Brain and Cognition 113 (2017) 56–64

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

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

ARTICLE INFO

Article history:
Received 21 June 2016
Revised 5 October 2016
Accepted 2 January 2017

Keywords:
N-back
Sternberg task
Female
Male
Prefrontal cortex

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).

http://dx.doi.org/10.1016/j.bandc.2017.01.001
0278-2626/Published by Elsevier Inc.

(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