Sex differences in verbal working memory performance emerge at very high loads of common neuroimaging tasks

Jessica L. Reed [a,b], Natalie M. Gallagher [a,c], Marie Sullivan [a], Joseph H. Callicott [b], Adam E. Green [a,*]

[a] Department of Psychology, Georgetown University, United States
[b] Clinical Brain Disorders Branch, National Institute of Mental Health, United States
[c] Department of Psychology, Northwestern University, United States

ABSTRACT

Working memory (WM) supports a broad range of intelligent cognition and has been the subject of rich cognitive and neural characterization. However, the highest ranges of WM have not been fully characterized, especially for verbal information. Tasks developed to test multiple levels of WM demand (load) currently predominate brain-based WM research. These tasks are typically used at loads that allow most healthy participants to perform well, which facilitates neuroimaging data collection. Critically, however, high performance at lower loads may obscure differences that emerge at higher loads. A key question not yet addressed at high loads concerns the effect of sex. Thoroughgoing investigation of high-load verbal WM is thus timely to test for potential hidden effects, and to provide behavioral context for effects of sex observed in WM-related brain structure and function. We tested 111 young adults, matched on genotype for the WM-associated COMT-Val[108/158]Met polymorphism, on three classic WM tasks using verbal information. Each task was tested at four WM loads, including higher loads than those used in previous studies of sex differences. All tasks loaded on a single factor, enabling comparison of verbal WM ability at a construct level. Results indicated sex effects at high loads across tasks and within each task, such that males had higher accuracy, even among groups that were matched for performance at lower loads.

Published by Elsevier Inc.

1. Introduction

Sprinters can run faster than marathoners, but the two groups may not appear different if they are only required to jog. Likewise, understanding differences in cognitive abilities depends on experimental paradigms that are challenging enough to distinguish between relevant groups. Verbal working memory (WM; i.e., mentally holding and using verbal information for a short-term goal) is an ability with wide-ranging importance for human cognition (Conway, Kane, & Engle, 2003). Effective assays have been developed to test verbal WM at multiple levels of WM demand (load), including the N-back paradigm in which participants must remember an item presented a specified number of trials (N) previously, where higher values of N reflect higher WM load. Multi-load paradigms are widely used, and currently predominate brain-based WM research (Curtis & D'Esposito, 2003) because they are suited to neuroimaging requirements and constraints. Notably, multi-load paradigms are generally tested at loads that are low enough to allow generally high accuracy among healthy participants (Anderson-Schmidt et al., 2009; Ceaser, Csernansky, & Barch, 2013; Cornelisse, van Stegeren, & Joëls, 2011; Goldstein et al., 2005; Lejbak, Crossley, & Vrbancic, 2011) indicating that they do not challenge the limits of WM capacity (i.e., jogging, not sprinting). Individual differences in WM performance appear to be largely determined by differences in WM capacity (Engle, 2002; Jarrold & Towse, 2006; Just & Carpenter, 1992), so paradigms that challenge capacity by using high WM loads are likely to reveal performance differences that may not emerge at lower loads. Indeed, increasing cognitive demand frequently yields increased individual differences in WM-related tasks (Bielinski & Davison, 1998; Drew & Vogel, 2008; Grabner et al., 2007; Kane & Engle, 2003).

A relatively broad gap in present understanding of WM at high load concerns the effect of sex. Only a small proportion of WM studies have considered sex. Many have not reported performance by sex, or have tested single-sex cohorts (typically all male) (Jonides et al., 1997; Petrides, Alivisatos, Meyer, & Evans, 1993; Veltman, Rombouts, & Dolan, 2003). Neuroimaging and behavioral verbal WM studies that have considered effects of sex have frequently reported no behavioral sex differences

* Corresponding author at: Department of Psychology, Georgetown University, 302C White-Gravenor Hall, 3700 O Street, NW, Box 571001, Washington, DC 20057, United States.
  E-mail address: aeg58@Georgetown.edu (A.E. Green).

(Anderson-Schmidt et al., 2009; Bell, Willson, Wilman, Dave, & Silverstone, 2006; Ceaser et al., 2013; Goldstein et al., 2005; Gur et al., 2012; Haut & Barch, 2006; Koch et al., 2007; Lejbak et al., 2011). However, consistent with the broader literature, these studies have not employed highly demanding WM loads and have shown rather high performance across groups (Anderson-Schmidt et al., 2009; Ceaser et al., 2013; Goldstein et al., 2005; Lejbak et al., 2011). Even when cognitive load was considered to be high, accuracy ranged from about 80–90% (Lejbak et al., 2011) suggesting that load was not demanding relative to available WM capacity. In rare instances, studies using conventional, relatively low levels of WM load have reported sex effects on verbal WM, with better performance found in males in some cases (Longenecker, Dickinson, Weinberger, & Elvevag, 2010) and in females in others (Speck et al., 2000). But these reports are difficult to interpret in the context of a majority of findings of no difference at these same relatively low loads, especially because these reports each relied on single tasks rather than construct variables. Intriguing evidence indicates longer digit span among men (Grossi, Matarese, & Orsini, 1980), even in participants who did not differ on a separate verbal WM task at standard loads (Evans & Hampson, 2015), which is notable because digit span measures the highest load at which a person remains accurate. This evidence suggests the hypothesis that sex differences in verbal WM may emerge at high loads of the multi-load tasks prevalent in neuroimaging and elsewhere, though span tasks reflect a separate underlying cognitive construct (Kane, Conway, Miura, & Colflesh, 2007), leaving this question unresolved. In addition, because span paradigms typically emphasize measurement of the longest total span, rather than multiple discrete levels of load, it is difficult to assess potential load-dependence of sex effects.

Exploring whether and how sex impacts verbal WM is important for understanding the biological and social mechanisms that influence verbal WM, for honing WM intervention/training paradigms, and for avoiding sex-related confounds in WM research designs, including interactions of sex with genotype, development, and clinical status. A more complete understanding requires testing at high loads that challenge WM capacity to identify whether previously undetected differences emerge. The presence of high-load differences in behavioral WM performance would inform interpretation of sex differences in the structure and connectivity of WM-linked frontal and parietal brain regions (Filippi et al., 2013; Frederikse, Lu, Aylward, Barta, & Pearlson, 1999; Ingalhalikar et al., 2014; Sowell et al., 2007), and data indicating differences in WM-related brain activity (Bell et al., 2006; Goldstein et al., 2005; Haut & Barch, 2006; Li, Luo, & Gong, 2010; Speck et al., 2000; Valera et al., 2010). If high-load behavioral tests indicate sex equivalence, it is unlikely that neural sex differences substantially impact WM capacity. By contrast, if behavioral sex differences emerge at high-loads, the observed neural differences that have been shown in previous studies would be implicated.

Summarizing the above, little research has investigated multi-load WM paradigms at high loads, relatively little research has investigated sex differences in verbal WM, and no study to our knowledge has investigated both. Thus, understanding of the relationship of sex to working memory remains incomplete, and extant brain-based findings of WM-relevant sex effects lack sufficient behavioral context for interpretation. The lack of clarity may also be due to the absence of a single "gold-standard" measure of verbal WM. Without a gold standard, a fruitful approach is to administer multiple tasks from which construct-level factors can be derived. It is also possible that genetic variables implicated in WM function, most prominently the COMT Val[108/158]Met polymorphism, which appears to impact WM-related dopaminergic prefrontal function (Winterer & Weinberger, 2004), and shows interactions by gender (Harrison & Tunbridge, 2008), may have confounded prior studies

that did not match sex groups by genotype. While the direction of COMT-by-gender interactions has varied in previous research related to effects on specific phenotypes, working memory was one area in which the association of the Met allele with better cognitive function was found only in males (Harrison & Tunbridge, 2008).

Given the general lack of evidence for sex effects on WM in studies that did not test very high loads, our primary question was whether previously undetected sex differences would emerge at higher loads even when they are not evident at lower loads. We tested 111 healthy women and men matched for COMT genotype on three classic WM tasks using verbal information, and derived a verbal WM factor across these tasks. Critically, we elevated WM load substantially beyond the levels used in previous studies in order to enable a more stringent investigation of between-group differences. Also critically, we were able to test for load-dependence of WM differences by controlling for performance at lower loads.

2. Material and methods

2.1. Participants

Participants were 111 healthy volunteers (females: N = 58; mean age = 21.33 years, SD = 3.15; males: N = 53; mean age = 21.38 years, SD = 3.80) reporting no history of psychiatric illness or psychotropic drug use. Participants were primarily university undergraduates recruited through posted advertisements and through the Georgetown Research Volunteer Program. All participants gave written informed consent, and all recruitment and study procedures were conducted in accordance with the Georgetown University Institutional Review Board. Individuals who participated in the study received credit towards course requirements or monetary compensation ($10 per hour). Female and male participants did not differ on age, frequency of English as a first language, ethnicity, handedness, or self-reported SAT/ACT, which has demonstrated validity in similar cohorts (Cole & Gonyea, 2010), all p > 0.250.

Data were quality-controlled for each task by removing participants from analyses who performed below chance accuracy at the lowest load, or on the lowest two loads of the Sternberg task, which was less difficult than the other tasks. One participant was removed from the Gevins N-back analysis due to average response times less than 100 ms, which indicated a recording error or non-meaningful responding. Slightly different subsets of participants were included for analysis of each task because a small number of participants showed performance that met quality control standards in one task but not another and because data from one (but not the same one) of the three tasks were not properly recorded by the presentation software in seven participants (Cohen N-back: N = 109, 57 female, 52 male; Sternberg task: N = 105, 57 female, 48 male; Gevins N-back: N = 106, 55 female, 51 male). A cross-task sample combined performance for the three tasks, and included participants who were not removed by quality control from any task (N = 98, 53 female, 45 male). The primary goal of this study was to test for performance differences specifically at high WM loads among participants who did not show sex differences at low loads. Among participants who differ at low loads, differences at high loads might reflect differences in WM ability more globally, and might not be attributable to the effect of high load. This was a particularly important consideration given that we sought to extend previous research, which has frequently observed similar performance between sexes at relatively low loads. To examine effects of sex specifically on high levels of WM load, while controlling for performance at low load, we selected female and

male participants matched for low load accuracy (average accuracy across loads 1 and 2). Individuals were paired one-to-one from lists of female and male participants ordered by low load accuracy. In a first pass, we sought to match as many male participants as possible – there were fewer males in our sample – with a female participant on low-load accuracy within 1 standard deviation, selecting the highest-performing unmatched female participant within this range. If this did not yield group-level equivalence (p > 0.05), a second pass was implemented in which matches were required to be within ½ standard deviation of each other. This matching procedure was implemented within each task and in the cross-task sample. Matching yielded groups that did not differ in accuracy at either of the two lowest loads or in average low-load accuracy (Cohen N-back matched sample: N = 104, 52 female, 52 male; Sternberg task matched sample: N = 96, 48 female, 48 male; Gevins N-back matched sample: N = 102, 51 female, 51 male; cross-task matched sample: N = 86, 43 female, 43 male). Thus, matched samples were representative of a lack of sex differences found in the majority of prior studies (Anderson-Schmidt et al., 2009; Cornelisse et al., 2011). Prior to matching, females showed lower accuracy than males on two of the six low-load levels across the three tasks, the lowest load of the Cohen N-back, and the lowest load of the Sternberg task, at an uncorrected $\alpha < 0.05$. Although we did not anticipate low-load differences, they do not present an interpretive hazard in the context of sex-matched analyses (see Supplementary Statistical Analysis), and are consistent with at least one prior study that used a digit N-back in a large sample (Longenecker et al., 2010). The lowest-performing female participants in each task were largely not included in the cross-task matched sample because few appropriate matches were available. Since female participants who showed low performance at low loads generally showed low performance across all loads, sex differences were smaller at both low and high loads in the cross-task matched sample than in the full cross-task sample. Thus, targeting the cross-task matched sample for our primary analysis was a conservative approach for identifying high-load effects of sex and, again, enables inferences about high-load effects in the absence of low-load effects.

Participants were initially recruited for a study of COMT genotype effects on WM (Ihne, Gallagher, Sullivan, Callicott, & Green, 2016). Nonetheless, even the smallest subsample of the dataset (86 cross-task matched participants) provided 80% power to detect a moderate effect size of $d = 0.61$ with a two-tailed $\alpha < 0.05$ for between group t-tests, and a small effect size of $\eta^2 = 0.02$ at $\alpha < 0.05$ for F-tests in a mixed ANOVA model.

Sexes were genotyped for the COMT-Val$^{108/158}$Met polymorphism (Supplementary Methods), which has been linked to individual differences in WM-related dopaminergic brain function (Winterer & Weinberger, 2004), and has shown interactions by gender (Harrison & Tunbridge, 2008). Equal proportions of Val/Val, Val/Met, and Met/Met genotypes were present among females and males in the full data set, X^2 (2, N = 111) = 0.63, $p = 0.727$, and in all subsamples of the data (all $p > 0.250$). The present study did not address COMT genotype-related effects (or lack of effects) on WM. COMT-related findings from this sample are reported by Ihne et al. (2016).

To obtain a measure of general cognitive ability in our sample, we invited participants to perform an online follow-up assessment (Supplementary Methods). The Similarities and Vocabularies subscales of the Multidimensional Aptitude Battery-II (MAB-II; Jackson, 1998) were used to obtain a preliminary measure of verbal ability, including an estimated total Verbal IQ. A version of the redrawn Vandenberg and Kuse Mental Rotation Test (Peters et al., 1995; Vandenberg & Kuse, 1978) was used to assess spatial ability. The primary motivation for this follow-up assessment was to provide greater confidence that the relative abilities of the women and men in our sample were not unrepresentative of the broader population (e.g., that better verbal WM performance among males was not due to unrepresentatively high verbal ability among males in our sample relative to females in our sample). Performance on the follow-up assessment was generally consistent with prior evidence in broader populations (Supplementary Results). Additionally, female participants were asked to provide the start date of their most recent period either via survey at the end of the lab testing session or via online follow-up survey following study participation, which enabled us to preliminarily determine that no phase-of-cycle was overrepresented in our sample (Supplementary Results).

2.2. Working memory tasks

Participants performed three widely-used WM tasks that involve processing verbal information: The Cohen N-back (Cohen et al., 1997), the Sternberg Task (Sternberg, 1966), and the Gevins N-back (Callicott et al., 1998; Gevins et al., 1987). Tasks were programmed using E-Prime Professional 2.08.90 software (Psychology Software Tools, Inc.) and were modified to test participants across four WM loads, including higher loads than those used in prior research that has considered sex differences. Task order was randomized and counterbalanced across all participants. Before each task, participants were given instructions about how to do the task and were familiarized via a shortened practice version.

2.2.1. Cohen N-back

Popularized by Cohen and colleagues (Cohen et al., 1997; Jonides et al., 1997), this task is frequently used in the behavioral and neuroimaging WM literature. In this task, upper-case letters (consonants only) were sequentially and pseudorandomly presented on the screen one at a time for 800 ms, with a 1200-ms inter-stimulus interval. Participants were instructed to press a key to indicate, "Yes," if the letter onscreen was the same as the letter displayed an instructed number (N) screens back, and a different key to indicate, "No," if the letter onscreen was not the same as the letter displayed N screens back (Fig. 1). There were four loads such that N was equal to 2, 3, 4, and 5 on the respective loads (i.e., 2-back, 3-back, 4-back, and 5-back). There were 21 key responses for each load, resulting in a slightly increased number of letters presented for each increasing load: 23 for 2-back (since there are no responses made for the first 2 stimuli), 24 for 3-back, 25 for 4-back, and 26 for 5-back.

2.2.2. Sternberg task

A modified version of the classic and widely-used Sternberg WM task (Sternberg, 1966) was also employed, in which memory sets of simultaneously presented upper-case letters (consonants only) were displayed onscreen, followed by a fixation circle, and then by a single letter probe (Fig. 2). Participants were instructed to press a key to indicate "Yes" if the single letter was in the previous memory set, or a different key to indicate, "No," if it was not. Each memory set was displayed for 4 s, and each fixation circle (delay) was displayed for 6 s. The size of the memory sets varied, containing 4, 6, 8, or 10 letters, with WM load increasing as the number of letters increased. There were 12 trials at each load, all pseudorandomized.

2.2.3. Gevins N-back

We employed a version of the N-back task developed by Gevins and colleagues for electrophysiological studies (Gevins et al., 1987), which we adapted in consultation with its design team, and which is used extensively in neuroimaging (Bertolino et al., 2006; Callicott et al., 1998, 1999; Egan et al., 2001; Mattay et al., 2003). In this task, the numbers 1, 2, 3, and 4 were continually pre-
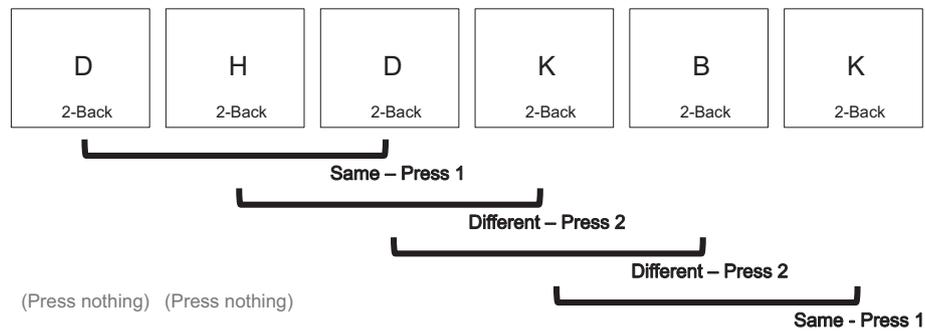
D	H	D	K	B	K
2-Back	2-Back	2-Back	2-Back	2-Back	2-Back

Same – Press 1

Different – Press 2

Different – Press 2

(Press nothing) (Press nothing)

Same - Press 1

Fig. 1. Cohen N-back. Schematic representation of the Cohen N-back task, showing responses for a 2-back load.

FNSVLPGT

(4 sec)

•

(6 sec)

V

(press response key)
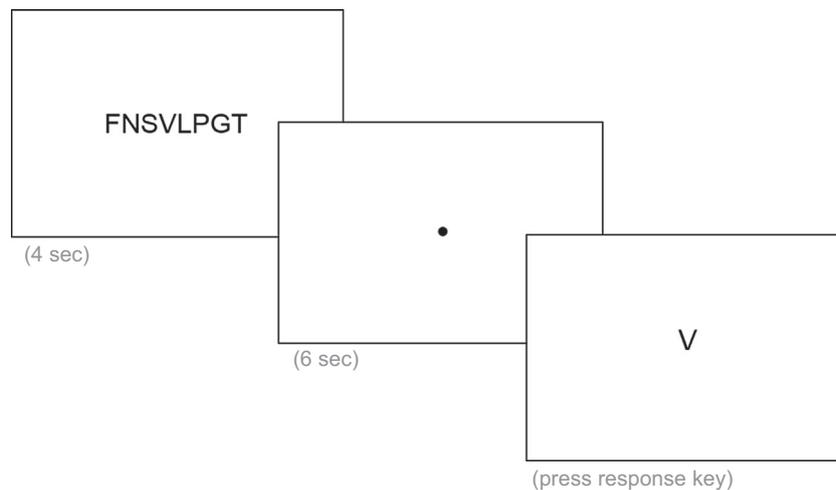
Fig. 2. Sternberg Task. Schematic representation of the Sternberg task, showing an 8-letter load.

sented (one at a time for 500 ms, with a 1500 ms ISI) at the points of a diamond (Fig. 3). The numbers were always presented in the same locations: 1 on the top, 2 on the left, 3 on the right, and 4 on the bottom. Participants were instructed to press the specially marked key on a keyboard corresponding to the number seen N trials back. The response keys were in a diamond figure, matching the locations of the stimuli. Participants were tested at 2-back, 3-back, 4-back, and 5-back loads. For each of the four loads, there were 45 total trials presented in a pseudorandomized continuous sequence.

The Gevins N-back was included as a verbal WM measure because it involves verbal information (i.e., remembering the names of the numbers presented) and can be successfully performed using only this verbal information. However, this task also involves spatial information (i.e., the positions of the letters on the diamond). It was not possible to precisely determine the extent to which participants used verbal vs. spatial information. Encoding of characters (e.g., numbers) is automatic when characters are presented (Shiffrin & Schneider, 1977), whereas attending visuospatial information may be more cue-dependent and more difficult at high loads (Soto, Hodsoll, Rotshtein, & Humphreys, 2008), suggesting that verbal information is processed during this task. Debriefing after the testing session indicated that most participants were explicitly aware of using verbal information, and this is consistent with extensive work we have done previously with this task (Bertolino et al., 2006; Callicott et al., 1998, 1999; Egan et al., 2001; Mattay et al., 2003). Nonetheless, while we deemed the Gevins N-back appropriate to include in the present study, we considered that it would provide weaker evidence concerning verbal WM than the other two tasks. This task has played a prominent role in the cognitive neuroscience and cognitive neurogenet-

ics of WM (Bertolino et al., 2006; Callicott et al., 1998, 1999; Egan et al., 2001; Mattay et al., 2003). Thus, characterizing performance and sex differences on this task at high loads, which has not been previously been tested above 3-back load to our knowledge, was independently of interest.

3. Results

3.1. Accuracy

As expected, accuracy (% correct) decreased as WM load increased. One-way ANOVA for accuracy at the 1st, 2nd, 3rd, and 4th loads averaged across the three tasks revealed a significant effect of load in the cross-task sample (N = 98; quality-controlled for all tasks; Fig. 4 line graph) and the cross-task matched sample (N = 86; sex-matched for performance at low loads across the three tasks; both $p < 0.001$). Corresponding models in the matched samples for each task, and in all quality-controlled participants for each task showed similar effects of load (all $p < 0.001$). Although average accuracy remained above chance for all loads tested, McNemar's repeated measures χ^2 tests indicated that near-chance accuracy (defined as chance ±1 standard deviation of the mean for each load) was much more frequent at high loads than low loads for the cross-task sample (Fig. 4 bar graph), cross-task matched sample, and for each individual task (all $p < .002_{corrected}$[1]). Near-chance accuracy accounted for >25% of participants at the

---

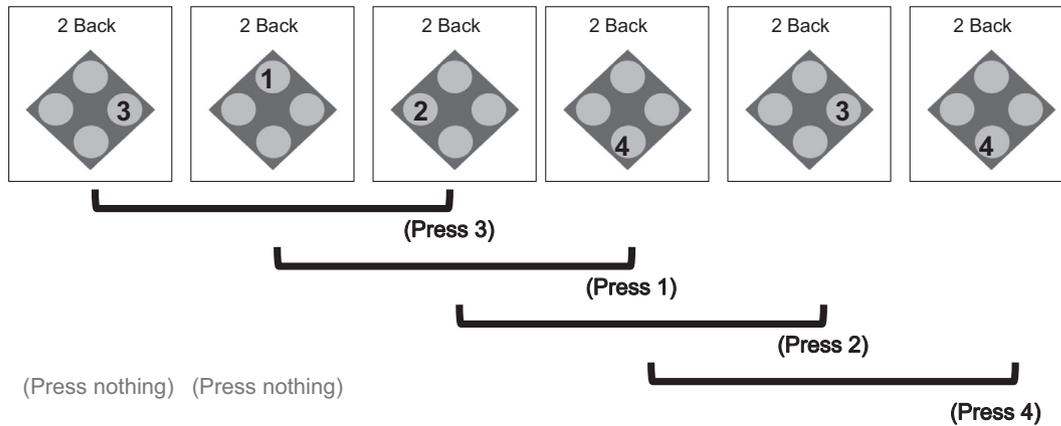[1] Hochberg correction for family-wise false discovery rate used for all corrected p-values (Benjamini & Hochberg, 1995)

Fig. 3. Gevins N-back. Schematic representation of the Gevins N-Back task, showing correct responses for trials at a 2-back load.
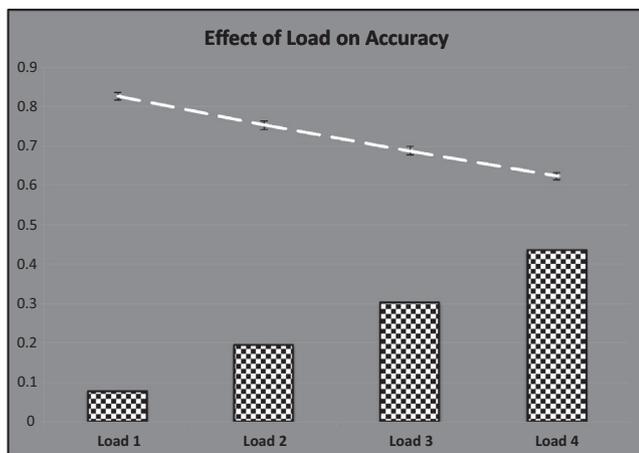

Fig. 4. Effect of verbal WM load on accuracy. Increasing WM load yielded steadily decreasing accuracy. The dashed line represents a steady decrease in accuracy averaged across the three tasks for the 1st, 2nd, 3rd, and 4th levels of WM load (i.e., 2-, 3-, 4-, and 5-back for N-back tasks; 4-, 6-, 8-, and 10-letter memory sets for the Sternberg task). Error bars represent one standard error of the mean. The bar graph indicates a steadily increasing percentage of participants exhibiting near-chance performance (chance ± 1SD) as WM load increased.
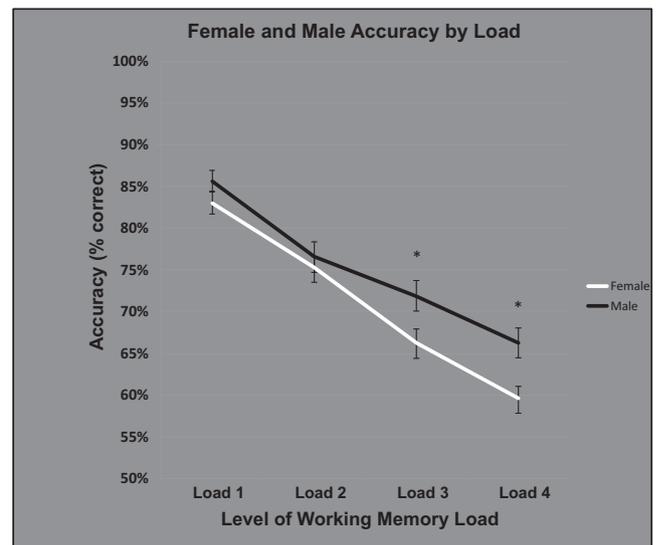

Fig. 5. Female and male accuracy by load. Males showed significantly higher accuracy at the highest WM loads (i.e., 4- and 5-back in the N-back tasks; 8- and 10-letter memory sets in the Sternberg task) among participants who showed no sex differences at the lower WM loads. ANOVA revealed a significant sex-by-load interaction. Error bars indicate one standard error of the mean.

highest load of each individual task, indicating that the high-load manipulation succeeded in challenging WM capacity.

Accuracy scores for each task and each load were analyzed using a Sex X Load X Task ANOVA model (see Supplementary Results). This analysis showed a sex-by-load interaction, $F(3, 82) = 3.27$, $p = 0.025$, $\eta^2 = 0.11$, in which there were greater sex differences at higher loads (Fig. 5). Specifically, at loads three and four, males had higher accuracy than females.

Because working memory is a broad construct, which is not captured by a single gold-standard task, we undertook a factor-analytic approach in order to determine whether performance on the three tasks could be characterized in terms of a working memory factor or factors. Such a cross-task factor would support inferences closer to the construct level of working memory. Accuracy was averaged across the two highest loads on each of the three WM tasks to produce an overall high-load accuracy score for each task. These scores were factor analyzed to determine whether a verbal WM factor could be identified underlying high-load performance on the three tasks. We compared regression factor scores derived from models of high-load performance in the cross-task matched sample, and in the full cross-task sample. Comparing matched participants at high loads was of the greatest interest

because it provided the most direct test of our central question: do sex differences emerge at high loads even when they are not present at low loads? In each model, a single factor with eigenvalue >1 emerged, and all three tasks loaded heavily on this factor. The model for the cross-task matched sample yielded a main factor with eigenvalue = 1.57, 52.23% variance explained, and factor loadings for each task as follows: Cohen N-back = 0.85, Sternberg task = 0.63, Gevins N-back = 0.68. The model for the full cross-task sample yielded a main factor with eigenvalue = 1.63, 54.20% variance explained, and factor loadings for each task as follows: Cohen N-back = 0.85, Sternberg task = 0.65, Gevins N-back = 0.69. Regression factor scores for the main factor in each model were then compared between the sexes via independent samples t-tests. Each comparison revealed significantly higher factor scores among men (cross-task matched sample, $t(84) = 3.08$, $p = .005_{corrected}$, $d = 0.67$, Fig. 6; cross-task sample, $t(96) = 3.95$, $p < .001_{corrected}$, $d = 0.81$).

We considered the main factor in these models to represent verbal WM because all tasks involved processing verbal information. However, because the Gevins N-back involved both verbal and spatial information, it was possible that the availability of
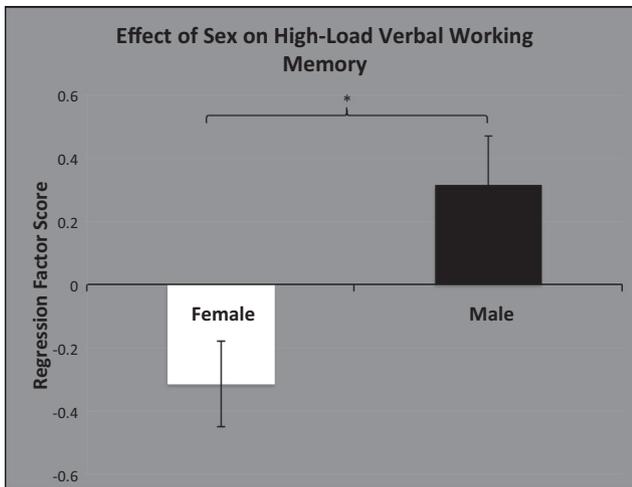
Fig. 6. Effect of sex on high-load verbal working memory. Regression factor scores were derived for a verbal WM factor underlying accuracy at the two highest loads for the three verbal WM tasks. Men exhibited significantly higher high-load factor scores than women among participants who were sex-matched for performance at lower loads. Error bars represent one standard error of the mean.

spatial information affected our results. We thus repeated the factor analysis and comparison of factor scores for the cross-task matched sample and cross-task sample using only the Cohen N-back and Sternberg task, neither of which involved spatial information. Results were very similar to those obtained in the models that included all three tasks: a single factor with eigenvalue >1, heavy loadings of both tasks on this main factor, and sex differences in regression factor scores for this main factor (cross-task matched: $p = .011_{corrected}$; cross-task sample: $p = .002_{corrected}$). Bartlett's test of sphericity indicated that each of the factor models was appropriate (all $p \leqslant 0.001$), and the Kaiser-Meyer Olkin measure indicated that the sample sizes were acceptable (all KMO $\geqslant 0.50$).

For each task, we compared high-load accuracy between women and men among participants sex-matched for performance at low loads, and among all quality-controlled participants. Each of these six comparisons revealed significantly higher accuracy among men (all $p < .05_{corrected}$). A superset of between-group comparisons is reported at uncorrected significance levels in Supplementary Table 1 to provide a broader characterization of the dataset.

Performance on WM tasks is often reported using the sensitivity index (d'). Of the tasks used in our study, d' was only an appropriate measure for the Cohen N-back task (see Supplementary Statistical Analysis). Sex differences in Cohen N-back d' emerged in the same models used to analyze accuracy: high–load performance in matched participants, $t(102) = 2.46$, $p = 0.021_{corrected}$, $d = 0.49$, and all quality-controlled participants, $t(107) = 2.59$, $p = 0.019_{corrected}$, $d = 0.51$).

Working memory has also been measured in prior research using capacity or span calculations, such as Cowan's k (Cowan et al., 2005; Rouder, Morey, Morey, & Cowan, 2011). For this study, Cowan's k was only applicable to the Sternberg task, since the N-back tasks do not use sets of multiple stimuli presented at the same time, as in the Sternberg, and would not be used as measure of capacity. (Rather, the N-backs involve more of the continuous updating component of working memory.) Calculating k for Sternberg task performance in our dataset revealed sex differences consistent with those indicated by our primary analyses, with higher scores in males (see Supplementary Statistical Analysis), $F(1, 96) = 5.73$, $p = 0.019$, particularly at high loads ($t(96) = 2.23$, $p = 0.028$).

The above models addressed our central aim to test performance at high loads and to determine whether sex differences emerged at high loads, even when participants did not differ by sex at low loads. A secondary question concerned characterization of change in sex differences as load increased. We explored this first in the cross-task matched sample because this subsample allowed for the strongest conclusions about performance at high load when differences are not present at low load. ANOVA models confirmed better performance among males, and revealed sex-by-load interaction effects indicating greater sex differences at higher loads (Supplementary Results).

3.2. Response time

Accuracy was the primary outcome measure for our tasks, and we emphasized accuracy over speed in our task instructions. Thus, there was less theoretical motivation to investigate response time (RT) than accuracy and far fewer models were tested for RT. As expected, one-way ANOVAs for average RT at the 1st, 2nd, 3rd, and 4th loads across the three tasks revealed a significant effect of load in the cross-task sample and the cross-task matched sample (both $p < 0.001$). Exploratory tests revealed that neither average RT nor the association between RT and accuracy (Pearson's r correlation coefficient) differed between sexes at high loads, across all loads, or at any individual load on any task at an uncorrected $\alpha = 0.05$ significance level.

4. Discussion

We modified three classic WM tasks to become quite difficult, including higher loads than those tested in previous studies of sex differences. All tasks loaded on a single verbal WM factor and comparison of factor scores indicated better performance among males at high loads. Indeed, we found better performance among males at the high loads of each task even when controlling for performance at lower loads that were similar to the loads tested in prior studies. These results demonstrate the utility of testing WM tasks at high loads to elicit individual differences that may not otherwise emerge.

These findings provide new behavioral context for prior observations of sex differences at the neural level. Though it is unfortunately not possible to determine the influence of publication bias, reported studies that have tested for sex differences in WM-related brain structure and function have generally found brain-based differences (Bell et al., 2006; Filippi et al., 2013; Frederikse et al., 1999; Goldstein et al., 2005; Haut & Barch, 2006; Ingalhalikar et al., 2014; Li et al., 2010; Sowell et al., 2007; Speck et al., 2000; Valera et al., 2010). Anderson-Schmidt et al. (2009) reported overlapping WM-related activation between sexes, but did not compare amplitude of activation. A recent study, which parsed network connectivity in the resting brain by the cognitive functions resting-state networks subserve, reported substantial sex differences in the WM network among a relatively large group of participants (Filippi et al., 2013). A fairly consistent finding among studies of sex differences in functional activity during visually-presented verbal WM tasks similar to those used in the present study, is that males exhibit greater activation in WM-related prefrontal and parietal regions (Bell et al., 2006; Koch et al., 2007; Li et al., 2010; Speck et al., 2000; Valera et al., 2010). Though greater neural activation is sometimes interpreted to indicate less efficiency, this interpretation is based on misguided or absent theory (Poldrack, 2015), and does not account for considerable extant neuroimaging data (e.g., (Gray, Chabris, & Braver, 2003; Green, Kraemer, Deyoung, Fossella, & Gray, 2013; Osaka et al., 2003)). Indeed, in the case of WM, there is evidence indicating that greater

activation may be advantageous in at least some cases. Olesen, Westerberg, and Klingberg (2004) trained participants extensively on WM tasks and demonstrated that improvements in performance were associated with increasing activity in WM-associated prefrontal and parietal regions, which were observed at low WM loads during scan sessions. Gray et al. (2003) found that greater verbal WM-related brain activation was associated with higher fluid intelligence. It is worth noting that, because neuroimaging paradigms have generally used relatively low WM loads, it is possible that observed sex differences in activation might be more robust at higher load. However, implementing high-load WM during neuroimaging poses substantial challenges for data collection and interpretation, especially for collecting enough accurate responses to power neuroimaging analyses. Thus, given that the preponderance of available neuroimaging data suggest brain-based sex differences, a more thoroughgoing assessment of behavioral sex effects on WM is timely. Neural differences do not always have clear impacts on behavior, but the present data support the hypothesis that previously observed sex effects on the brain may have effects on verbal WM performance, which have not been adequately detected by studies that have not explored very high loads.

The healthy human brain has tremendous capacity for WM (Anderson-Schmidt et al., 2009). Thus, even if brain-based differences bear on verbal WM capacity, the demands used in prior studies of verbal WM may not have exceeded the levels at which most people of both sexes can perform well (i.e., jogging, not sprinting). It is also possible that broader modal effects of sex on information processing help to explain prior findings. Undergraduate psychology students often learn that women tend to be better at verbal tasks while men tend to be better at spatial tasks. Though this appears to be an oversimplification (Hyde & Linn, 1988; Weiss, Kemmler, Deisenhammer, Flieschhacker, & Delazer, 2003), there is considerable evidence for at least some instances of this conventional verbal/spatial sex differences account (Halpern & LaMay, 2000; Johnson & Bouchard, 2007). Indeed, WM appears to be better in men than in women when the information is spatial (e.g., (Kaufman, 2007; Lejbak et al., 2011; Pauls, Petermann, & Lepach, 2013)). This evidence is strongest for tasks of mental rotation in which a representation of an object must be held in mind and rotated in space (see Voyer, Voyer, and Bryden (1995) for review). It is thus possible that sex differences at relatively low loads have been less evident for verbal than spatial WM because spatial WM differences are augmented by a putative male advantage in spatial processing (Voyer et al., 1995). It is also possible that a female advantage in verbal processing, suggested by substantial prior research (Halpern & LaMay, 2000; Johnson & Bouchard, 2007), may compensate for a male advantage in WM, making verbal WM performance appear similar at relatively low loads. The present findings suggest that sex may have effects on WM that extend beyond the processing of explicitly spatial stimuli in WM (Voyer et al., 1995) to at least some forms of verbal information processing. This would be consistent with evidence that underlying neural systems, cognitive mechanisms, and genetic structures are substantially shared between spatial and verbal WM (Ando, Ono, & Wright, 2001; Curtis & D'Esposito, 2003).

Sex differences in high-load WM may also be related to effects of stress. Our tasks became extremely difficult, and the N-back tasks may have been particularly stressful because they involved continually changing stimuli and the need for quick responses. Stress has been shown to improve WM performance in males but either not affect performance (Cornelisse et al., 2011) or impair performance (Schoofs, Pabst, Brand, & Wolf, 2013) in females. In a rat model, estrogen was shown to increase a prefrontal stress response associated with decreased WM performance (Shansky et al., 2004). Direct measures of stress (e.g., blood cortisol) associ-

ated with high-load WM will be required to address this possibility in future research. Additionally, the appraisal of stress and the perception of one's own cognitive abilities during challenging and stressful situations could play a role in the sex differences found. For example, stereotype threat may contribute towards impairment of performance in females (Schmader, Forbes, Zhang, & Mendes, 2009). Cognitive stress appraisal and also concept of one's competence has been shown to be associated with WM in females but not in males (Zandara et al., 2016). Another study showed that while negative emotion impaired WM in both males and females, there were differences in fMRI activation in which females showed greater activation in emotion-related brain regions (Koch et al., 2007). While our data do not allow us to address these factors directly, it is possible that there was some influence of a reaction to stress or frustration that impacted females more than males. In our analyses of Cowan's k measurements of capacity, we found that k did vary across loads, which would not be expected for this overall measure of capacity. In Cowan's model, WM capacity is thought to be a construct that remains constant across varying set sizes (Cowan et al., 2005; Rouder et al., 2011). However, our data showed variation in k across loads, particularly for females, such that there was a decrease in k at the highest load. This could potentially signify an emotional stress response interfering with performance, since the k value itself would not be expected to change. Perhaps at this load, the performance does not reflect true WM capacity due to emotion or stress negatively affecting task performance, thus resulting in lower accuracy and the apparent decrease in k. Further research is needed to investigate these types of factors and potential effects underlying sex differences of WM at high loads.

Extensive psychometric individual differences research has richly characterized the relationship of WM, including verbal WM, to other intelligence-related abilities (Conway et al., 2003; Engle, 2002; Jarrold & Towse, 2006). This work has generally not considered effects of sex, and has almost exclusively employed WM "span" tasks that measure the largest number of items an individual can retain in WM. As noted by Jarrold and Towse (2006), individual differences research on WM has rarely used N-back multi-load designs. Span tasks draw on a different underlying cognitive construct than the multi-load N-back WM paradigm (Kane et al., 2007). Thus, it is difficult to interpret data obtained via multi-load paradigms, including the great majority of neuroimaging data, in the context of the extant individual differences literature centered on span tasks. Also, because span tasks typically do not emphasize multiple measurements at discrete levels of WM load, they often do not afford clear assessment of load-dependence (e.g., the extent to which differences are present at all loads or only emerge at high loads). The present research demonstrates the potential for using multi-load tasks at high load to test individual differences. Future work should seek to enrich the psychometric characterization of individual differences in high-load performance with respect to a broader suite of cognitive abilities to provide a richer interpretive context for brain-based data.

5. Conclusions

In a cohort of healthy young adults, we identified effects of sex on accuracy at high verbal WM loads for each of three classic neuroimaging tasks, and for a cross-task verbal WM factor. Differences at high loads remained when participants were matched for performance at lower loads. These findings inform interpretation of brain-based evidence that sex affects the structure and function of WM-associated brain regions, especially in prefrontal and parietal cortex. More broadly, the findings highlight the consideration

that effects of sex and other individual and group differences may be load-dependent. Thus, exploring cognitive abilities at levels of difficulty high enough to challenge capacity may reveal effects that are otherwise hidden or underappreciated. High-load behavioral implementations of neuroimaging tasks may be helpful to expand the interpretive context of brain-based data.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandc.2017.01.001.

References

Anderson-Schmidt, H., Jogia, J., Fast, K., Christodoulou, T., Haldane, M., Kumari, V., & Frangou, S. (2009). No gender differences in brain activation during the N-back task: An fMRI study in healthy individuals. *Human Brain Mapping, 30*(11), 3609–3615.

Ando, J., Ono, Y., & Wright, M. J. (2001). Genetic structure of spatial and verbal working memory. *Behavior Genetics, 31*(6), 615–624.

Bell, E. C., Willson, M. C., Wilman, A. H., Dave, S., & Silverstone, P. H. (2006). Males and females differ in brain activation during cognitive tasks. *Neuroimage, 30*(2), 529–538.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*, 289–300.

Bertolino, A., Blasi, G., Latorre, V., Rubino, V., Rampino, A., Sinibaldi, L., ... Dallaopiccola, B. (2006). Additive effects of genetic variation in dopamine regulating genes on working memory cortical activity in human brain. *Journal of Neuroscience, 26*(15), 3918–3922.

Bielinski, J., & Davison, M. L. (1998). Gender differences by item difficulty interactions in multiple-choice mathematics items. *American Education Research Journal, 35*, 455–476.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., ... Weinberger, D. R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex, 9*(1), 20–26.

Callicott, J. H., Ramsey, N. F., Tallent, K., Bertolino, A., Knable, M. B., Coppola, R., ... Weinberger, D. R. (1998). Functional magnetic resonance imaging brain mapping in psychiatry: Methodological issues illustrated in a study of working memory in schizophrenia. *Neuropsychopharmacology, 18*(3), 186–196.

Ceaser, A., Csernansky, J. G., & Barch, D. M. (2013). COMT influences on prefrontal and striatal blood oxygenation level-dependent responses during working memory among individuals with schizophrenia, their siblings, and healthy controls. *Cognitive Neuropsychiatry, 18*(4), 257–283.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*(6625), 604–608.

Cole, J. S., & Gonyea, R. M. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education, 51*(4), 305–319.

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.

Cornelisse, S., van Stegeren, A. H., & Joëls, M. (2011). Implications of psychosocial stress on memory formation in a typical male versus female student sample. *Psychoneuroendocrinology, 36*(4), 569–578.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42–100.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*(9), 415–423.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience, 28*(16), 4183–4191.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., ... Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences USA, 98*(12), 6917–6922.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Evans, K. L., & Hampson, E. (2015). Sex differences on prefrontally-dependent cognitive tasks. *Brain and Cognition, 93*, 42–53.

Filippi, M., Valsasina, P., Misci, P., Falini, A., Comi, G., & Rocca, M. A. (2013). The organization of intrinsic brain activity differs between genders: A resting-state fMRI study in a large cohort of young healthy subjects. *Human Brain Mapping, 34*(6), 1330–1343.

Frederikse, M. E., Lu, A., Aylward, E., Barta, P., & Pearlson, G. (1999). Sex differences in the inferior parietal lobule. *Cerebral Cortex, 9*(8), 896–901.

Gevins, A. S., Morgan, N. H., Bressler, S. L., Cutillo, B. A., White, R. M., Illes, J., ... Zeitlin, G. M. (1987). Human neuroelectric patterns predict performance accuracy. *Science, 235*(4788), 580–585.

Goldstein, J. M., Jerram, M., Poldrack, R., Anagnoson, R., Breiter, H. C., Makris, N., ... Seidman, L. J. (2005). Sex differences in prefrontal cortical brain activity during fMRI of auditory verbal working memory. *Neuropsychology, 19*(4), 509–519.

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage, 38*(2), 346–356.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*(3), 316–322.

Green, A. E., Kraemer, D. J., Deyoung, C. G., Fossella, J. A., & Gray, J. R. (2013). A gene-brain-cognition pathway: Prefrontal activity mediates the effect of COMT on cognitive control and IQ. *Cerebral Cortex, 23*, 552–559.

Grossi, D., Matarese, V., & Orsini, A. (1980). Sex differences in adults' spatial and verbal memory span. *Cortex, 16*(2), 339–340.

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., ... Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21. *Neuropsychology, 26*(2), 251–265.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*(2), 229–246.

Harrison, P. J., & Tunbridge, E. M. (2008). Catechol-O-methyltransferase (COMT): A gene contributing to sex differences in brain function, and to sexual dimorphism in the predisposition to psychiatric disorders. *Neuropsychopharmacology, 33*(13), 3037–3045.

Haut, K. M., & Barch, D. M. (2006). Sex influences on material-sensitive functional lateralization in working and episodic memory: Men and women are not all that different. *Neuroimage, 32*(1), 411–422.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological bulletin, 104*(1), 53–69.

Ihne, J., Gallagher, N. M., Sullivan, M., Callicott, J. H., & Green, A. (2016). Is less really more: Does a prefrontal efficiency genotype actually confer better performance when working memory becomes difficult? *Cortex, 74*, 79–95.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., ... Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences USA, 111*(2), 823–828.

Jackson, D. N. (1998). *Multidimensional aptitude battery-II: Manual)*. Port Huron, MI: Sigma Assessment Systems.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139*(1), 39–50.

Johnson, W., & Bouchard, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence, 35*, 23–39.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*(4), 462–475.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(3), 615–622.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47–70.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35*(3), 211–223.

Koch, K., Pauly, K., Kellermann, T., Seiferth, N. Y., Reske, M., Backes, V., ... Habel, U. (2007). Gender differences in the cognitive control of emotion: An fMRI study. *Neuropsychologia, 45*(12), 2744–2754.

Lejbak, L., Crossley, M., & Vrbancic, M. (2011). A male advantage for spatial and object but not verbal working memory using the n-back task. *Brain and Cognition, 76*(1), 191–196.

Li, T., Luo, Q., & Gong, H. (2010). Gender-specific hemodynamics in prefrontal cortex during a verbal working memory task by near-infrared spectroscopy. *Behavioural Brain Research, 209*(1), 148–153.

Longenecker, J., Dickinson, D., Weinberger, D. R., & Elvevag, B. (2010). Cognitive differences between men and women: A comparison of patients with schizophrenia and healthy volunteers. *Schizophrenia Research, 120*(1–3), 234–235.

Mattay, V. S., Goldberg, T. E., Fera, F., Hariri, A. R., Tessitore, A., Egan, M. F., ... Weinberger, D. R. (2003). Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences USA, 100*(10), 6186–6191.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75–79.

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). The neural basis of individual differences in working memory capacity: An fMRI study. *Neuroimage, 18*(3), 789–797.

Pauls, F., Petermann, F., & Lepach, A. (2013). Gender differences in episodic memory and visual working memory including the effects of age. *Memory, 21*(7), 857–874.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Sciences USA, 90*(3), 878–882.

Poldrack, R. A. (2015). Is "efficiency" a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience, 11*, 12–17.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review, 18*(2), 324–330.

Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35*(5), 584–596.

Schoofs, D., Pabst, S., Brand, M., & Wolf, O. T. (2013). Working memory is differentially affected by stress in men and women. *Behavioural Brain Research, 241*, 144–153.

Shansky, R. M., Glavis-Bloom, C., Lerman, D., McRae, P., Benson, C., Miller, K., ... Arnsten, A. F. (2004). Estrogen mediates sex differences in stress-induced prefrontal cortex dysfunction. *Molecular Psychiatry, 9*(5), 531–538.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127–190.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*(9), 342–348.

Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., ... Toga, A. W. (2007). Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cerebral Cortex, 17*(7), 1550–1560.

Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *NeuroReport, 11*(11), 2581–2585.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736), 652–654.

Valera, E. M., Brown, A., Biederman, J., Faraone, S. V., Makris, N., Monuteaux, M. C., ... Seidman, L. J. (2010). Sex differences in the functional neuroanatomy of working memory in adults with ADHD. *American Journal of Psychiatry, 167*(1), 86–94.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604.

Veltman, D. J., Rombouts, S. A., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: An fMRI study. *Neuroimage, 18*(2), 247–256.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250–270.

Weiss, E., Kemmler, G., Deisenhammer, E., Flieschhacker, W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences, 35* (4), 863–875.

Winterer, G., & Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences, 27*(11), 683–690.

Zandara, M., Garcia-Lluch, M., Pulopulos, M. M., Hildalgo, V., Villada, C., & Salvador, A. (2016). Acute stress and working memory: The role of sex and cognitive stress appraisal. *Physiology & Behavior, 164*, 336–344.