The Development of Working Memory: Sex Differences in Accuracy and Reaction Times

Paul Ibbotson & Ernesto Roque-Gutierrez

To cite this article: Paul Ibbotson & Ernesto Roque-Gutierrez (2023): The Development of Working Memory: Sex Differences in Accuracy and Reaction Times, Journal of Cognition and Development, DOI: 10.1080/15248372.2023.2178437

To link to this article: https://doi.org/10.1080/15248372.2023.2178437

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 02 Mar 2023.

Submit your article to this journal

Article views: 95

View related articles

View Crossmark data
The Development of Working Memory: Sex Differences in Accuracy and Reaction Times

Paul Ibbotson and Ernesto Roque-Gutierrez
The Open University, Milton Keynes, UK

ABSTRACT
Small but robust differences in cognition exist between the sexes in adult populations. Studying sex differences in children’s cognition can bring insights into when, where and how these differences might emerge in development. Here, we focus on differences in working memory because of its importance in underpinning a wide range of complex cognitive tasks and developmental outcomes for children. Using two levels of difficulty on a standard test of working memory (N-Back), data from 104 6- to 7-year-olds in Cuba showed that boys have quicker reaction times, but girls provide more accurate responses. With a comparable true positive rate between boys and girls, the sex differences in both accuracy and reaction times were limited to false-positive responses. Sex differences were consistent across levels of task difficulty and persisted after speed-accuracy trade-offs were considered. We argue that avoiding false positives requires a particularly strong role for inhibitory control and that this emerges in development according to a different maturational schedule for girls than it does for boys, underpinned by quantitative and qualitative differences in the development of brain areas that support this function.

Data from adults reveal that men and women do not significantly differ on most cognitive traits, and where they do, the differences are often small (Halpern, 1992). However, on average, women have higher verbal abilities, and men have higher spatial and mathematical abilities (Kimura, 1999). With respect to memory, sex difference differentiation is more equivocal, at least in adults, with some studies reporting a significant difference (e.g., McGivern et al., 1998; Voyer, Saint Aubin, Altman, & Gallant, 2021) and others failing to find a difference (e.g., Freides & Avery, 1991; McCarty, Siegler, & Logue, 1982). The picture is further complicated by the different ways in which memory is measured across studies (e.g., N-back, Complex Span, Test of Memory and Learning (TOMAL)) and which subcomponents of memory are targeted by the test (e.g., visuospatial, auditory, and verbal).

Studying sex differences in children’s cognition can bring a valuable perspective to this issue. The different factors that have been put forward to explain sex differences in cognition – societal barriers of discrimination, discouragement, differential child-rearing time investment, a predisposition of talents, temperaments, and interests – make different
predictions as to when, where and how differences might emerge in development. For example, Galsworthy, Dionne, Dale, and Plomin (2000) examined the genetic and environmental origins of sex differences in verbal and non-verbal cognitive ability in over 3000 2-year-old twin pairs. Girls scored significantly higher on both measures ($p < .0001$), although sex only accounted for approximately 3% of the variance in verbal ability and 1% of the variance in non-verbal cognitive ability. The authors concluded that genetic and environmental influences differ for girls and boys for early verbal but not non-verbal development (Galsworthy et al., 2000). Ardila, Rosselli, Matute, and Inozemtseva (2011) found that sex differences accounted for a similar amount of the variance in performance (1–3%) when they gave a battery of neuropsychological tests to 788 monolingual children aged 5–16-years-old from Mexico and Colombia. In this study, on average, boys outperformed girls on measurements of language (both expression and production and contrary to the adult norm), and in spatial and visual abilities, with girls outperforming boys only on a tactile task (e.g., identifying objects by touch).

Here, we focus on working memory because of its importance in underpinning a wide range of complex cognitive tasks and developmental outcomes for children, including reading comprehension (Cain, 2006), reasoning (García-Madruga et al., 2007), arithmetic calculations (Deschuyteneer, Vandierendonck, & Muylraert, 2006), mathematical problem solving (Passolunghi & Pazzaglia, 2004), academic performance (Alloway & Alloway, 2010) and fluid intelligence (Friedman et al., 2006). In our everyday lives, working memory is employed when we need to coordinate the topic of conversation, chunk numerical operations to perform basic arithmetic, find objects in space, and so it is a basic part of normal cognitive and social function. The essential underlying cognitive process that supports this range of behavior is one that enables active maintenance and regulation of a limited amount of task-relevant information (Baddeley & Logie, 1999).

With respect to working memory sex differences, the general picture that has emerged thus far from the developmental literature is that boys are quicker to respond but girls are more accurate, again, with subtle variation based on which test is used, what components of memory the test targets and for which ages (e.g., Lowe, Mayfield, & Reynolds, 2003; Lynn & Irwing, 2008; Pelegrina et al., 2015; Voyer, Voyer, & Saint-Aubin, 2017; Vuontela et al., 2003). In a meta-analysis of 180 effect sizes from healthy males and females drawn from 98 samples ranging in mean age from 3 to 86 years, Voyer et al. (2017) showed a small but significant male advantage for visual-spatial working memory (mean $d = 0.155$, 95% confidence interval = 0.087–0.223). Moreover, they found an increase in the magnitude of sex differences with age, with a significant difference emerging around puberty. Vuontela et al. (2003) tested the visuospatial working memory of 66 schoolchildren from Helsinki, Finland, aged 6–13 years of age. They found that boys performed the task significantly faster than the girls, but girls made fewer errors than boys. Pelegrina et al. (2015) found the same basic pattern with 3722 Spanish school children aged 7–13 years, with girls outperforming boys although taking more time to respond on a test of linguistic working memory (see Lowe et al., 2003; Lynn & Irwing, 2008 for a similar pattern with different tests and slightly different age ranges).

Sex differences in children’s spatial cognition have also been investigated in the context of drawing and reading. For example, Lange-Küttner and Ebersbach (2013) investigated children’s ability to draw cubes in two- and three-dimensions and found that boys were more likely to attend to the projective appearance of the cubes (the visual object), while girls
were more likely to explore aspects of the identity of the cubes (the real object). In girls, the best predictor of drawing was their score on the Mental Rotation Test (MRT). For boys, their drawing was best predicted by the MRT reaction time and the Embedded Figures Test (EFT), with performance on the EFT itself being highly correlated with working memory. Huestegge, Heim, Zettelmeyer, and Lange-Küttner (2012) demonstrated that only in boys, visual abilities, and in particular Visual Short-Term Memory (VSTM) skills, were closely related to reading performance. The authors argue this is evidence that reading development in boys generally relies on the use of VSTM in a more pronounced way than do girls, which in turn is linked with a whole array of visual cognitive processes.

The current study

The present study seeks to establish whether the general pattern – boys quicker, girls more accurate – generalizes to a test of visual working memory with no spatial-location or linguistic content, a datum that is missing from our understanding of the developmental timeline. For example, Vuontela et al. (2003) relied on children remembering the location of objects in space; Pelegrina and colleagues (2015) relied on children remembering letters (linguistic stimuli); Lowe and colleagues (2003) and Lynn & Irwing, (2008) likewise did not have a pure visual measure of working memory. Using stimuli that varied in both shape, place and familiarity, Lange-Küttner and Küttner (2015) found that shape and place learning accuracy increased only during repeated identical memory sets and that whereas shape reaction times were initially fast and then slowed down, place reaction times were slow and speeded up.

We know that one of the most robust findings from the adult literature is that women, on average, have higher verbal abilities, and men have higher abilities on most measures of spatial cognition (see Voyer, Postma, Brake, & Imperato-McGinley, 2007 for a female advantage in object location memory tasks above the age of 13). Therefore, our rationale is that by using a visual, non-linguistic measure, it allows us to neutralize both of these influences on performance and ask the question whether sex differences exist in working memory above and beyond these factors. While there is spatial information in the task we employ – in the sense that the images have a two-dimensional form – memory for where these objects are in relation to the background does not contribute to a participant’s accuracy in the test (more details in Methods).

As a measure of working memory, we use the well-known N-back test, which taps into this ability by asking participants to decide whether a stimulus was previously presented in certain conditions, taxing working memory to hold some information “in mind” (i.e., the sequence of stimuli) while simultaneously performing a different task (i.e., do the stimuli match?). We also included two levels of difficulty on the working memory test to see if this would further differentiate between the sexes.

We chose to concentrate on the accuracy and reaction times of working memory for the following reasons. First, implicit in the notion of working memory is that information is temporarily stored, thus there is some time-dependent window to provide an appropriate response and quicker reaction times are indicative of a more efficient working memory system. More generally, reaction times are an indicator of the neural maturity of children’s information processing system, thus depending on the task complexity, adults have on average reaction times between 500 and 1,200 ms but children can show reaction times between 1,000 and 10,000 ms (Lange-Küttner, 2012).
Second, regardless of how quick children are to react, their responses need to be accurate. Accuracy itself is not a simple unitary concept but one which can be understood by its component parts: providing a response when one is required (a true positive) and a response when one is not required (a false positive). Thus, the two different subcomponents of accuracy can capture something different about how sensitive working memory is (true positives) and how specific it is (false positives) to a given stimulus. In theory, participants could be perfectly sensitive to a stimulus by constantly responding; however, this strategy would pay a penalty for not being specific enough. Likewise, they could avoid false positives entirely by never making a decision, but this strategy would pay a penalty for not being sensitive enough. The two components of accuracy require different cognitive resources and strategies to succeed, and potentially unfold according to different schedules in development. For example, age-related improvement has been reported in the performance of a variety of tasks including working memory, susceptibility to interference, and inhibition of inappropriate responses (e.g., Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002). These behavioral changes to the functioning of working memory occur at the same time as reorganization and maturation of the frontal lobes and the prefrontal cortex, a location that has been strongly associated with working memory (see Carlson et al., 1998 for the N-back specific associations). In their review, Grissom and Reyes (2019) found little support for significant sex differences in overall executive function, although on impulsive action they note several sources of evidence in support of the idea that boys are more impulsive than girls and show associated quicker reaction times too. If this sex difference in impulsiveness is present in our participants, we would predict this would lead to relatively more false positives being committed by boys than girls, as well as boys having quicker reaction times.

**Hypothesis and summary**

We know from the adult literature that there are sex differences between men and women on measures of verbal abilities and spatial abilities and evidence that differences are already apparent in childhood (e.g., Ardila et al., 2011; Lange-Küttner & Ebersbach, 2013; Pelegrina et al., 2015; Vuontela et al., 2003). We explore this difference in the context of working memory performance with non-linguistic and non-spatial stimuli. Our hypothesis is that if sex differences are observed in children engaged with such stimuli, then they are for reasons other than direct linguistic or spatial. A plausible explanation for this difference is that the false positives rate and reaction time performance are likely to have a strong inhibitory control component. This cognitive resource is not equally distributed between the sexes at this age (Bunge et al., 2002; Luna, Garver, Urban, Lazar, & Sweeney, 2004) and the way in which these differences emerge in development are likely to be driven by qualitatively different maturation schedules of the relevant brain areas for boys and girls as well as quantitatively different neurological regions that are recruited during working memory tasks.

**Methods**

**Participants**

One hundred and twelve participants were initially considered for the study and were recruited from two schools located in different municipalities of Havana, Cuba. All children
aged 6 and 7 enrolled at the time were considered eligible to be part of the study. Of the initial 112, six did not complete the full schedule of N-back tests and two dropped out of school during the experiment. The remaining 104 children consisted of 54 girls (Mean Age = 7.2, SD = 0.5 Range 6.3–7.9) and 50 boys (Mean Age = 7.1, SD = 0.5 Range 6.3–7.9). We were interested in this age group as working memory is still undergoing significant change through childhood and adolescence (Gathercole, Pickering, Ambridge, & Wearing, 2004), including the effectiveness with which information in WM is updated (Belachi, Carretti, & Cornoldi, 2010).

Participants were recruited from two private English language schools: Britannia 1 (N = 42, M = 7.2, Female = 25, Male = 17) and Britannia 2 (N = 62, M = 7.2, Female = 29, Male = 33) located in two different municipalities of Havana. At these language schools, children attended two 60-min English classes per week, in addition to their weekly or twice weekly English classes at their regular school. Both private language schools are part of the same academic network, under the same management, so they follow similar pedagogical approaches, shared principles, and teachers and students often collaborate on classes and projects. Tuition fees are standard for both schools at 250 Cuban pesos per month, which means access is limited to those parents who can afford this. While the evidence of the effect of Social Economic Status on WM is equivocal, we present this information to add some background context to the location of the study.

**Ethics**

The study was approved by the Human Research Ethics Committee at The Open University (HREC Reference 3368). Interested parents and children were briefed on site in Cuba with via oral presentation, with opportunities to ask questions of the lead researcher. Together with a written Participant Information Sheet (in Cuban Spanish), this informed potential participants of the aims of the project, who were conducting the experiment and their affiliations, why they were being invited to take part, what the study involves if they decide to take part, what we would be observing and the main research methods. They were also informed how their data will be used, stored and that no individual will be identifiable from the published results. All communications with parents and children stressed their right to withdraw from the study at any point as well as the lead experimenter monitoring for signs on non-verbal assent or otherwise during the experiment.

**Working memory measures**

In the N-Back test, children were presented with a series of visual non-linguistic shapes and their task was to determine whether it matched a shape N shapes before. The shape sequence was generated and presented on a laptop using the Inquisit 5 software by Millisecond [https://www.millisecond.com/products/inquisit6/eduoverview.aspx]. Participants were encouraged to respond as quickly and as accurately as possible, and responses were automatically recorded using the same software. If the children thought they had detected a match, they pressed the letter “A” on the laptop. If they thought there was no match, they were not required to respond. A correct (true positive) response was recorded if participants pressed the “A” key when the target shape repeated itself N positions back. An incorrect response (false positive) was recorded if participants pressed
the “A” key when there was no target shape N positions back. Target correct matches represented approximately 28% of the presented stimuli, and each test block took approximately 25 minutes to complete. An N-Back sequence was created by pseudo randomly generating a sequence of shapes, taken from a pool of eight types of polygon (Figure 1) based on the stimuli used in Jaeggi et al. (2010).

Only one shape appeared on screen at a time for 2.7 seconds, followed by a 0.3 second gap (no shape) after which the next shape would be displayed and so on until the sequence was finished, Figure 2.

**Practice phase**

All children took part in one practice block for the 1-back and one for the 2-back so that they understood the rules of the game. The practice blocks contained 11 tokens of 8 types of shape for 1-back and 12 tokens of 8 types of shape for the 2-back. In the practice phase, participants received feedback on the accuracy of their answers.

**Test phase**

In the test phase, blocks contained 21 tokens of 8 types of shapes for the 1-back and 22 tokens of 8 types of shapes for the 2-back; furthermore, participants received no feedback as to the accuracy of their answers.

At the start of the test phase, all children took part in N-backs, which were equally weighted between 1-back and 2-back: three blocks of 1-back and three blocks of 2-back. For the remaining 16 N-back blocks, we introduced the possibility of altering the balance between 1-back and 2-back, depending on how the participants performed. If children scored 80% on three consecutive blocks of 1-back, then the balance of difficulty changed, such that they were then presented with two blocks of 1-back and four blocks of 2-back. We did this in case, as a result of repeated exposure to the task, some children’s performance

---

**Figure 1.** The pool of 8 irregular polygon types used for WM assessment.
would reach ceiling on the 1-back. In such cases, we could keep taxing their working memory at the upper limit with a regime more heavily weighted toward the more difficult 2-back. Of the 104 children involved in the study, only 5 children reached this 80% threshold (3 girls and 2 boys) and thus for the vast majority of participants, they stayed on the equally weighted three blocks of 1-back and three blocks of 2-back for the entire study.

Data were collected between 17/10/2019 and 28/11/2019. All true positives (pressing the A button on a target) and false positives (pressing the A button on a non-target) were extracted, and in total, 7888 responses were recorded between the 0–2999-ms response window.

Results

To get a global picture of children’s accuracy on the N-back test, we plotted their performance on a Receiver Operator Curve (ROC). The curve is created by plotting the true positive rate (TPR), also known as sensitivity, against the false-positive rate (FPR), which can be calculated as 1 – specificity. The greater the area under the curve (AUC), the better the overall accuracy, which in this context means participants were maximizing their TRP while minimizing FPR on the N-back. Thus, the ROC analysis is a natural way to combine the trade-offs involved in decision-making on a working memory task and reflects our interest in the idea that different components of accuracy require different cognitive resources and that these might unfold in development differently for boys and girls.

From participants’ raw number of responses on the N-back, we calculated their TPR and FPR as a ratio of all possible appropriate responses. For example, Participant 1 provided 49 true-positive responses from possible 64 opportunities to provide a positive response; 64 in this case would represent a perfect score. The TPR for this participant was thus (1/64)*49 = .76 and was compared against the true value of the stimulus, which in the case of true

![Figure 2](image-url). A summary of the possible participant responses across the stimuli presentations.
positives was 1 (the corresponding false-positive cases were compared to 0). This process was repeated for all participants, and the output of this analysis is displayed in Figure 3.

We can see that both boys and girls on both levels of task difficulty are performing significantly better than chance. To simplify the ROC interpretation, lines that are mostly to left of another line are closer to perfect performance of 1. As would be expected, the more difficult 2-back task depresses overall performance, reducing the AUC compared to the 1-back. From the AUC values it appears girls are performing marginally better than boys, particularly so in the 1-back test, and by 5.3% better across both N-backs. To unpack the notion of accuracy into its components, we next analyze sex differences with respect to TRP and FPR independently, to better understand this global picture.

**True-positive rate and false-positive rates**

To assess the effect of sex and task difficulty on TPR, we performed a 2 (1-back, 2-back) by 2 (boys, girls) ANOVA with repeated measures on task difficulty (all participants took part in both levels of the test). There was a significant main effect of task difficulty such that $F(1,102) = 1910.49, p < .001, \eta^2 = 0.94$, and no main effect of sex $p = .103$ and no significant interaction between task difficulty and sex $p = .368$. This means that there were significantly more true positive responses for 1-back than 2-back and a failure to find an effect of sex on true positives, Figure 4 A.
To assess the effect of sex and task difficulty on FPR we performed a 2 (1-back, 2-back) by 2 (boys, girls) ANOVA with repeated measures on task difficulty. There was a significant main effect of task difficulty such that $F(1,102) = 197.88, p < .001$, $\eta^2 = 0.68$ and a significant effect of sex (boys vs. girls) $F(1,102) = 4.52, p = .036$, $\eta^2 = 0.04$ and no significant interaction between task difficulty and sex $p = .750$. This means boys gave significantly more false-positive responses than girls regardless of task difficulty. There were also more false positives for 1-back than 2-back, regardless of sex, as was the case for TPR, Figure 4 B.

Overall, sex differences in working memory response types were limited to false positives.

**Reaction times**

Our other dependent measure of interest was reaction times. This reflects our interest in the time-dependent window to provide an appropriate response in working memory and, more generally, reaction times as an indicator of the neural maturity of children’s information processing system.

The reaction time distribution was significantly non-normally distributed ($\text{Kolmogorov–Smirnov} = .241, \text{df} = 7809, p < .001$). Therefore, before we performed statistical analyses that assumed normality, we first log-transformed the underlying distribution, keeping the mean and standard deviation of the sample the same, which produced a normal distribution ($\text{Kolmogorov–Smirnov} = .002, \text{df} = 7809, p = .200$). Using this transformed distribution, we assessed the effect of sex and task difficulty on reaction times following the same approach we took for accuracy, namely, performing a 2 (1-back, 2-back) by 2 (boys, girls) ANOVA with repeated measures on task difficulty, for TPR and FPR, respectively. Analyzing reaction time broken down by TPR and FPR reflects our interest in decomposing the global notion of accuracy into its component parts, and for underlying cognitive reasons that these might pattern differently between the sexes.
For TPR, there was no significant main effect of sex (boys vs. girls), \( p = .080 \), no significant effect of task (1-back vs. 2-back), \( p = .632 \) and no significant interaction between task difficulty and sex \( p = .592 \), Figure 5 A. For FPR, there was a significant main effect of sex (boys vs. girls) \( F(1,102) = 4.02 \), \( p = .047 \), \( \eta^2 = 0.039 \) and no significant effect of task (1-back vs. 2-back), \( p = .229 \) and no significant interaction between task difficulty and sex \( p = .688 \), Figure 5B. This means boys were significantly quicker to respond than girls for FPR but not for TPR, regardless of task difficulty.

**Speed-accuracy trade-offs**

Finally, participants completing the N-back are encouraged to respond as quickly and as accurately as possible. Facing such a speed-accuracy trade-off, participants could adopt a strategy of fast and careless or careful and slow, for example (Liesefeld, Fu, & Zimmer, 2015). To assess whether such a trade-off was evident in our study, and if so, whether sex was a differentiating factor, we used the Balanced Integration Score (BIS) advocated by Liesefeld and Janczyk (2019). This measure effectively controls for speed-accuracy trade-offs while retaining true effects. First, we standardized Reaction Times (RTs) and Percent Correct (PC; the proportion of true positives over all blocks) to bring them to the same scale and then subtracted one standardized score from the other:

\[
BIS_{ij} = z_{PC_{ij}} - z_{RT_{ij}}
\]

\[
with z_{ij} = \frac{x_{ij} - \bar{x}}{S_x}
\]

Thus, BIS is given by the difference in standardized mean correct RTs and PCs, with \( \bar{r}, PC, S_{RT} \) and \( S_{PC} \) calculated across all observed mean RTs and all PCs from our experimental data (including all participants and all conditions) that contribute relevant variance. A summary of the distribution of BIS by sex and task difficulty is given in Figure 6.
For 1-back, an independent samples t-test revealed a significant difference on BIS between boys and girls, \( t(102) = -2.218, p = .029, d = 1.45 \) (CI: \(-1.19, -0.066\)), such that girls scored a mean BIS of \(-0.303\) (SD: 1.40) and boys \(0.328\) (SD: 1.49).

For 2-back, an independent samples t-test revealed a significant difference on BIS between boys and girls, \( t(102) = -2.378, p = .019, d = 1.5 \) (CI: \(-1.07, -0.097\)), such that girls scored a mean BIS of \(-0.281\) (SD: 1.27) and boys \(0.304\) (SD: 1.23).

BIS reflects performance corrected for any variation in the speed-accuracy trade-off. Thus, if we found no difference on the BIS between boys and girls, this would indicate that the effect of sex on reaction times and accuracies that we demonstrated earlier was likely due to a speed-accuracy trade-off. However, this analysis demonstrates boys have a significantly higher BIS than the girls, indicating boys performed better than the girls on this measure.

**Discussion**

Using two levels of difficulty on a standard test of working memory, we found boys have quicker reaction times, but girls provide more accurate responses. With a comparable true positive rate between boys and girls, the sex differences in both accuracy and reaction times were isolated to false-positive responses, and this pattern of results was consistent across task difficulty. When performance was corrected for any variation in speed-accuracy trade-off, we found boys were at a significant advantage. To some extent, this is not surprising as the BIS balances RTs against proportion correct and our analysis has shown that with comparable true positive rates between boys and girls, sex differences were limited to false positives. However, the value-add of such analysis is that it shows that the quicker reaction times demonstrated by boys were not the result of a speed accuracy trade-off but a reflection of genuinely better performance on this task. In this regard, our results are in line with the general pattern of data seen in other developmental studies of working memory, where boys demonstrated faster reaction times and girls were more accurate (e.g., Lowe et al., 2003; Lynn & Irwing, 2008; Pelegrina et al., 2015; Vuontela et al., 2003). Where our study goes
beyond the established pattern, is demonstrating that these differences persist over two levels of task difficulty on a test of working memory with non-linguistic, non-spatial stimuli and pinpointing the difference in accuracy to a false-positive component. Our hypothesis was that if sex differences are observed in children who engaged with such stimuli, and then the differences are not directly due to linguistic or spatial factors. We controlled for such effects because we know from the adult literature that there are sex differences between men and women on measures of verbal abilities and spatial abilities (Kimura, 1999) and evidence that differences are already apparent in childhood (e.g., Ardila et al., 2011; Lange-Küttner & Ebersbach, 2013; Pelegrina et al., 2015; Vuontela et al., 2003). In what follows, we discuss some of the relevant cognitive mechanisms and underlying neurological substrates that might explain the pattern of behavior demonstrated in this study and its emergence in development.

As outlined in the introduction, the different components of accuracy (true positive/sensitivity vs. false positive/specificity) place different demands on cognition, and the resources needed to meet these demands might not be equally distributed between the sexes in development. Not producing inappropriate behavior (false positives) requires inhibiting a response, and as such we would expect a role for effortful control or inhibition. Inhibitory control is part of executive function; a collection of abilities that allows an individual to organize their behavior to respond to the current environment, to plan for future situations and is related to social competence during the preschool years (Caporaso, Boseovski, & Marcovitch, 2019; Ibbotson, 2023). The central executive essentially asks, “what should I do now” and a subtask of performing that function requires first answering the question “can I afford to wait.” In tests of impulsivity, such as the Go/NoGo task, girls of the age we studied here are typically less impulsive than boys and consequently, have slower reaction times too (Bezdjian, Baker, Lozano, & Raine, 2009). Outside of the laboratory, boys tend to be more impulsive in their risk-taking during play and are more likely to injure themselves (Rosen & Peterson, 1990). In more formal assessments of risk-taking (e.g., the Iowa Gambling Task), there are sex differences in tolerance for frequency of loss and magnitude of gains, which is thought to relate more generally to the greater sensitivity of women to both rewarding and punishing outcomes (e.g., van den Bos, Homberg, & de Visser, 2013). In taxing situations, boys are more likely to enact impulsive, aggressive responses instead of socially competent responses during peer conflict episodes (Caporaso & Marcovitch, 2021).

We did not test different age groups or follow the same children longitudinally within this study, so any developmental conclusions must be caveated by that fact. However, we did find important differences within the sample we used and that raises the question of why, and what is it that develops for these children that would explain such a difference? We offer the following possibility not based on neurological data that was generated within this study (see Limitations and future directions) but with reference to the wider established literature on sex differences in brain development. With respect to the current study, we showed working memory is developing more sensitivity for girls and is getting more specific for boys. Age-related improvement has been reported in the performance of a variety of tasks including working memory, susceptibility to interference, and inhibition of inappropriate responses (Bunge et al., 2002; Luna et al., 2004). These behavioral changes to the functioning of working memory occur at the same time as reorganization and maturation of the frontal lobes and the prefrontal cortex,
a location that has been strongly associated with working memory (Baddeley, 1986; Carlson et al., 1998). Wolf et al. (2018) showed that a test of visual selective attention activated canonical fronto-parietal attention systems and motion-sensitive areas in children as young as 7 years of age. Performance improved with age, together with stronger recruitment of parietal attention areas and a shift from low-level to higher-level visual areas. Wolf and colleagues argued improvements in selective visual attention and in the resolution of attention are characterized by an increased use of more functionally specialized brain regions during the course of development.

The general picture is that, although both sexes share a lot of the same networks when involved in working memory tasks, females consistently activate more limbic (e.g., amygdala and hippocampus) and prefrontal structures (e.g., right inferior frontal gyrus), and males activate a distributed network inclusive of more parietal regions (Hill, Laird, & Robinson, 2014). Not only are there quantitative difference in the areas of the brain that boys and girls recruit, there are qualitative differences too. Cerebral and gray matter volume in the frontal and parietal lobes peak earlier in girls than in boys (though the exact ages vary depending on the subregion), a pattern that may relate to sex differences in timing of puberty (Lenroot et al., 2007). The neural processes subserving working memory and brain structures underlying this system continue to develop during childhood with the prefrontal cortex one of the last brain regions to mature (Luna et al., 2004). In summary, one plausible explanation for the differences in WM we demonstrated is that the neural substrates that support impulsivity/inhibitory control are developing according to different developmental schedules for boys and girls. This causes boys to have quicker reaction times and to commit more false positives too. The way in which these differences emerge in development is likely to be driven by quantitatively different maturation schedules of the brain areas that underpin the relevant cognitive functions for boys and girls as well as qualitatively different neurological regions that are recruited during the task.

This raises the question whether one should conceive of the sex difference we establish here as one of working memory or impulsivity/inhibition. The answer to that question very much depends on the theoretical position one takes of what working memory is and how domain specific it is. For example, the theoretical framework of Cowan (1988) places greater emphasis on the possibility of domain interference within working memory than does Baddeley (1986). Although we only presented stimuli here in a non-verbal way, there is a more general controversy about the extent to which modes (verbal, nonverbal), and components of the executive function (attention, working memory) interfere with one another. That sex differences also depend greatly on the modality of testing and the parameters tested, has led some to suggest that apparent differences in the abilities in executive functions may in fact reflect different strategies employed by each sex when confronted with a challenge or ambiguous situation (e.g., Grissom & Reyes, 2019). For example, altering the types of information available in spatial working memory tasks reveals sex-specific behavioral strategies such that women appear to rely on allocentric cues to direct behavior (from the perspective of others), whereas removal of global spatial or directional cues has a greater detrimental effect in men (Chai & Jacobs, 2010). However, the distinction between strategy and ability may not always be so clear when the strategy is the ability one is interested in testing. That is, whether one defines global spatial cues or inhibitory control as a necessary part of a broader working memory function.
Limitations and future directions

We have argued that one plausible explanation for the differences in WM we demonstrated is that the neural substrates that support impulsivity/inhibitory control are developing according to different developmental schedules for boys and girls. However, we did not directly measure the neurological function of children with fMRI or EEG, for example. Stronger support for the causal explanation we support here, would be a demonstration that sex differences in brain activation (particularly, amygdala, hippocampus, right inferior frontal gyrus, and parietal regions) predict false-positive rates for the same children with the WM task we employed, something that we hope future studies will investigate.

There are of course other sex differences above and beyond boys’ impulsivity that could explain the differences we observed, for example, girls are better able to attend to the detail of the object rather than the gestalt whole (e.g., Lange-Küttner & Ebersbach, 2013), and this may make them better at recognizing which shape has gone before in the N-back. However, if that were true across the board, we would expect the advantage to be seen in TPR as well as the FPR. The difference we observed was restricted to FPR, which suggests a particular role for inhibitory control. The fact that females in general encode memories with greater detail, leading to better recognition abilities, could, however, help to explain the differences in reaction times. That is, girls pay a greater time-penalty in reaction times than boys because they encode more detail of the shapes. We used relatively complex geometric shapes, but a control condition with simpler shapes could assess how significant this factor is in explaining our pattern of results.

Conclusions

Drawing together the issues that this study has raised, we argue that differences in accuracy and reaction times are likely to have a strong impulsivity/inhibitory control component, and this cognitive resource is not equally distributed between the sexes at this age. This causes boys to have quicker reaction times and to commit more false positives too. The way in which these differences emerge in development is likely to be driven by quantitatively different maturation schedules of the brain areas that underpin the relevant cognitive functions for boys and girls as well as qualitatively different neurological regions that are recruited during the task.

It is important to recognize that the sex differences we found here were relatively small and, as such, were in line with the proportion of variance attributable to sex in other developmental studies (e.g., 1–3% in Ardila et al., 2011; Galsworthy et al., 2000). Furthermore, with respect to working memory performance, boys and girls are not deterministic categories but probabilistic ones; some boys are slower than girls and some girls are less accurate than boys. However, where there are differences, it can tell us something about developmental processes. No one explanation is likely to a give complete answer, with some sex differences showing degrees of contextual sensitivity while also showing similarities across different cultures (e.g., Mann, Sasanuma, Sakuma, & Masaki, 1990); some sex differences appearing to pattern the same across species and some differently (e.g., Herrmann, Hare, Call, Tomasello, & Emery, 2010); and in general, effect sizes appearing to be very dependent on the exact age of participants, the standardized test being used and the component of cognition under examination.
Where cognition is not indistinguishable between the sexes it is important to understand how and why they are different, partly because some of these differences are part of a bigger story of different fitness pressures played out over evolutionary timescale (Halpern, 1992; Kimura, 1999) and partly because, if we want to change the way we interact with boys and girls – be that clinical interventions, teaching or parenting – in order to attain more equitable outcomes, then knowing what biases and predispositions we have to work with is a good starting point.

Acknowledgments

Thank you very much to the children who took part in this study, the staff at the schools where the children were tested, and the time and expertise committed to improving this paper by the editor and two helpful reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References


