tests, ANOVA and ANCOVA on measures of Gatekeeper performance (Morey & Rouder, 2012; see Rouder, Morey, Speckman, & Province, 2012, for mathematical details). This approach has a particular advantage in our context where evidence for gender differences is controversial, in that it can provide evidence for a null effect (Wagenmakers, 2007), and so provides an even-handed evaluation of whether or not males and females differ in multitasking ability. Given our relatively large samples the Bayesian approach is also advantageous as it is not subject to the bias in traditional frequentist approaches towards finding all effects significant in large samples (Raftery, 1995).

Method

Participants A total of 245 participants completed the Gatekeeper task. We analysed the data from 222 participants who responded on more than 90% of trials and who did not take extended breaks during the task. Self-report indicated an age range of 13-83 years with 115 females (mean age of 36.7 years) and 107 males (mean 32 years).

Procedure The task took approximately 30 minutes and was administered online using Flash Macromedia (URL: https://psych.newcastle.edu.au/~ae273/GateKeeper/GateKeeper.html). Participants were first asked to record their gender, age, nationality and occupation. They were then told that the task involved acting as the door person at a nightclub and that their task was to block or allow entry to a person trying to enter the club based on whether they were 'cool' or 'uncool'. As illustrated in Figure 1 at the start of each trial one of the three doors turned red, and one of the letters "Y", "P" or "O" were spoken through the computer speakers in a female voice.

Auditory and visual stimuli were selected randomly and independently with the constraint that stimuli never repeated. Hence, no target, visual target only, auditory target only and double target trials occurred with equal frequency on average. Responses were made via the keyboard using the "z" and "/" to allow or block entry, with the mapping counterbalanced across participants. Participants were asked to respond as quickly and accurately as possible. A trial terminated with the response or after 2.5sec if no response was given, and a new trial would begin after a 1sec interval.

Participants were told that the initial two entries on each block of trials were the manager and the barman, who were allowed entry. Thus, they did not have to respond, but had to nevertheless remember the doors and passwords used. Before starting the experiment, participants performed two 12-trial single-task (visual only then auditory only) blocks. Feedback was provided at the top of the screen indicating whether responses were correct or incorrect. They then performed two practice dual-task blocks of 27 trials, the first with feedback and the second without. Practice was followed by 16 experimental dual-task blocks of 27 trials each without feedback. Participants were required to press the space key to move on to the next block, but could only do so after a mandatory 1-minute break between blocks had elapsed. At the conclusion of the task participants were given feedback about their overall performance.

Results

We quantified response-choice (i.e., "block" vs. "allow") data both in terms of the overall accuracy (i.e., percentage of correct responses) and using signal detection theory measures (Stanislaw & Todorov, 1999).

Split-half reliabilities were calculated using the Spearman-Brown formula for RT based and accuracy based performance statistics. Table 1 shows split-half reliabilities of data from n = 400 trials (i.e., all 16 blocks of 25 experimental trials), and subsets (randomly selected from all experimental trials) of n = 200, 100 and 50 trials. Reliabilities were averaged over 100 random splits, and with this number of splits the standard error of the mean was negligible. Table 1 shows that for most measures reliability was very good for 400 and 200 trials, and in some cases this was even the case for lesser numbers.

Table 1. Average Spearman-Brown split-half reliabilities for: PC = overall percentage correct, MRT = overall mean RT. d' = signal detection sensitivity. Subscripts indicate statistics calculated based on double-target (av), and auditory (a) or visual (v) single target trials (relative to non-target trials in the case of d') and non-target (n) trials.

n	400	200	100	50
PC	0.98	0.95	0.91	0.83
d'_{av}	0.9	0.85	0.77	0.67
d'_{a}	0.94	0.89	0.81	0.72
d'_{v}	0.94	0.9	0.84	0.76
MRT	0.99	0.98	0.97	0.94
MRT_{av}	0.97	0.94	0.89	0.80
MRT_{a}	0.97	0.94	0.88	0.79
MRT_{v}	0.97	0.95	0.9	0.81
MRT_{n}	0.97	0.95	0.9	0.81

In ANOVA and ANCOVA analyses we fit all possible hierarchical models, that is, all additive combinations of main effects and interactions with the restriction that when higher-order terms are included so are all of their lower-order constituents. The restriction corresponds to a Type-II sums of squares approach in traditional ANOVA. We first report the best model, that is, the model with the strongest evidence indicated by the largest Bayes factor (BF) relative to the intercept-only (grand mean) model.

We then examined the strength of evidence for each term based on the BF for a simpler model with the term dropped relative to the best model. For example, BF = 0.1 indicates the data increase the odds in favour of the inclusion of the term in the best model by a factor of 10 (i.e., the inverse of 0.1). Jeffreys (1961, p. 432) described a factor of 10 or larger as indicating strong evidence (i.e., BF ≤ $^1/_{10}$), whereas a factor of less than 3 (i.e., BF > $^1/_3$) provides equivocal evidence (or, more colourfully, evidence "barely worth mentioning"), with values in between (i.e., $^1/_{10}$ > BF ≤ $^1/_3$) indicating substantial evidence.

Figure 2 plots overall accuracy and mean RT for correct responses as a function of age for male and female participants. For further analysis we removed 10 participants with overall accuracy less than 55% (triangles in Figure 2) because they were likely responding randomly or had misunderstood the response instructions. The remaining 212 participants had the same age range as the full sample, with 109 females (mean age 36.6 years) and 103 males (mean age 31.1 years).

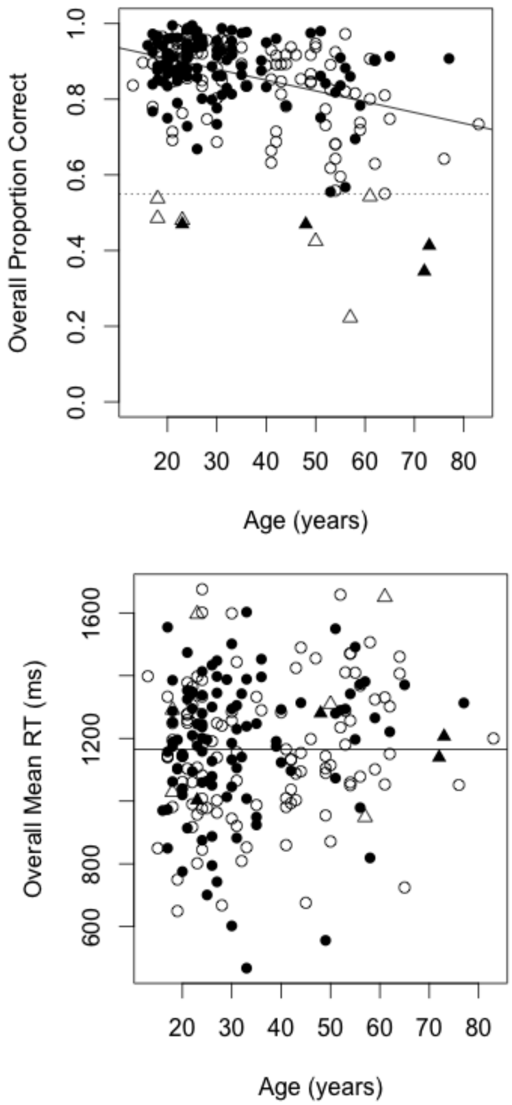

Figure 2. Overall accuracy and mean response (RT) for correct responses as a function of age and gender (female: open circles and triangles; male: solid circles and triangles) for the 222 participants with less than a 10% non-response rate. The horizontal dotted line in the top panel indicates the accuracy cut-off of 55% correct for the subset of 212 participants used in all further analyses (included participants: open and closed circles; excluded participants: open and closed triangles). Solid lines are predictions for the 212 participants for the regression model selected by ANCOVA (age main effect for accuracy and intercept only for RT).

For correct mean RT, the intercept only model (intercept = 1166ms) was selected, with substantial evidence for a null gender effect (BF = 0.16) and equivocal evidence for a null age effect (r = 0.12, BF = 0.69). For accuracy, in contrast, the age main effect model was selected (BF = 2.1×10^9) with accuracy and age having a substantial negative correlation (r = -0.43). There was substantial evidence against the addition of a gender main effect (BF = 0.13) and strong evidence against also adding the interaction (BF = 0.03).

Figure 3 displays the probability of responding "block" and mean RT results for the 2 × 2 within-subject design (auditory target present vs. absent × visual target present vs. absent). Consistent with the predominance of targets, there was strong evidence for a target bias (c = -0.17, BF = 2.7×10^9). We used Bayesian t-tests to examine differences in sensitivity between the single-target and dual-target blocks. There was substantial evidence for a null difference in sensitivity between single visual (d' = 2.23) and auditory (d' = 2.17) conditions (BF = 0.17), and strong evidence for greater sensitivity in the dual condition (d' = 2.95, BF = 8.2×10^{35} and BF = 87.8×10^{48}, respectively).

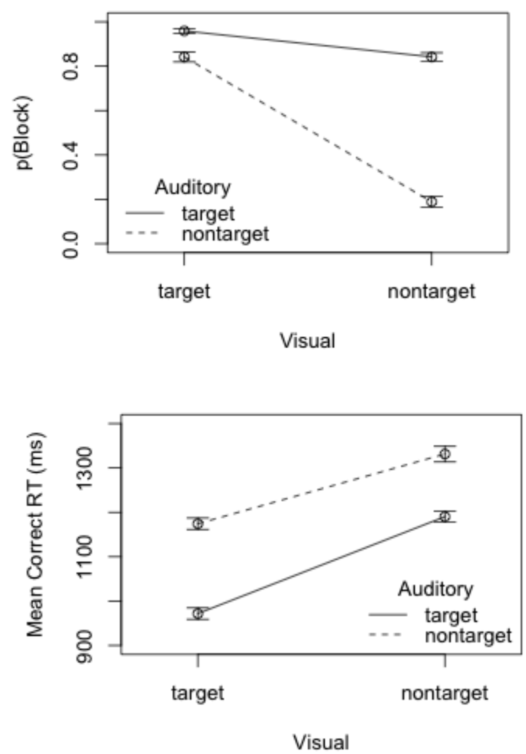

Figure 3. Average probability of detecting a target (responding "Block") and mean correct RT with Morey (2008) bias-corrected within-subject 95% confidence intervals.

Discussion

Overall, measures derived from the Gatekeeper task were quite reliable. This was particularly the case for overall accuracy, which remained highly reliable (> .9) down to 100 trials, and for mean correct RT, which remained highly reliable even with 50 trials. The results for RT contrast somewhat with those of Jaeggi et al. (2010b), who found modest reliability (0.5 – 0.74) based on experiments using

40-66 trials in a dual 2-back task. Jaeggi et al. did not require a non-target response and so could not look at overall accuracy as we did, but did calculate a high-threshold theory sensitivity measure, hit minus false-alarm rate. Our reliabilities for sensitivity results are more similar to, but still somewhat higher than, theirs for similar numbers of trials (0.55 – 0.63). These results suggest that the response method and smaller number of stimuli used in the Gatekeeper task produce more reliable measurements. A particular advantage of responding to all types of stimuli in Gatekeeper is that it enables collection of the two most reliable measures, mean RT and overall accuracy.

As expected, when two targets were present performance was more accurate and faster than when only one target was present. There was also a bias towards target responses ("block") reflecting the predominance of target stimuli. Most importantly, neither gender nor age were correlated with mean RT and there was no indication of a gender difference in accuracy, but accuracy was negatively correlated with age, decreasing at a rate of 0.3% per year. Because of the large number of trials performed by each participant, and consequently the highly reliable nature of the accuracy (0.98) and mean RT (0.99) measures, there is little downward bias in these estimates due to measurement error that might have spuriously lead to a null effect on the theoretically controversial issue of gender effects in multitasking. Further, our Bayesian analysis enabled us to avoid the inability of traditional approaches to confirm a null hypothesis and provide positive and substantial evidence in favour of their being no gender effect.

Taken together, the null effects of gender and the age-related declines in performance on our novel Gatekeeper task represent a dissociation that is consistent with a larger psychological literature on individual differences in attentional control. More specifically, it is well understood that there are age-related breakdowns in working memory capacity and attentional control, the ability to stay on task and to avoid cognitive distractions, an idea that is nicely illustrated by the finding of age-related declines in performance on divided attention tasks (see Watson et al., 2011, for a review) and the ability to handle interruptions (Clapp & Gazzaley, 2012). In this light, one might argue that the older adults who did our Gatekeeper task performed less accurately than the young adults while multitasking due to an impaired ability to control and/or divide their attention, as such attentional abilities are necessary to resolve interference from distractions and attend to stimuli from both modalities. It is also possible that there was an age-related decline in performance on some or all of the components of the Gatekeeper task rather than in participant's ability to perform the components together. Future work might investigate this possibility by measuring single 2-back as well as dual 2-back performance.

In contrast, evidence for gender differences in these same attentional control abilities is more controversial (see Mäntylä & Todorov, 2013; Strayer, Medeiros-Ward, & Watson, 2013). Our findings, based on a large sample and highly reliable measures derived from the Gatekeeper task, strengthen the case for their being little or no gender difference in attentional control and multitasking.

More broadly, our results our results are consistent with individual difference variables that capture underlying variability in working memory capacity and attentional control as being useful in explaining individual differences in multitasking performance (cf., Watson & Strayer, 2010). The Gatekeeper task provides new and reliable measurements that particularly target the impact of attentional control in working memory on multitasking. However, further work is required to explore the relationship of multitasking as measured by the Gatekeeper task and multitasking involving more distinct tasks with separate goals. Borst, Taatgen and van Rijn (2010) noted that participants attempt to merge two tasks into a single task unless they are already practiced and familiar with each task separately, so it is possible that the performance we observed on Gatekeeper will differ from multitasking based on familiar tasks.

Acknowledgments

We acknowledge support from ARC Professorial Fellowship DP110100234 to AH and an ARC Discovery Project DP120102907 to AE and AH.

References

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 363–382.

Clapp, W. C., & Gazzaley, A. (2012). Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiology of Aging*, *33*, 134–148.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity, Behavioral & Brain Sciences, 24, 87-185.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence, 41*, 244-262.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience*, 7, 75–89.

Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., & Gutbrod, K. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. *NeuroImage*, 19, 210–225.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6829–6833.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010a). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010b). The concurrent validity of the N-back task as a working memory measure. Memory, 18(4), 394–412.

Jeffreys, H. (1961). *The Theory of Probability* (3 ed.). Oxford University Press: Oxford.

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

Keppel, G. & Underwood, B.J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1, 153-161.

Mantyla, T. (2013). Gender Differences in Multitasking Reflect Spatial Ability. *Psychological Science*, *24*(4), 514–520.

Mantyla, T., & Todorov, I. (2013). Questioning anecdotal beliefs and scientific Findings: A reply to Strayer, Medeiros-Ward, and Watson (2013). *Psychological Science*, *24*(5), 811–812.

McElree, B. (2001). Working memory and the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 817–835.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*, 296-301.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64.

Morey, R. D. & Rounder, J. (2012). *BayesFactor: An R package for computing Bayes factors in common research designs*. http://bayesfactorpcl.r-forge.r-project.org/

Oberauer, K. (2005). Binding and Inhibition in Working Memory: Individual and Age Differences in Short-Term Recognition. *Journal of experimental psychology: General*, 134, 368–387.

Ophir, E., Nass, C. I., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, *106*, 15583-15587.

Rabb, D. (1962). Statistical facilitation of simple reaction time. *Transaction of the New York Academy of Science, 43,* 574-590.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–164.

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.

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), e54402.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.

Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver dis- traction. *Current Directions in Psychological Science*, **16**, 128-131.

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**, 23-52.

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*, 462-466.

Strayer, D.L., Medeiros-Ward, N., & Watson, J.M. (2013). Gender invariance in multitasking: A comment on Mantyla (2013). *Psychological Science*, *24,* 809- 810.

Strayer, D. L., & Watson, J. M. (2012). Supertaskers and the multitasking brain. *Scientific American Mind*, *23*, 22-29.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.

Watson, J.M., Lambert, A.E., Miller, A.E., & Strayer, D.L. (2011). The magical letters P, F, C, and sometimes U: The rise and fall of executive attention with the development of prefrontal cortex. In K. Fingerman, C. Berg, J. Smith, & T. Antonucci (Eds.), *Handbook of Lifespan Psychology* (pp. 407-436). New York: Springer.

Wallis, C. (2006, March 27). The multitasking generation. Time, *163*.

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479–485.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and performance VIII (pp. 239-257). Hills- dale, NJ: Erlbaum.