1 + 1 = 3: Factors influencing error detection with mathematical and reasoning problems

Carly Pymont (carly.pymont@anu.edu.au)
Department of Psychology, Bld 39, The Australian National University
Canberra, ACT 2617 Australia

Michael Smithson (michael.smithson@anu.edu.au)
Department of Psychology, Bld 39, The Australian National University
Canberra, ACT 2617 Australia

Abstract

Mathematics and reasoning are two skills which appear to be related based on observation and research examining cognitive capacities important in children’s mathematical ability. However, it is not clear if these two skills are linked. The present study investigated the relationship between different forms of mathematical problem error detection and deductive reasoning error detection with undergraduate students. Surprisingly performance on mathematical and reasoning error detection tasks were found to have little in common, despite the fact that participants viewed their abilities on the two tasks as related. Factors important for each type of task including mathematical experience and the task type were also examined. Results suggest that mathematical experience, while important for mathematical error detection has little effect on reasoning. Further implications for the manipulation of the context in the WST and reliance on reasoning were found.

Keywords: error detection; mathematics; reasoning; Wason Selection task; syllogisms.

Mathematical ability is consistently noted as important to many aspects of life including school and employment (e.g. Rivera-Batiz, 1992; Parsons & Bynner, 1997) yet there is still little known about the cognitive structure of mathematical abilities. In particular, it is not clear what other abilities should be linked to the ability to perform mathematics or what other abilities are particularly needed for mathematics. A candidate for an ability which may be related to mathematics is reasoning ability. Generally it appears that mathematics and reasoning should be linked and that they represent similar types of abilities, however there is a lack of research particularly investigating this question.

Examining the relationship between mathematics and reasoning, studies have attempted to determine the cognitive factors, including reasoning, which are related to children’s mathematical abilities. Taub, Floyd, Keith and McGrew’s (2008) examined applied problem and calculation abilities of children aged 5 to 19 years of age. They found that fluid reasoning, as measured by the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001), was related to mathematical ability. Further, Stylianides and Stylianides (2008) suggest that reasoning generally, but also more specifically deductive reasoning, is required for students to be able to use mathematical proofs. Reasoning and mathematical ability have also been linked such that the two terms are used almost interchangeably or as combined “mathematical reasoning” (e.g. Cirino, Morris & Morris, 2007; Varma & Schwartz, 2008). However, despite these findings and suggestions, Kroger, Nystron, Cohen, and Johnson-Laird (2008) provide some evidence that reasoning and mathematics are actually more distinct than originally thought. Kroger and colleagues (2008) presented participants with logic problems involving a series of statements which was either followed by a logical conclusion or a mathematical problem. Participants had to either indicate whether the logic followed from the statements or solve the mathematical problem. fMRI activation revealed that there were a large number of brain areas which were significantly more activated for the reasoning problem including right prefrontal cortex and inferior parietal lobe areas. Further, there were also areas which showed greater activation to the mathematical problem, including the frontoparietal network. However, joint activation of reasoning and mathematics problems could not be determined due to the lack of a control task as an indicator of general task requirements (Kroger et al., 2008). Despite mathematical tasks used by Kroger and colleagues as a control to separate activation due to processes such as working memory this suggests that processing of some mathematics and reasoning require different brain areas.

While mathematics and reasoning have been superficially linked, as outlined above there is not yet any clear evidence whether or not they are related skills. Indeed there are many forms of mathematics and many forms of reasoning, so the relationship may not be clear. What is true for probability theory for example may not be true for complex algebra. Hence the current examination will include a range of simple mathematical and reasoning tasks in order to begin to understand any links between mathematics and reasoning. The form of reasoning important for the current study is deductive reasoning, which has been measured in many ways including using the Wason Selection Task (WST; Wason, 1966). The WST has been used in deductive reasoning research since its invention in the 1960s (Wason,
This is particularly useful for mathematics and reasoning as both have clear right and wrong answers to problems such that there are correct answers to mathematical problems and statements which logically follow from reasoning problems. Error detection has also been used to examine linguistic processing, in particular processing of semantic and syntactic rules. In these studies syntactic and semantic errors result in increased word and sentence reading time (e.g. Wright & Garrett, 1984). This may indicate re-reading of the stimuli when an error is encountered, as eye movements show longer eye fixations and increased regressions to errors (Frazier & Rayner, 1982; Rayner, Sereno, Morris, Schmauder & Clifton, 1989).

The current study will examine both mathematical and reasoning error detection in order to examine whether any relationship exists between these two abilities. To overcome potential issues with reasoning tasks, two deductive reasoning tasks will be included: syllogisms and the WST. Further, participants will complete both the abstract and social versions of the WST, which will allow for an examination of whether improved performance on the social version will be generalized to the abstract version.

Four types of mathematical problems will be used to examine differential effects dependent on the mathematical requirements of the task. Calculations will be of a similar form to those used by Martin-Loeches, Casado, Gonzalo, DeHeras, and Fernandez-Drias (2006) in their examination of brain potentials to mathematical stimuli. While they do not report behavioural data comparing responses to correct and erroneous stimuli the difficulty appears appropriate for a university sample with an 11.5% error rate for long embedded calculations as will be used in the current study. Fractions will consist of a visual based comparison of a shaded segment of a circle with a fraction and again should be of a level of difficulty which is appropriate for the sample. Statements will be linguistically based rote recall problems where participants will need to detect errors in statements about numbers. Finally, power problems will involve participants verifying the outcome of a number or fraction raised to power. Power problems were selected to represent higher order or more complex mathematical problems which require the application of more than one mathematical rule (for example negative power and fractional power). Power problems, however, should be familiar to participants as such calculations are covered in mathematical high school classes.

Using these stimuli the current study will explore the error detection in different mathematical and reasoning tasks in order to examine the relationship between mathematical and reasoning error detection. Specifically two aims will be addressed in turn which are to:

1. Examine the relationship between mathematical and reasoning error detection to determine whether these two tasks may require a similar underlying ability. Further, it is of interest to examine whether participants feel that their performance on mathematical and reasoning tasks is related. While there is no evidence to link the two skills, due to the
assumptions made by researchers and individuals in the mathematical field it will be hypothesized that a common ability will underlie these two skills in this case.

2. Examine the factors which influence error detection including: stimuli type (including different types of mathematics and different types of reasoning), the participants’ confidence about doing the task and their mathematical experience. This will be conducted for each stimuli type.

Method

Participants
Seventy four first year students from the Australian National University (49 females; \( M_{\text{age}} = 18.93 \) years, \( SD = 1.27 \) years) participated in the current study for research credit. A further two participants were excluded due to not having complete enough mathematics training to ensure that they could complete the problems. All participants had normal or corrected to normal vision. Participants had a range of mathematical experience and the highest level of mathematical training was measured. The mathematical experience of participants was divided into the following three groups: completed a basic/standard level of mathematics in Yr. 12 (54.7%), completed an advanced level of mathematics in Yr. 12 (30.7%) or completed mathematical training after Yr. 12 (9%).

Apparatus
Stimuli were presented using Presentation (Neurobehavioural Systems inc.) on a Dell 21 inch CRT monitor with a refresh rate of 85 Hz. A chin rest was used to maintain viewing distance at 75cm

Stimuli
Stimuli were black figures, equations or words presented on a light grey screen. For fractions the section of the circle which was to be judged was a dark grey.

In order yo cover a range of different forms of mathematics four types of mathematical stimuli were examined; arithmetic, mathematical statements, fractions and power calculation, and Table 1 presents examples of correct and erroneous stimuli for each mathematical type. Calculation stimuli were generated from those used by Martin-Loeches et al. (2006) and were included as a simple and common for of mathematics. For fraction errors the label differed from the actual portion by at least 1/8 and all numbers in fractions, statements and power calculations were less than 30. Fractions were included in order to examine a more ‘visually’ based mathematics, while statements were designed to tap into communication aspects of mathematics. Finally power calculations were designed to represent a ‘higher order’ mathematical problem. Power problems were of a level similar to problems completed in high school mathematics classes and difficulty was introduced as to complete the problems more than one mathematical rule needed to be applied.

<table>
<thead>
<tr>
<th>Stimuli Type</th>
<th>Correct</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculation Statement</td>
<td>6 + (3 x 5) = 21</td>
<td>6 + (3 x 5) = 16</td>
</tr>
<tr>
<td>Fraction</td>
<td>Eight is an even number</td>
<td>Fourteen is an odd number</td>
</tr>
<tr>
<td>Power</td>
<td>( 8^{\sqrt{3}} = \frac{1}{4} )</td>
<td>( 6^{\sqrt{3}} = 36 )</td>
</tr>
</tbody>
</table>

Reasoning stimuli were either syllogism problems or the WST (Johnson-Laird & Wason, 1970). Syllogism problems presented syllogisms with two statements about a person, group of people or object, followed by a conclusion. Participants detected errors in the logic of the conclusion given the statements. In the WST all information was presented simultaneously and included a statement of how the cards were constructed and the rule (such as ‘if P then Q’) which the cards should obey was presented above fixation. At fixation height the four cards (P, not-P, Q and Not-Q) were presented next to each other in a random order. Below the cards a sample answer containing two cards was presented and was the source of potential error. The correct answer was the P and not-Q cards while the error was the most common error found in previous research on this task; the P and Q cards (Johnson-Laird & Wason, 1970). Social and abstract versions of the task were also presented following from Cosmides’ (1989) original manipulation. Social scenarios were all selected to be familiar to the participants.

Design
For the mathematical stimuli error (correct or error) and domain (calculation, statement, fraction or power) were manipulated within subjects. Problem type (mathematical or reasoning) was also manipulated within subjects as all participants completed one type of reasoning task and all the mathematical tasks. Type of reasoning (syllogism or WST) was manipulated between subjects. For the WST, context (social or abstract) and error (correct or error) were manipulated within subjects. Order of the WST context (abstract first or social first) was manipulated between subjects. For all analyses experience refers to mathematical experience which consists of the three levels outlined in the participants section.
Procedure

All 64 mathematical stimuli were randomly presented in a single block. For reasoning tasks participants either completed a single block containing all 32 syllogism problems randomly presented or two blocks: one containing all 16 abstract WST problems and containing all 16 social WST problems. Participants only completed one type of reasoning task (syllogism or WST) due to time constraints but the total number of reasoning tasks was identical.

Blocks were randomly presented and a break was given between blocks, in which the types of errors in the upcoming block were explained to the participant. Before the mathematics block participants were given definitions of the terms used in the statement problems to ensure that mistakes were not made simply due to the participants not remembering/knowing definitions.

Each trial began with a blank screen presented for 500ms followed by a fixation for 500ms and then the stimuli. Trials ended when the participant indicated via a keyboard button press whether or not they detected an error. Response time and accuracy were measured for each stimulus.

At the end of all blocks participants were asked to rate their confidence, enjoyment and ability on each of the tasks using an 11-point Likert scale. Participants were also asked to report the highest level of mathematical training they had completed or were completing.

Results

Aim 1 – Relationship between mathematics and reasoning

Bivariate correlations were computed between mean accuracy on all mathematical problems and the two reasoning problems. Mathematical performance was not found to significantly correlate with reasoning performance, $r = -.20, p > .05$ for WST and $r = .32, p > .05$ for syllogisms. Further, as it may be argued that certain types of mathematical tasks may be more related to deductive reasoning than others, bivariate correlations between accuracy on each mathematical task and reasoning were calculated and are presented in Table 2. While some mathematical problem types correlated with each other there were no significant correlations between any mathematical problem and reasoning performance.

In order to examine the structure of participants’ confidence on the different tasks, their confidence, liking, and ability ratings were subjected to a principal components analysis. The analysis revealed two components with eigenvalues exceeding one, explaining 72.49% and 17.86% of the variance respectively. After inspection of the scree plot one factor was extracted. This resulted in a single factor with all loadings of mathematical and reasoning ratings being greater than .8.

Examining the relationship between this confidence component and accuracy on the tasks revealed that confidence was a good predictor of performance on the mathematical and syllogism tasks, $r = .47, p < .001$ and $r = .481, p < .01$ respectively, but not on the WST task, $r = -.07, p > .1$.

Table 2: Bivariate correlations between accuracy on mathematical and reasoning tasks.

<table>
<thead>
<tr>
<th></th>
<th>State</th>
<th>Fract.</th>
<th>Power</th>
<th>Abs. WST</th>
<th>Soc. WST</th>
<th>Syll.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calc.</td>
<td>.37**</td>
<td>.18</td>
<td>.33**</td>
<td>.22</td>
<td>.13</td>
<td>-.17</td>
</tr>
<tr>
<td>State</td>
<td>.35**</td>
<td>.17</td>
<td>.04</td>
<td>-.05</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Fract.</td>
<td>.19</td>
<td></td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td>-.09</td>
<td>.13</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.77***</td>
<td>-</td>
</tr>
<tr>
<td>WST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p<.01, *** p<.001

A univariate ANOVA was conducted which found that experience significantly predicted scores on the confidence factor, $F(2,58) = 5.46, p < .01$, $\eta^2 = .16$. Mean confidence for each experience level is presented in Table 3. Comparisons revealed that experience beyond Yr. 12 increased confidence compared to advanced Yr. 12, $t(58) = -2.46, p < .05, d = -0.65$. Levels within Yr. 12 also made no difference, $t(58) = -1.01, p > .1, d = -0.27$.

Table 3: Mean confidence for each experience level.

<table>
<thead>
<tr>
<th>Experience</th>
<th>Basic Yr. 12</th>
<th>Adv. Yr. 12</th>
<th>Beyond Yr. 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>-.21</td>
<td>.06</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Aim 2a - Mathematics

To examine which factors influenced performance on mathematical tasks a 4 (domain: calculation, statement, fraction or power) x 2 (error: present or absent) x 3 (experience: Basic Yr. 12, Advanced Yr. 12, or Beyond Yr. 12) mixed ANOVA was conducted with domain and error manipulated within subjects. The analysis was run separately with mean accuracy as the DV and mean RT as the DV. Confidence was also added as a covariate.

Accuracy A significant interaction was found between the domain and error, $F(3,55) = 12.70, p < .001$, $\eta^2 = .41$, and Table 4 presents the means for this interaction. Bonferroni corrected comparisons revealed that errors were more accurately detected for calculation and power problems, $t(57) = -2.00, p < .05, d = -0.53$ and $t(57) = -5.36, p < .001, d = -1.42$ respectively, while correct problems were easier for fractions, $t(57) = 2.49, p < .01, d = 0.66$. In comparison there was no accuracy difference between correct and errors for statements, $t(57) = -0.48, p > .1, d = -0.01$. 

<table>
<thead>
<tr>
<th>Domain</th>
<th>Basic</th>
<th>Advanced</th>
<th>Beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fract.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Mean accuracy for each domain.
In general errors (\(M = .84, SE = .02\)) were also more accurate than correct stimuli (\(M = .76, SE = .02\)), \(F(1, 57) = 8.21, p = .006, \eta^2_p = .13\), and mathematical domain had a significant effect on performance, \(F(3, 57) = 69.06, p < .001, \eta^2_p = .79\). Examining the effect of domain contrasts, revealed that performance was significantly better for calculation (\(M = .92, SE = .02\)), \(t(57) = 10.49, p < .001, d = 2.78\). and for statements (\(M = .88, SE = .02\)), \(t(57) = 4.32, p < .001, d = 1.14\). Performance was significantly worse on power (\(M = .59, SE = .02\)), \(t(57) = -10.85, p < .001, d = 2.26\), while statements and fractions (\(M = .81, SE = .02\)) did not significantly differ, \(t(57) = 0.58, p > .1\).

While experience was approaching significance \(F(2, 57) = 2.88, p = .064, \eta^2_p = .09\), confidence had a significant positive effect on performance, \(F(1, 57) = 10.36, p = .002, \eta^2_p = .15\). All other effects were non-significant, \(p > .05\).

### RT

As confidence significantly interacted with the mathematical domain, \(F(6, 56) = 4.61, p < .01, \eta^2_p = .20\), and it did not predict performance \(F(1, 57) = 2.19, p > .1\), \(\eta^2_p = .04\), it was removed from the analysis. However, an examination of the correlations between confidence and the mathematical domains revealed a negative relationship with RT on calculation tasks, \(r = -32, p < .05\) and a positive relationship with RT on power tasks, \(r = .33, p < .01\) while the other correlations were non-significant.

In the overall ANOVA a significant interaction was found between experience and domain, \(F(3, 63) = 2.16, p = .01, \eta^2_p = .11\). Means for this analysis are presented in Table 5. Comparisons revealed that for participants with only basic Yr. 12 experience, statements were faster than calculations, \(t(38) = -5.17, p < .001\), which were not faster than fractions, \(t(38) = -1.79, p = .08\), and that fractions faster than power, \(t(38) = -4.51, p < .001\). For individuals with advanced Yr. 12 statements and calculations did not differ, \(t(18) = -0.38, p > .1\), while calculations were slower than fractions, \(t(18) = -2.63, p = .02\) and fractions were slower than power, \(t(18) = -3.09, p < .01\). For participants with experience beyond Yr. 12 only the difference between fractions and power was approaching significance, \(t(3) = -2.55, p = .05\), while statements and calculation, and calculation and fractions did not differ, \(t(3) = -1.31, p > .1\) and \(t(3) = -2.05, p = .09\).

<table>
<thead>
<tr>
<th></th>
<th>Calc.</th>
<th>State</th>
<th>Fract.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>.87 (.02)</td>
<td>.88 (.03)</td>
<td>.87 (.03)</td>
<td>.41 (.05)</td>
</tr>
<tr>
<td>Error</td>
<td>.94 (.02)</td>
<td>.89 (.03)</td>
<td>.75 (.03)</td>
<td>.78 (.04)</td>
</tr>
</tbody>
</table>

### Table 4: Mean accuracy (and standard error) for correct and erroneous stimuli by mathematical type.

Overall domain had a significant effect \(F(3, 63) = 20.54, p < .001, \eta^2_p = .49\), with calculations (\(M = 40.97, SE = 1.9\)) and statements (\(M = 35.35, SE = 2.16\)) responded to more quickly, \(t(63) = -6.09, p < .001\) and \(t(63) = -7.80, p < .001\), respectively, and power (\(M = 100.46, SE = 12.0\)) more slowly, \(t(63) = 7.32, p < .001\). In comparison fractions (\(M = 54.87, SE = 4.67\)) did not differ \(t(63) = -1.29, p > .1\), \(d = -0.32\). Further, there was no significant difference in reaction time for correct and erroneous stimuli, \(F(1,65) < 1, \eta^2_p = .01\). All other effects were non-significant, \(p > .05\).

### Aim 2b - Syllogism

To examine factors affecting performance on the syllogism task a 2 (error: present or absent) x 3 (experience: Basic Yr. 12, Advanced Yr. 12, and Beyond Yr. 12) mixed ANOVA was conducted with error manipulated within subjects. This was run for accuracy and RT and confidence was added as a covariate.

#### Accuracy

Correct stimuli were more accurately identified (\(M = .75, SE = .03\)) than errors (\(M = .50, SE = .04\)), \(F(1, 27) = 20.54, p < .001, \eta^2_p = .43\). Confidence also had a significant positive effect, \(F(1, 27) = 10.78, p < .01, \eta^2_p = .29\). Experience was not related to performance, \(F(2, 27) < 1\) and no interaction effects were significant, \(p > .1\).

#### RT

As confidence did not have a significant effect, \(F(1, 27) = 1.89, p = .18\), it was removed from the model. Correct stimuli were responded to more quickly (\(M = 1.68, SE = .04\)) than errors (\(M = 1.78, SE = .05\)), \(F(1, 29) = 6.25, p = .018, \eta^2_p = .18\). Experience or its interaction with error presence did not predict performance, \(F(2, 29) < 1\) for both.

### Aim 2c – WST

To examine which factors affected performance on the WST a 2 (error: present or absent) x 2 (context: abstract or social) x 3 (experience: Basic Yr. 12, Advanced Yr. 12, or Beyond Yr. 12) x 2 (order: social first or abstract first) mixed ANOVA was conducted with error and context manipulated.
within subjects. As above, this was run for accuracy and RT and confidence was added as a covariate.

Accuracy As confidence did not significant predict performance, $F(1, 20) = 17.44, p > .1$, it was removed from the model. The order that the problems were presented in did not significantly affect the performance, $F(1,21) = 1.44, p > .1$. Error and context significantly interacted, $F(1,21) = 17.44, p < .001, \eta^2_p = .45$, such that when the correct answer was displayed there was no difference between abstract ($M = .32, SE = .04$) and social ($M = .31, SE = .04$) versions of the task, $t(20) = -0.30, p > .1$ but when the incorrect answer was displayed the social ($M = .73, SE = .04$) version was easier than the abstract version ($M = .49, SE = .03$), $t(20) = -4.40, p < .001$.

Generally abstract ($M = .41, SE = .02$) was harder than social ($M = .52, SE = .03$), $F(1,21) = 6.30, p = .02, \eta^2_p = .23$, and errors ($M = .61, SE = .03$) were easier than correct problems ($M = .32, SE = .03$), $F(1,21) = 25.67, p < .001, \eta^2_p = .55$. Finally, experience interacted significantly with error, $F(1, 21) = 6.02, p < .01, \eta^2_p = .36$. Means for this interaction are presented in Table 6. Contrasts indicated that for individuals with beyond Yr. 12 errors and correct stimuli were of the same difficulty, $p > .1$, while errors were easier for those with basic Yr. 12 and advanced Yr. 12, $t(14) = -4.44, p = .001$ and $t(7) = -13.36, p < .001$ respectively.

Table 6: Mean accuracy and standard errors for the effect of experience on correct and erroneous WSTs

<table>
<thead>
<tr>
<th>Experience</th>
<th>Correct</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>neat Yr. 12</td>
<td>.37 (.04)</td>
<td>.58 (.03)</td>
</tr>
<tr>
<td>Adv. Yr. 12</td>
<td>.23 (.05)</td>
<td>.74 (.04)</td>
</tr>
<tr>
<td>Beyond Yr.</td>
<td>.38 (.10)</td>
<td>.41 (.09)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RT As confidence did not significantly predict RTs, $F(1, 20) < 1$ it was removed from the model. The only effect which was significant was the interaction between the order that the contexts were presented and the contexts, $F(1, 21) = 7.25, p < .05, \eta^2_p = .26$. The means for this interaction are presented in Table 7. Comparisons revealed that the abstract version of the task was slower when presented first, $t(12) = 3.27, p < .01$ and that the social version of the task was slower when it was presented first, $t(12) = -2.36, p < .05$. The main effects of error, experience and context were non-significant, $F(1,21) = 1.09, p > .05, \eta^2_p = .05, F(2,21) = 1.2, p > .1$ respectively, and all other interactions were non-significant.

Table 7: Mean RT in seconds (and standard error) for the interaction between context order and context

<table>
<thead>
<tr>
<th>Order</th>
<th>Abstract first</th>
<th>Social first</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>16.6 (1.62)</td>
<td>13.7 (1.70)</td>
</tr>
<tr>
<td>Social</td>
<td>12.6 (2.16)</td>
<td>19.0 (2.26)</td>
</tr>
</tbody>
</table>

Discussion

The current study aimed to examine the link between mathematical and reasoning skills through error detection tasks in these domains. Further the dynamics of performance on these tasks was examined. In comparing performance on reasoning and mathematical tasks, surprisingly little relationship between the different tasks was found, such that no mathematical problem types were correlated with performance on either reasoning task. Hence, claims of an overlap between reasoning and mathematical ability may not be accurate and reasoning may not be needed for mathematics. However, participants did view their performance on these two types of task as related, with a single factor being formed. This could relate to our views about such tasks and what the tasks should involve, but these views may not actually be related to the skills required. The lack of a relationship, however, may also be a reflection of the limited range of mathematical abilities of the participants and range of mathematical tasks used. Therefore reasoning may be related to expert level mathematical abilities, or particular forms of mathematics. It should be mentioned that while a range of mathematical tasks and two different reasoning tasks were included the findings maybe specific to those domains. It may be that reasoning is particularly related to certain forms of mathematics, such as proof based mathematics as suggested by Stylianides and Stylianides (2008). Further the findings may be specific to the particular task which participants performed. In asking participants to complete a forced choice error detection task, the demand on reasoning skills may have been decreased. These considerations allow room for future research predominantly in examining the cognitive processes underlying such tasks and the generalisability of the findings.

The other aim was to examine factors influencing error detection across the three different tasks. For the mathematical task, predictors of performance were the mathematical domain, error presence, experience and confidence. Interestingly, performance was increased for errors for the easiest and hardest mathematical problems, so this finding cannot be explained in terms of difficulty. Instead, the form of the problem appears to be important in that both of these tasks were equation based problems which combine mathematical elements which then require solving. Fraction problems, on the other hand, were visual based problems where the size of the segment needed to be determined and compared to the printed fraction and in this case errors were more difficult. Also, statements were based

---

Article DOI: 10.5096/ASCS200943
on linguistic recall of information and so did not require the same “working out” as calculation and power problems. This may suggest that errors were only more accurate when the problem required explicit solving rather than just recognition.

Both confidence and experience had a similar effect in predicting increased mathematical performance. However, confidence was more important as it predicted both how fast tasks were completed as well as accuracy, while experience only predicted reaction time. This suggests that while experience does to some degree increase confidence and predicts how quickly mathematical tasks can be completed, the individual needs to be confident in order to actually perform better. The effect of experience may also have been a consequence of the amount of time passed since participants were required to complete mathematical problems. Individuals who completed mathematics training after high school and were possibly completing some at the time of testing had a more consistent reaction time across domains while individuals who stopped mathematics after high school had a more varied reaction time profile. Hence the domains were more varied for those who had more time away from doing such problems and probably were less automated in their solutions.

In terms of reasoning, confidence was related to syllogism accuracy but did not relate to performance on the WST. This suggests that participants were not good at predicting how they would do on the WST. Also as experience was not related to performance on either reasoning tasks, the skills required for these tasks may not be taught or built on in mathematical education.

Syllogism error detection was affected by the task type, such that syllogism errors took longer than correct stimuli following a similar pattern as observed to sentence processing errors (e.g. Wright & Garrett, 1984). This may be indicative of the linguistic component of the task and the potential need to re-read when an error is encountered.

For the WST Cosmides’ (1989) findings were replicated in that performance was improved on the social version of the task over the abstract version. However, this difference disappeared when the correct answer was displayed. This suggests that the social version of the task does not facilitate performance when recognition of the correct answer is needed, but does when the correct answer needs to be generated or at least the given answer disproved. Hence there may be separate processes involved for the different contexts. Indeed Cosmides’ results indicate that the processes involved in the two versions may be different but particularly recognition competence may not help in the abstract version of the task. Overall, however, most of the findings with the WST can be accounted for by the cards cueing particular responses rather than actually involving reasoning. This finding follows from the arguments raised in the introduction about what the WST actually measures and adds to the argument that the processes underlying the WST may in fact not be reasoning. While the current study, due to its design, could not examine the relationship between the WST and syllogisms, it would be interesting to examine the relationship between the different forms of the WST and others forms of reasoning tasks. This may give some indication of what skills are required to complete the task and in particular, whether one version of the task is more reliant on a certain type of reasoning.

Interestingly, the order in which the contexts were presented was not found to impact performance, such that facilitating performance on the social version did not generalise to the abstract version. The only effect observed due to the order of the tasks was that the context which was completed first was slower than the context completed second, which is not unexpected as it suggests that participants recognised them as similar tasks. So there is a lack of transference of skills learnt in one context to the other and correctly solving the social version does not help understand the underlying task. Alternatively it may be that different skills are required for solving the two different versions of the task. However participants did recognised the task as similar to the one they completed earlier, or at least some practise effects were evident.

As mentioned the range of tasks and mathematical competence in this study which may have resulted in some of the effects being weakened. Hence, applying this experimental set up to other domains as well as experts or high functioning individuals in mathematics may strengthen these results further, or provide some counter evidence.

The main finding, however, was the lack of support for the relationship between mathematics and reasoning. Despite the participants believing that their performance on the tasks was linked there was no evidence of a relationship between performance on any of the mathematical tasks and the reasoning tasks. Further mathematical experience did not influence reasoning task performance suggesting that reasoning skills are not learnt as a part of mathematical education. While the WST may not be a true reasoning task, as evidenced by previous findings as well as the current ones, the effect was also not found for syllogisms. It is yet to be suggested that syllogistic problems do not involve reasoning and an examination of most basic cognition books will show that syllogisms are used as the main example of deductive reasoning. Hence the results should not be simply discounted and suggest a need to examine the relationship between mathematics and reasoning rather than assuming it should exist. In particular research with different forms of mathematics as well as reasoning tasks (beyond the deductive reasoning used here) would be of interest. However in order to fully understand whether mathematics and reasoning are linked, research should examine the cognitive processes underlying the tasks. Then a real understanding of the tasks individually, as well as any link between them can be gained.
References


Citation details for this article:
