Two Routes to Expertise in Mental Rotation

Alexander Provost, Blake Johnson, Frini Karayanidis, Scott D. Brown, Andrew Heathcote

School of Psychology, University of Newcastle
ARC Centre of Excellence in Cognition and Its Disorders, Macquarie Centre for Cognitive Science, Macquarie University

Received 31 October 2011; received in revised form 24 July 2012; accepted 27 September 2012

Abstract

The ability to imagine objects undergoing rotation (mental rotation) improves markedly with practice, but an explanation of this plasticity remains controversial. Some researchers propose that practice speeds up the rate of a general-purpose rotation algorithm. Others maintain that performance improvements arise through the adoption of a new cognitive strategy—repeated exposure leads to rapid retrieval from memory of the required response to familiar mental rotation stimuli. In two experiments we provide support for an integrated explanation of practice effects in mental rotation by combining behavioral and EEG measures in a way that provides more rigorous inference than is available from either measure alone. Before practice, participants displayed two well-established signatures of mental rotation: Both response time and EEG negativity increased linearly with rotation angle. After extensive practice with a small set of stimuli, both signatures of mental rotation had all but disappeared. In contrast, after the same amount of practice with a much larger set both signatures remained, even though performance improved markedly. Taken together, these results constitute a reversed association, which cannot arise from variation in a single cause, and so they provide compelling evidence for the existence of two routes to expertise in mental rotation. We also found novel evidence that practice with the large but not the small stimulus set increased the magnitude of an early visual evoked potential, suggesting increased rotation speed is enabled by improved efficiency in extracting three-dimensional information from two-dimensional stimuli.

Keywords: Electroencephalography; Mental rotation; Cognitive strategy; Training

1. Two routes to expertise in mental rotation

In their seminal experiments on spatial cognition, Shepard and Metzler (1971) studied reaction time (RT) for judgments of a match between matching and mirror image pairs of
perspective drawings, a shape-matching task. They found that mean RT increased linearly with angular displacement between objects for both picture plane and depth rotations. Shepard and Metzler proposed that participants use a mental rotation process analogous to the physical act of rotating one object in an attempt to match to another object. These results had strong implications for our understanding of the nature of mental imagery: They indicated that mental representations have some degree of isomorphism with real-world objects; and that mental transformations of these representations involve analogue, rather than propositional processes.

Mental rotation—as indexed by Shepard and Metzler’s elegant experimental paradigm—is now considered to be a prototypical process of spatial thought (Corballis, 1997). Spatial thought is fundamentally different from linguistically based thought and underpins human activities ranging from everyday navigation, reaching, and grasping, to expert performance in the sports, arts, and sciences. As mental rotation is considered to be a critical component of spatial processing, there is considerable interest in the extent to which mental rotation skills can be improved and the techniques by which expertise can be best accomplished. It is known that mental rotation does improve markedly with practice, as evidenced by performance in shape-matching tasks (e.g., Kail & Park, 1990). Although intensive study has revealed much about the cognitive and physiological underpinnings of mental rotation, an explanation of its plasticity remains controversial (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008).

We used converging behavioral and event-related potential (ERP) evidence to demonstrate two routes to expertise in a shape-matching task. The high temporal resolution of ERPs measured by electroencephalography (EEG) provided a fine-grained characterization of how the time course of processing in each route changed with practice. We indexed mental rotation during a shape-matching task using an ERP modulation called “rotation related negativity” (RRN). The RRN is defined by greater EEG negativity for greater rotation angles, and it typically emerges at a latency of 300–600 ms at parietal electrodes (Heil, 2002; Milivojevic, Johnson, Hamm, & Corballis, 2003; Peronnet & Farah, 1989; Thayer & Johnson, 2006). Character stimuli have been used to demonstrate a functional link between mental rotation and the RRN in a shape-matching task; in a contrast between ERPs for character classification and match judgments about rotated characters, only the matching task caused an RRN (Heil, Bajric, Rosler, & Hennighausen, 1996).

Neuroimaging and neurostimulation studies support conclusions from ERP studies about the location and time course of mental rotation processing. Consistent with the parietal locus of the RRN, and evidence that spatial transformations are processed in the parietal cortex, fMRI studies investigating mental rotation show activation in the intraparietal sulcus and surrounding areas during mental rotation (Zacks, 2008). Regional blood flow in the right parietal lobe measured by PET is also correlated with mental rotation (Harris et al., 2000). Also right parietal cortex functional lesions induced by repetitive transcranial magnetic stimulation have revealed a critical time window for mental rotation 400–600 ms after stimulus presentation (Harris & Miniussi, 2003).

Practice effects in shape-matching task performance, which include both a decrease in overall RT and a decrease in the effect of rotation angle on RT, have been explained by
single- and dual-process accounts. The single-process account attributes the reduced angle effect to participants becoming faster at rotation (Bethell-Fox & Shepard, 1988). The dual-process account assumes that rotation speed remains constant, whereas practice engages a second process that is not sensitive to angle: direct retrieval from memory of the correct response associated with each particular stimulus (Tarr & Pinker, 1989).

The dual-process approach is a special case of Logan’s (1988) instance theory of automaticity. On each practice trial an instance, which contains a stimulus–response association, is stored in memory. A new stimulus triggers retrieval of stored instances, which race with each other and with an algorithmic process (e.g., mental rotation and shape matching) to produce a response. Instance retrieval time varies randomly, so as more instances accrue, the time for the first instance to be retrieved decreases. Initially, when only a few instances join the race, the algorithm usually wins, but as more instances join the race, memory retrieval controls the response more and more often until eventually responding is always based on retrieval.

The single-process account suggests that practice effects will transfer to previously unpracticed stimuli, whereas the item-specific learning associated with retrieval should result in little or no transfer. Evidence on this point is mixed, with some studies finding transfer (Leone, Taine, & Droulez, 1993) and others not (Heil, Rösler, Link, & Bajric, 1998; Kail & Park, 1990). Evidence is also mixed for transfer to a previously unpracticed task that shares a mental rotation component with the practiced task (e.g., De Lisi & Cammarano, 1996; Terlecki, Newcombe, & Little, 2008; Uttal et al., in press; Wright et al., 2008 for supportive evidence vs. Heil et al., 1998; Kail & Park, 1990; Sims & Mayer, 2002 for transfer failure).

We propose that these inconsistencies occur because practice can engage two different strategies that improve performance. Consistent with this proposal, Wright et al. (2008) provided evidence of both strategies operating in the same task; responses to practiced stimuli were faster than to unpracticed stimuli, consistent with the dual-process account, but responses to unpracticed stimuli were faster after practice than before practice, consistent with the single-process account. Neuroimaging evidence also points to individual differences in strategies used in the matching task (Logie, Pernet, Buonocore, & Sala, 2011; Riečanský & Jagla, 2008).

We test this proposal in two shape-matching experiments designed to bias participants toward either retrieval or rotation strategies. In both experiments, participants completed six 1-h sessions. EEG measures were taken in the initial and final sessions, which were separated by approximately a week during which they performed four practice sessions. The experiments used stimulus pairs similar to Shepard and Metzler’s (1971), with the angle between a standard and rotated objects manipulated over five equally spaced levels from 0 to 180 degrees.

The first experiment was designed to maximize improvement via the retrieval route. It used a small set of 10 stimulus pairs all with the same standard object. The small set size made it quick and easy to learn the correct stimulus–response mapping for every pair, encouraging improvement via retrieval. The retrieval account predicts that a reduced angle effect on mean RT should be accompanied by a reduced RRN. Pilot testing
indicated that the level of practice used was sufficient to severely attenuate the angle effect on mean RT, so we predicted that, after practice, there would be little or no modulation of the ERP by angle.

The second experiment was designed to maximize improvement via the mental rotation route. Participants practiced with a large set of 320 stimulus pairs generated from 32 different standard objects. Each pair was used only once in each session, minimizing item-specific practice effects, and hence improvement via the retrieval route. Pilot testing indicated that substantial improvements in performance with practice still occurred with the large set. On the assumption that improvement occurred via the rotation-speed route, we predicted that the RRN would remain after practice and may even increase or occur earlier.

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were post-graduate students or staff members (N = 12, 7 male) from the University of Newcastle with a mean age of 34 years (range 25–47). All were in good health, had normal or corrected to normal vision, and gave informed consent.

2.1.2. Procedure

The four practice sessions all occurred within one working week. The first EEG session occurred either on the Friday before or the Monday of the practice week. The final EEG session occurred either on the Friday of the practice week or on the following Monday. Trials were grouped into eight blocks in the EEG sessions and ten blocks in the practice sessions. Each block had 80 trials, consisting of eight repetitions of each stimulus pair randomly ordered. Use of a mental rotation strategy was encouraged by showing participants a video of a stimulus pair with the left-hand member (standard) stationary and the right-hand member rotating from 180° to 0°. Participants were shown this video of the same (i.e., matching pair) condition followed by a video of the different (i.e., mirror pair) condition.

Fig. 1a illustrates the time course of a trial. Stimuli for all participants were the same set of 10 two-dimensional projections of three-dimensional block–object pairs with a longest axis (spine) of four blocks and two right angle projections from the spine (arms) made of either one or two blocks. Each pair had the same standard object presented on the left with a vertically oriented spine and a rotated (0°, 45°, 90°, 135°, 180°) version of it, which was either identical to (same condition) or a mirror image of (different condition) the standard, presented on the right.

Participants made a match judgement by pressing a response button with their index finger on their right hand to indicate a same pair and their middle finger to indicate a different pair, with performance feedback. Participants had to make the correct response to
progress. Once an accurate response was made, the stimulus remained on screen for 2 s. Participants were supine in the EEG sessions. Stimuli were displayed using an LCD projector on a plastic screen located 37 cm above participants. Practice sessions were completed seated in front of a CRT display located 60 cm in front of participants with stimuli scaled to subtend the same visual angle as in the EEG sessions (10° from the left to right edges of the stimulus pair).

MEG as well as EEG recording occurred in the first and last sessions, with results for only the behavioral and EEG data analysis reported here. For EEG measurement, a 64-channel BrainAmp recording cap (Brainproducts, Munich, Germany) was positioned according to the 10/20 system. The frontal central midline (FCz) electrode was the reference channel, vertical electrooculogram (EOG) was measured with a single electrode placed below the eye, and participants’ electrocardiogram (ECG) was also measured. EEG data were sampled at 1,000 Hz and bandpass filtered from 0.1–1,000 Hz with impedances set below 20 kΩ. EEG data were analysed off-line using BESA version 5.3.4 (MEGIS Software GMbH, Grafelfing, Germany). Data were re-referenced to the average reference and epoched around stimulus presentations of correct trials between −200 and 1,000 ms. EOG channel correction used a regression algorithm (Ille, Berg, & Scherg, 2002) and baseline correction was relative to a ±50 ms interval around stimulus onset. Averaged data were filtered with a bandpass of 0.2–30 Hz. Equipment failures caused 0.04% of responses to be lost.

3. Results

The following analysis describes the set of stimuli used throughout the experiment as “novel” when tested in the first session and “practiced” when tested in the final session. As is commonly observed in many tasks, behavioral performance improved at a decreasing
rate over practice sessions (Heathcote, Brown, & Mewhort, 2000). Because our interest focuses on the relationship between EEG and behavior, we focus on results from the first and last sessions (plots of practice curves are provided as supporting information available online).

Responses faster than 200 ms (0.015%) and slower than 7,000 ms (0.02%) were excluded from further analysis as outliers. Approximately, 2.3% of trials were removed from further analysis due to artifacts resulting in EEG amplitudes over 200 μV. We report significant ($p < .05$) results from ANOVA analyses of the angle × match (same vs. different pairs) × practice (first vs. last session) design. In all inferential analyses of effects with more than one degree of freedom, Greenhouse–Geisser corrections $e$ and corresponding $p$ values are reported.

To minimize floor effects, error rates were calculated using Snodgrass and Corwin’s (1988) recommended procedure and inferential analyses applied to the $z$ transformation of these error rates. Angle effects were quantified by linear regression estimates of the time required for a $180^\circ$ rotation; associated correlation coefficients were close to 1 when angle effects were significant, indicating that the regression estimates captured most of the angle effects. Bars on Fig. 2 indicate a bias-corrected within-subject standard error (Morey, 2005).

3.1. Accuracy and mean RT

Table 1 reports significant ANOVA effects for accuracy and mean correct RT analyses, with Greenhouse–Geisser corrections where appropriate. Fig. 2 summarizes behavioral results for stimuli when they were novel (Session 1) and after they were practiced (Session 6). Accuracy was high both before and after practice, with error rates generally well below 5% except for angles greater than $90^\circ$, particularly in the first session. The main effect of angle on accuracy was significant, as was its interaction with match, due to a stronger angle effect for same than different pairs. Mean RT was significantly affected by all factors in the three-way ANOVA. Practice interacted with the other two factors due to a reduction in both the angle effect and the advantage for same over different pairs with practice. In contrast, the relationship between match and angle was additive. Before practice it took, on average, 500 ms to perform a $180^\circ$ rotation, whereas after it took only 90 ms. However, analysis showed that the angle effect remained significant after practice for same stimuli; see Table 1.

3.2. Event-related potentials

Following previous research, we examined the RRN at the central parietal electrode (Pz) (Heil, 2002). Fig. 3 shows grand average waveforms for same and different pairs in the first (Novel) and last (Practiced) sessions. A strong RRN was evident before practice, but it was greatly reduced after practice. To quantify the RRN, we calculated mean amplitudes for each condition before and after practice in 100 ms windows from 200 ms to 1 s. Fig. 4 displays these measures, averaged over participants, as black lines, and points in a Trellis plot (Cleveland, 1993).
A linear regression analysis was performed on the mean amplitudes within each window as a function of angle, with average results depicted as gray lines in each panel of Fig. 4. The slopes of these lines summarize the strength of the RRN in each panel of Fig. 4. In each time window we tested mean amplitudes with angle × match ANOVAs. The strip at the top of each panel in Fig. 4 reports the null-hypothesis probability (p) for the main effect of angle in these ANOVAs. Significant angle main effects only occurred

Fig. 2. Experiments 1 and 2 mean RT results for same (S) and different (D) pairs of stimuli tested in Session 1 (Novel) and for the same stimuli tested in Session 6 (Practiced), and also for stimuli in Experiment 2 that were only ever tested in Session 6 (Transfer). Note that large y-axis range is required, so the results of Experiments 1 and 2 can be presented on the same scale.
before practice, in the 500–600 and 600–700 ms windows (N 500–600 and N 600–700 in Fig. 4). Even with a strict Bonferroni correction to control Type 1 error given nine windows were tested (i.e., a critical $p = .05/9$), these two effects remained significant. This result was confirmed by significant angle × practice interactions in separate within-subjects ANOVAs for each time window ($p < .005$ and $p < .002$, respectively). This interaction was not significant for any other mean amplitude window.

### 4. Discussion

Practice caused a large decrease in overall RT and in the effect of angle on RT, whereas accuracy stayed relatively constant. Practice attenuated the angle effect on mean RT by a factor of more than 5, but a significant angle effect remained. One interpretation of the behavioral findings is that practice shifted participants to a predominant retrieval strategy that is unaffected by angle, but with some participants on at least some trials persisting with a mental rotation strategy to complete the shape-matching task, resulting in a small residual angle effect. A second interpretation is that participants learned to greatly increase rotation speed.

If participants largely adopted a retrieval strategy, mental rotation would rarely if ever be required to complete the shape-matching task following practice. Hence, if the RRN is a marker of mental rotation, it should be present before practice, but attenuated or eliminated after practice. This is exactly what we observed, with significant angle effects on mean amplitude in 500–600 and 600–700 ms windows before practice, but no significant angle effects after practice. Fig. 4 quantifies the time course of the RRN using linear regressions of mean amplitude on angle (plotted as gray lines). For the 500–600 and 600–700 ms windows with significant angle effects before practice, negativity increased

<table>
<thead>
<tr>
<th>Effects</th>
<th>Accuracy</th>
<th>Mean Correct RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle</td>
<td>$F_{(4,40)} = 14.9, p &lt; .001, \eta^2 = 0.60$</td>
<td>$F_{(2,20)} = 25.6, p &lt; .001, \eta^2 = 0.72$</td>
</tr>
<tr>
<td>Match</td>
<td>$F_{(1,10)} = 32.9, p &lt; .001, \eta^2 = 0.77$</td>
<td>$F_{(1,10)} = 46.4, p &lt; .001, \eta^2 = 0.82$</td>
</tr>
<tr>
<td>Practice</td>
<td>$F_{(1,10)} = 23.4, p &lt; .001, \eta^2 = 0.7$</td>
<td></td>
</tr>
<tr>
<td>Angle × match</td>
<td>$F_{(4,40)} = 3.1, p = .02, \eta^2 = 0.24$</td>
<td>$F_{(1,19.6)} = 20.5, p &lt; .001, \eta^2 = 0.67$</td>
</tr>
<tr>
<td>Match × practice</td>
<td>$F_{(1,10)} = 23.4, p &lt; .001, \eta^2 = 0.7$</td>
<td></td>
</tr>
<tr>
<td>Angle × match × practice</td>
<td>$F_{(1,10)} = 23.4, p &lt; .001, \eta^2 = 0.7$</td>
<td></td>
</tr>
<tr>
<td>Practiced angle</td>
<td>$F_{(4,40)} = 4.47, p = .004, \eta^2 = 0.32$</td>
<td>$F_{(1,19.19)} = 5.2, p = .002, \eta^2 = 0.34$</td>
</tr>
<tr>
<td>Practiced match</td>
<td>$F_{(1,10)} = 6.1, p = .03, \eta^2 = 0.38$</td>
<td>$F_{(4,40)} = 5.5, p = .001, \eta^2 = 0.35$</td>
</tr>
<tr>
<td>Practiced angle × match</td>
<td>$F_{(1,10)} = 8.17 p &lt; .004, \eta^2 = 0.45$</td>
<td></td>
</tr>
<tr>
<td>Practiced angle (same)</td>
<td>$F_{(4,40)} = 3.5 p = .01, \eta^2 = 0.26$</td>
<td>$F_{(1,17)} = 8.17 p &lt; .004, \eta^2 = 0.45$</td>
</tr>
<tr>
<td>Practiced angle (different)</td>
<td>$F_{(4,40)} = 3.5 p = .01, \eta^2 = 0.26$</td>
<td>$F_{(1,17)} = 8.17 p &lt; .004, \eta^2 = 0.45$</td>
</tr>
</tbody>
</table>
Fig. 3. Grand average waveforms for same and different pairs. The left column gives results for pairs when they were novel (i.e., tested in Session 1) and the second column for pairs when they were practiced (i.e., tested in Session 6). The third column gives results for pairs tested for the first time in Session 6 (Transfer) of Experiment 2. The first two pairs of rows, respectively, give Pz grand averages for Experiments 1 and 2. The final pair of rows shows grand averages at Oz for Experiment 2.
Fig. 4. Dark lines and symbols (s = same pairs, d = different pairs) are mean amplitude at Pz in the first (N = novel) and final (P = practiced) sessions over successive 100 ms windows (e.g., N 200–300 ms = Novel set from 200–300 ms, N 300–400 ms = Novel 300–400 ms, and so on). Thick gray lines in each panel are fits of linear regressions as a function of angle based on all data points in a panel.
by 1.5 and 1.2 μV/180°, respectively. By comparison, after practice, the largest estimated angle effect, occurring in the 400–500 ms window, was only 0.7 μV/180°. This trend may be driven either by some participants on some trials persisting with mental rotation or, as previously suggested, increased rotation speed.

5. Experiment 2

The attenuation of the RRN with practice found in Experiment 1 is consistent with the dual-process account (Tarr & Pinker, 1989), with initial responding based on an algorithmic process eventually being replaced by retrieval of the correct response from memory after extensive practice (Logan, 1988). However, these results might also be compatible with the single-process account. That is, it is possible that practice both increases rotation speed, and so decreases the angle effect on mean RT, and decreases the energetic expenditure required to perform rotation, and so decreases the RRN angle effect. Experiment 2 tests this possibility.

Experiment 2 used many more stimuli than Experiment 1, and so provided fewer opportunities to remember the stimulus–response relationship for each item. Given the paucity of the item-specific practice that is necessary to support the retrieval strategy, we predict that performance improvements in Experiment 2 will be due to increased rotation speed, with a substantial linear angle effect on mean RT remaining after practice. To be consistent with a single-process explanation of the ERP results in Experiment 1, such improvements should be accompanied by a reduced RRN. In contrast, the dual-process explanation of Experiment 1 is supported if the RRN does not reduce. Comparison between results for practiced and transfer stimuli allows us to test whether participants in Experiment 2 used mixed strategies.

To the degree that the mental rotation process is general purpose (i.e., it works at the same speed for all stimuli, or at least stimuli with equivalent complexity), the single-process account predicts transfer of performance improvements to previously unpractised stimuli. To test this prediction, in the final session of Experiment 2, participants responded to a previously unpractised set of “transfer” stimuli. As for Experiment 1, we continue to refer to stimuli used throughout the experiment as “novel” when tested in the first session and “practiced” when tested in the final session.

6. Method

6.1. Participants

Participants were 12 members of the general public (7 male) with ages ranging from 23 to 36 years (\(M = 28\)). All were in good health, had either normal or corrected to normal vision, and gave written informed consent. One participant’s data were lost due to a computer error.
6.2. Stimuli

To create the much larger stimulus set required for Experiment 2, we used more complex block figures than in Experiment 1. Block stimuli, seen in Fig. 1b, differed from Experiment 1 in that (a) the longest section (spine) increased from four to six blocks, (b) the longest projection from the spine (arm) was increased from two to three blocks, and (c) the block elements were scaled so that the overall stimulus had the same size as Experiment 1 stimuli. We created a large set of standard images, with vertically oriented spines, by placing the short and long arms at different positions on the four sides of the spine. Only images that were unique under picture plane rotation and mirror image reflection were retained in this standard set. For each unique standard ten further images were generated, made up of five picture-plane rotations (0°, 45°, 90°, 135°, 180°) of the standard and five rotations of its mirror image. Stimulus pairs consisted of the standard presented on the left and one of the ten images generated from it on the right.

Pilot testing–guided selection of two subsets (denoted as A and B) of 32 standards on the basis that no pair generated from the standard (640 pairs in total) contained occlusions making the correct match response unclear. From the remaining pairs a further subset of 320 (denoted C) was selected that minimized any ambiguity due to occlusion. The A and B sets were used as practice and transfer stimuli, responses to which were analyzed, and the C set was used for “filler” stimuli, responses to which were not analyzed.

6.3. Procedure

Fig. 1b depicts the time course of a trial in Experiment 2. The procedure was identical to Experiment 1 with the following exceptions. Participants performed five blocks in the initial session, six in the second session, and eight blocks per session thereafter (shorter initial sessions equated all session times at approximately 1 h). Attribution of sets A and B to practice and transfer sets was counterbalanced over participants. In the first session, each of the 320 set A (respectively B) stimuli was used once, with the remaining 80 trials based on stimuli selected randomly and without replacement from set C. In the second session, the number of set C stimuli was increased to 160 and for all but the final session all set C stimuli were used. As the time to perform trials decreased with practice, set C stimuli were included in this way to approximately equate session times. In the final session, participants responded to a previously unpractised set of 320 transfer stimuli that were drawn from the same pool as the practiced stimuli (i.e., A if B was practiced or vice versa) that replaced set C stimuli.

7. Results

For Experiment 2 no responses faster than the lower outlier criterion occurred; 0.13% of responses faster than the upper criterion were removed, as were 2.2% of trials due to EEG artifacts. Table 2 reports significant ANOVA effects for accuracy and mean correct
RT results, and Table 3 reports significant effects from a mixed ANOVA comparing Experiments 1 and 2 results for accuracy, mean correct RT, and ERP mean amplitude collapsed across a 400–700 ms post-stimulus window.

7.1. Accuracy and mean RT

Fig. 2 shows error rates and mean RT averaged over participants for practiced stimulus pairs when they were novel (Session 1) and after practice (Session 6), and also for transfer pairs tested for the first time in Session 6, with Greenhouse–Geisser corrections. In contrast to Experiment 1, error rates decreased significantly with practice as did the angle effect on errors (see Fig. 2), and the angle effect remained after practice. Overall, for practiced pairs the angle and match effects interacted, and although there was a marginally significant decrease in this interaction with practice, the angle effect remained in the final session. No tests contrasting error rates for practiced and transfer pairs in the final session approached significance.

Mean RT for practiced pairs displayed main effects of practice, angle, and match. Both the angle and match effects decreased with practice, and the reduction in the angle effect for same pairs was significantly greater than that for different pairs. However, after practice, the match main effect and the effect of angle for both same and different pairs remained significant. No tests contrasting mean RT for practiced and transfer pairs in the final session approached significance.

7.2. Event-related potentials

Fig. 3 shows grand average waveforms at Pz for same and different practiced pairs in Session 1 (novel) and Session 6 (practiced), and for pairs in Session 6 that had not previously

<table>
<thead>
<tr>
<th>Effects</th>
<th>Accuracy</th>
<th>Mean Correct RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle</td>
<td>$F_{(1,5,15)} = 20, p &lt; .001, \eta^2 = 0.67$</td>
<td>$F_{(1,2,12)} = 50, p &lt; .001, \eta^2 = 0.83$</td>
</tr>
<tr>
<td>Match</td>
<td>$F_{(1,10)} = 20.8, p = .001, \eta^2 = 0.68$</td>
<td>$F_{(1,10)} = 37.4, p &lt; .001, \eta^2 = 0.79$</td>
</tr>
<tr>
<td>Practice</td>
<td>$F_{(4,40)} = 4.36, p = .005, \eta^2 = 0.3$</td>
<td>$F_{(1,10)} = 85.6, p &lt; .001, \eta^2 = 0.9$</td>
</tr>
<tr>
<td>Angle $\times$ match</td>
<td>$F_{(2,5,25)} = 4.7, p = .01, \eta^2 = 0.32$</td>
<td>$F_{(1,7,17)} = 17, p &lt; .001, \eta^2 = 0.63$</td>
</tr>
<tr>
<td>Angle $\times$ practice</td>
<td>$F_{(2,7,27)} = 2.89, p = .06, \eta^2 = 0.22$</td>
<td>$F_{(1,10)} = 7.3, p = .02, \eta^2 = 0.42$</td>
</tr>
<tr>
<td>Match $\times$ practice</td>
<td>$F_{(1,8,18)} = 4.9, p = .02, \eta^2 = 0.33$</td>
<td>$F_{(1,6,16)} = 4.3, p = .04, \eta^2 = 0.3$</td>
</tr>
<tr>
<td>Practiced angle</td>
<td>$F_{(1,5,15)} = 5.6, p = .02, \eta^2 = 0.36$</td>
<td>$F_{(1,3,13)} = 22.7, p &lt; .001, \eta^2 = 0.69$</td>
</tr>
<tr>
<td>Practiced match</td>
<td>$F_{(2,5,25)} = 3.9, p = .03, \eta^2 = 0.28$</td>
<td>$F_{(1,10)} = 25.4, p &lt; .001, \eta^2 = 0.72$</td>
</tr>
<tr>
<td>Practiced angle (same)</td>
<td>$F_{(1,5,15)} = 5.6, p = .02, \eta^2 = 0.36$</td>
<td>$F_{(1,2,12)} = 24.6, p &lt; .001, \eta^2 = 0.71$</td>
</tr>
<tr>
<td>Practiced angle (different)</td>
<td>$F_{(1,5,15)} = 5.6, p = .02, \eta^2 = 0.36$</td>
<td>$F_{(1,2,12)} = 16, p = .001, \eta^2 = 0.62$</td>
</tr>
</tbody>
</table>
been tested (transfer). A strong RRN is evident before practice, but, in contrast to Experiment 1, is not reduced by practice. Instead, practice caused the RRN to move into an earlier onset and a narrower time range (400–700 ms). To quantify these findings, we analyzed average amplitude over 100 ms windows as we did for Experiment 1 (see Fig. 5).

For novel (i.e., Session 1) pairs, significant main effects of angle occurred in the 500–600, 700–800, and 800–900 ms windows. For these intervals the linear regression angle effect estimates were 2.2, 1.1, and 1.4 μV/180°, respectively. After practice a significant angle effect for these stimuli emerged earlier, in the 400–500 ms window, and persisted into the 500–600 and 600–700 ms windows, with corresponding linear regression angle effect estimates of 1.5, 2.2, and 1.5 μV/180°, respectively. All of these effects, except the latter one, remained significant after Bonferroni correction. This pattern was supported by within-subjects ANOVA results, with significant angle × practice interactions observed in the 400–500 and 800–900 ms windows only (p = .02 and p < .005, respectively).

### Table 3

Experimental comparison with significant effects on accuracy, mean reaction time (RT), and ERP amplitude (between 400–700 ms) for correct responses from mixed one-way between (Experiment: 1 vs. 2) × three-way within (Angle × match × practice) ANOVAs

<table>
<thead>
<tr>
<th>Effects</th>
<th>Accuracy</th>
<th>RT</th>
<th>Mean amplitude 400–700 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>$F_{(1,20)} = 7$, $p = .02$, $\eta^2 = 0.26$</td>
<td>$F_{(1,20)} = 33.8$, $p &lt; .001$, $\eta^2 = 0.63$</td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td>$F_{(1,20)} = 16.5$, $p &lt; .001$, $\eta^2 = 0.45$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment × practice</td>
<td>$F_{(1,20)} = 30.4$, $p &lt; .001$, $\eta^2 = 0.6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle</td>
<td>$F_{(4,80)} = 5.4$, $p &lt; .001$, $\eta^2 = 0.21$</td>
<td>$F_{(1,88,37.7)} = 8.6$, $p = .001$, $\eta^2 = 0.3$</td>
<td>$F_{(2,8,56.6)} = 4.5$, $p = .008$, $\eta^2 = 0.18$</td>
</tr>
<tr>
<td>Experiment × angle</td>
<td>$F_{(1,88,37.7)} = 16$, $p &lt; .001$, $\eta^2 = 0.44$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match × practice</td>
<td>$F_{(4,80)} = 4.7$, $p = .002$, $\eta^2 = 0.19$</td>
<td>$F_{(2,6,51.7)} = 7$, $p &lt; .001$, $\eta^2 = 0.26$</td>
<td>$F_{(3,1,62.8)} = 3.8$, $p = .01$, $\eta^2 = 0.16$</td>
</tr>
<tr>
<td>Experiment × practice × angle</td>
<td>$F_{(2,6,51.7)} = 4.1$, $p = .01$, $\eta^2 = 0.17$</td>
<td>$F_{(3,1,62.8)} = 6.2$, $p &lt; .001$, $\eta^2 = 0.23$</td>
<td></td>
</tr>
<tr>
<td>Angle × match</td>
<td>$F_{(2,2,43.5)} = 3.2$, $p = .04$, $\eta^2 = 0.14$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment × match × angle</td>
<td>$F_{(2,2,43.5)} = 14.7$, $p &lt; .001$, $\eta^2 = 0.43$</td>
<td>$F_{(2,8,55.2)} = 3.3$, $p = .03$, $\eta^2 = 0.14$</td>
<td></td>
</tr>
<tr>
<td>Practice × match × angle</td>
<td>$F_{(2,5,50.9)} = 3.5$, $p = .03$, $\eta^2 = 0.15$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For transfer pairs, the angle main effect was significant in all the 100 ms windows spanning 300–700 ms. Corresponding linear angle effect estimates were 0.7, 1.9, 2.2, and 2.1 μV/180°, respectively, with all but the first effect surviving Bonferroni correction. As shown in Fig. 5, ERPs were very similar for practiced and transfer stimuli. As was the case for behavioral measures, no inferential tests on ERP measures involving a practiced versus transfer contrast approached significance.

Further investigation revealed a practice effect on an early perceptual (N1) component, which was prominent at occipital sites in both experiments (see Fig. 3). We quantified the practice effect on N1 by measuring peak latency and amplitude in a 100–200 ms window at Oz. Absolute peak amplitude increased with practice in Experiment 2, from −7.5 to −10.7 μV, \( F(1,10) = 40.9, p < .001 \). No other effects were significant for either peak amplitude or latency. Fig. 3 shows the corresponding grand averages at the central occipital electrode (Oz). Transfer stimuli produced an N1 of essentially the same magnitude as practiced stimuli (10.7 μV vs. 10.6 μV).

The effect of practice on N1 was specific to Experiment 2. Although a clear N1 component was also observed at Oz in Experiment 1, neither its latency nor peak magnitude changed significantly with practice. In particular, the trend for an increased peak amplitude with practice evident in Fig. 3 did not approach significance (\( F(1,20) = 0.6, p = .46 \)). Furthermore, when N1 peak amplitudes from both experiments were combined in a mixed ANOVA, there was a significant interaction between experiment and practice (\( F(1,20) = 28.19, p < .001 \)).

8. Discussion

Consistent with the increased complexity of the stimuli in Experiment 2, the average angle effect before practice (1300 ms/180°) and the advantage for same over different pairs (greater than 600 ms) were more than twice as large as in Experiment 1. Practice also had a very large effect on mean RT: an overall reduction of around 800 ms, a greater than 600 ms/180° reduction in the angle effect, and a more than 500 ms reduction in the match effect (i.e., the difference between same and different pairs). Despite these large improvements, after practice, the angle effect in Experiment 2 was still larger than before practice in Experiment 1 (see Table 3 for inferential results supporting these differences between experiments).

ERP results revealed an even more marked contrast between experiments. Before practice in Experiment 1 there was a clear RRN over the 500–700 ms post-stimulus range, but this largely disappeared after practice, consistent with a shift from a rotation to a retrieval strategy. In Experiment 2, the RRN occurred over a broader range before practice (500–900 ms) and remained after practice. Practice did not affect the magnitude of the RRN as measured by the linear angle effect within 100 ms time windows, which peaked at 2.2 μV/180° in the 500–600 ms interval for novel and practiced stimuli in Experiment 2. Instead, it caused the RRN to contract into an earlier time range (400–700 ms), with onset in the 400–500 ms range (see Table 3 for supporting inferential analyses comparing ERP magnitude in the 400–700 ms range between experiments).
Fig. 5. Dark lines and symbols (s = same pairs, d = different pairs) are mean amplitude at Pz practiced pairs in the first (N = novel) and final (P = practiced) sessions over, and for pairs tested for the first time in Session 6 (T = Transfer) in successive 100 ms windows (e.g., N 200–300 ms = Novel set from 200–300 ms, N 300–400 ms = Novel 300–400ms, and so on). Thick gray lines are linear regressions as a function of angle for all data points in a panel, and probability values.
The results of Experiment 2 are consistent with a single-process explanation where practice improves mental rotation speed, resulting in both an earlier RRN onset and a reduced effect of angle on mean RT. The close equivalence between behavioral and ERP results for practiced and transfer stimuli also support a single-process explanation. This equivalence suggests that participants in Experiment 2 engaged in little or no retrieval-based processing, but instead predominantly used a general-purpose mental rotation process whose efficiency does not depend on whether a specific stimulus has been previously practiced. The results of Experiment 2 stand in marked contrast to the support for the dual-process model found in Experiment 1.²

Analysis of an early visual N1 component indicated that the improved performance in Experiment 2, and in particular, the earlier onset of the RRN might have its basis, at least in part, in perceptual learning. Consistent with this possibility, earlier RRN onset has been reported as difficulty of perceptual and object encoding processes were decreased prior to rotation (Heil & Rolke, 2002). It has been found that N1 was affected by learning about viewpoint invariance in two-dimensional projections of three-dimensional shapes (Wang & Bingo, 2010). Furthermore, Gauthier, Curran, Curby, and Collins (2003) reported that participants with expertise about a class of objects showed an increased occipitotemporal N170 component for these objects. We observed a significant increase in N1 peak amplitude with practice in Experiment 2 but not in Experiment 1. The specificity of the N1 increase to Experiment 2, a cause associated with practicing with a variety of different stimuli, such as improved ability to extract three-dimensional representations from the two-dimensional wireframe images used in our experiments.

![Fig. 6. A state-trace plot of EEG measures versus the linear angle effect (linear regression coefficients for angle averaged over same and different pairs) on mean RT: (a) Plots the linear angle effect on average amplitude 400–700 ms post-stimulus at Pz, and (b) plots peak amplitude at Oz 100–200 ms post-stimulus. Note that error bars indicated a bias-corrected within-subject standard error (Morey, 2005).](image-url)
8.1. Reversed associations

The finding that increased rotation speed in Experiment 2 was unaccompanied by the RRN reduction seen in Experiment 1 supports the dual-process explanation of the Experiment 1 practice effect. Fig. 6a encapsulates this inference in the form of a state-trace plot (Bamber, 1979; Prince, Brown, & Heathcote, 2012) of linear-angle effects on RRN amplitude against linear-angle effects on mean RT. The decreased RRN in the first experiment, but no change (a single dissociation), or even an increase (a double dissociation) in the second experiment, is suggestive of a two-process explanation. However, dissociations do not logically compel this conclusion; to do so requires a reversed association (Dunn & Kirsner, 1988; Henson, 2006), such as is shown in Fig. 6a.

A reversed association occurs when a monotonic (i.e., always non-increasing or always non-decreasing) line cannot join the points in a state-trace plot. This is clearly the case for Fig. 6a, where the average amplitude for the RRN was calculated over a 400–700 ms window. The solid arrow pointing up and to the left indicates that under conditions encouraging item-specific learning (Experiment 1) the decreasing angle effect on mean RT is associated with a decreasing RRN. The dashed arrow pointing down and to the left indicates a decreasing angle effect on mean RT with an increasing RRN, under conditions discouraging item-specific learning (Experiment 2). Together these effects constitute a reversed association. The same failure of monotonicity applies to state-trace plots using earlier (i.e., 400–500 and 400–600 ms) windows that also reflect the earlier onset of the RRN after practice in Experiment 2.

Fig. 6b shows that there is also a reversed association between behavior and the peak amplitude of the N1 visual-evoked potential, further confirming that different brain mechanisms are involved in the development of expertise in the two experiments. Practice caused a decrease in the angle effect on behavior in both experiments and a strong increase in the peak negativity of the N1 in Experiment 2, but no change or a small decrease in Experiment 1. This pattern is consistent with N1 indexing the extraction of a three-dimensional interpretation from the two-dimensional stimuli. In Experiment 1, a three-dimensional representation was not required because a two-dimensional representation is sufficient to retrieve the correct answer. In contrast, mental rotation processing in Experiment 2 can only be accomplished based on a three-dimensional representation. Increased N1 negativity after practice in Experiment 2 appears to be a marker for the construction of a higher quality three-dimensional representation that enables the mental rotation process to begin earlier, complete more quickly, and to be less error prone (see Fig. 2).

9. General discussion

Spatial cognition, specifically mental rotation, is an important component of reasoning, decision making, and navigation. Hence, quantifying (e.g., Sutton, Heathcote, & Bore, 2007) and improving (e.g., Wright et al., 2008) these skills is an important goal for brain and behavioral sciences. Our behavioral results, and those of Wright et al. using a similar
level of training, demonstrate that people can greatly improve performance in shape matching through improved mental rotation skill, and that, with appropriate training, that skill can generalize to related stimuli and tasks.

We reported converging behavioral and neurophysiological evidence implicating two different routes to the development of expertise in mental rotation. Extensive practice with a small set of stimuli allowed participants to switch from a mental rotation strategy before practice, evidenced by increased mean RT and ERP negativity (rotation-related negativity, RRN) with rotation angle, to direct retrieval from memory the correct response associated with each stimulus, evidenced by minimal angle effects remaining after practice. The development of a retrieval strategy is generally consistent with Logan’s (1988) instance theory of automaticity. However, our ERP results shed new light on the ongoing debate (e.g., Bajic & Rickard, 2009, 2011) about whether the algorithmic process (in our case mental rotation) continues to race with retrieval as practice progresses. The disappearance of the RRN after practice suggests that after extensive practice there is a switch to purely retrieval-based responding (Rickard, 1997, 1999).

In a second experiment that minimized stimulus-specific learning by using a much larger stimulus set, the same level of practice as in the first experiment still markedly reduced the angle effect on mean RT but not the RRN; instead, the RRN was undiminished and emerged earlier. This reversed association between ERP and behavioral measures over the two experiments rules out a single-route explanation of the development of mental rotation expertise. Equivalent behavioral and RRN results for practiced and transfer (i.e., unpractised) stimuli indicate that the training methods used in the second experiment promoted the development of a generalized mental rotation skill. Taken together, the results of our experiments support an augmented dual-process theory where practice can both increase the efficiency of mental rotation (Bethell-Fox & Shepard, 1988) and cause a switch to direct retrieval of the correct answer from memory (Tarr & Pinker, 1989). Although our results are not definitive on this point, it seems likely that both types of improvement occur together, so that even with a large (but finite) stimulus set, such as was used in our second experiment, very long-term practice would eventually lead to responding based on retrieval.

Perhaps our most novel finding was that, in our second experiment, we observed an increased peak amplitude for an early visual (N1) component that also displayed stimulus generalization (i.e., the increase was equal for practiced and transfer stimuli). The increased N1 effect was specific to our second experiment (i.e., there was no significant increase in N1 in our first experiment). We argue that this specificity, and particularly the strong transfer, support the conclusion that participant’s experience with a large variety of two-dimensional wireframe images lead them to develop a general-purpose expertise in extraction of three-dimensional information. This, in turn, facilitated an earlier onset and faster rate of mental rotation, and so reduced the angle effect in the shape-matching task. Future studies could make use of this training methodology and new N1 index of perceptual proficiency to further explore the cognitive and neural processes underpinning the development and generalization of skill in mental rotation.
Acknowledgments

This work was supported by pilot grants from Macquarie University and the University of Newcastle and by Australian Research Council (grants DP110100234 and LEO668421). We gratefully acknowledge the collaboration of Kanazawa Institute of Technology and Yokogawa Electric Corporation in establishing the KIT-Macquarie MEG laboratory. Thanks to Graciela Tesan, Chris Sewell, Melanie Reid, David Elliot, and E/Prof Pat Michie.

Notes

1. For both experiments, we also analyzed the same mean amplitude windows at FPz to investigate whether our experimental manipulation was confounded by a difference in saccades. We reasoned that if the angle \times practice interaction at Pz was due to eye-motion artifact, this should be evident at FPz, and also possibly be larger and emerge earlier than at Pz. We found no significant angle–practice interaction, nor systematic effects such as those observed at Pz.

2. We ran a control experiment to test whether the behavioral results found in Experiment 1 would be replicated with the more complex stimuli and younger subject population used in Experiment 2, and to test whether transfer occurred when practice used a small set of the more complex stimuli. The control experiment confirmed the prediction made by the dual-process model that no transfer would be found with the small practice set, and also that the stimulus and age differences did not confound our results. For practiced stimuli, the average linear angle effect on correct mean RT in the final session of the control experiment was only 119 ms/180°, whereas for transfer stimuli it was almost an order of magnitude greater at 1059 ms/180°.

References


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

- Figure S1. Learning curves for mean RT averaged over participants, at each angle by session, for Experiments 1 and 2.
- Data S1. PDF containing naming conventions for behavioral data from Experiment 1 and 2.
- Data S2. Behavioral data, Experiment 1.
- Data S3. Behavioral data, Experiment 2.

**Appendix:**

**Mixed-design ANOVAs**

This appendix contains a table of the results of mixed two-way between (Experiment 1, Experiment 2) × three-way within (Angle × match × practice) ANOVAs. This analysis complements and confirms the state trace analysis reported in the main body of the study. In particular, an experiment by practice by angle interaction was observed for mean RT. We also applied the mixed ANOVA to all of the ERP measures and found an experiment by practice by angle interaction between 400–500 ms and 600–700 ms. This pattern is observed, and strengthened, when we collapse across 400–700 ms time windows; results for the collapsed analysis are reported in the following table.