Training generalized spatial skills

Rebecca Wright
Oxford University, Oxford, England

William L. Thompson
Harvard University, Cambridge, Massachusetts

Giorgio Ganis
Harvard University, Cambridge, Massachusetts,
Harvard Medical School, Boston, Massachusetts,
and Massachusetts General Hospital, Charlestown, Massachusetts

Nora S. Newcombe
Temple University, Philadelphia, Pennsylvania

AND

Stephen M. Kosslyn
Harvard University, Cambridge, Massachusetts
and Massachusetts General Hospital, Charlestown, Massachusetts

Spatial transformation skills are an essential aspect of cognitive ability. These skills can be improved by practice, but such improvement has usually been specific to tasks and stimuli. The present study investigated whether intensive long-term practice leads to change that transcends stimulus and task parameters. Thirty-one participants (14 male, 17 female) were tested on three cognitive tasks: a computerized version of the Shepard–Metzler (1971) mental rotation task (MRT), a mental paper-folding task (MPFT), and a verbal analogies task (VAT). Each individual then participated in daily practice sessions with the MRT or the MPFT over 21 days. Postpractice comparisons revealed transfer of practice gains to novel stimuli for the practiced task, as well as transfer to the other, nonpracticed spatial task. Thus, practice effects were process based, not instance based. Improvement in the nonpracticed spatial task was greater than that in the VAT; thus, improvement was not merely due to greater ease with computerized testing.

Spatial skills play a key role in many types of reasoning and communication and are important in domains such as mathematics, natural sciences, and engineering. Discovering how to increase one's level of spatial functioning is therefore an important goal. Fortunately, performance of spatial tasks can be improved through practice and training (see, e.g., Baenninger & Newcombe, 1989); this has wide implications for instruction and education.

However, the nature of this improvement is unclear. Some researchers have claimed that practice leads people to make fundamental changes in how they process spatial stimuli, which allows practice to transfer to novel stimuli and new tasks (Bethell-Fox & Shepard, 1988). Specifically, researchers have claimed that improvement in spatial processing can generalize to novel stimuli within the same task (e.g., Leone, Taine, & Droulez, 1993), to other tasks of the same general type (e.g., De Lisi & Cammarano, 1996), and to tasks that share underlying cognitive processes with the practiced task (e.g., Wallace & Hofe-

lich, 1992). Others, however, have reported that improvement in one spatial task does not transfer to other spatial tasks (e.g., Stericker & LeVesconte, 1982). In fact, practice has often been studied in paradigms using the same stimuli multiple times (e.g., Kail, 1986), thus leaving open the possibility that gains are achieved by instance-based memory rather than by skill improvement. Recently, a National Academy of Sciences panel concluded that transfer of improved spatial skills has not been convincingly demonstrated, and called for research aimed at improving spatial performance in a generalizable way (Committee on Support for Thinking Spatially [CSTS], 2006).

According to the present analysis, studies of the development of spatial skill must have five characteristics in order to document general improvement. First, they should permit analysis of tasks' component processes. Tasks such as mental rotation are generally thought to consist of four processes (Cooper & Shepard, 1973), only one of which (rotation) involves spatial transformation: (1) the initial

visual encoding of the stimuli; (2) rotating one object (typically into congruence with another); (3) comparing objects to decide whether they are the same or different; (4) and finally, responding. Tests that only offer an overall measure of performance cannot specify the individual contributions of particular underlying processes. In this study, we used computerized testing in order to facilitate the componential analysis of response times (RTs). RTs and errors were each decomposed to reflect two components: transformation processes (slope) and other processes (intercept).

Second, gains within a specific task should be assessed by testing with novel stimuli following practice, in order to rule out instance-based improvement (i.e., memory for specific items). Such assessments have been limited by small sets of stimuli (e.g., the Vandenberg test has only 6 basic block configurations; we used 48 unique block configurations in our task).

Third, transfer to another spatial task should be assessed symmetrically so that one group is trained on Task A and transfer is assessed on Task B, whereas another group's training is the opposite. Here, we used a mental rotation task (MRT) along with a mental paper-folding task (MPFT) (based on one devised by Shepard & Feng, 1972). We hypothesized that gains through practice in mental rotation would transfer to the transformational aspects of the MPFT and vice versa. However, the two tasks overlap in additional subprocesses (e.g., spatial encoding; Wallace & Hofelich, 1992), and thus transfer might occur for other reasons.

Fourth, to differentiate domain-specific improvement from transfer effects caused by factors such as familiarity with the test situation, one must include a nonspatial task. Verbal and spatial skills are distinct (Halpern, 1992); hence, improvement in spatial ability should not generalize to verbal tasks. We used a verbal analogies task (VAT) as the control condition.

Fifth, training should be intensive enough to produce large gains, which in turn can produce clear transfer.

Thus, we designed this study first and foremost to discover whether the effects of practice transfer between different spatial tasks. In addition, if such transfer does occur, the design allows us to break down the RTs and errors into component parts, enabling us to identify the locus of such transfer.

METHOD

Participants

Participants were 38 volunteers (18 female, 20 male) recruited via a Harvard University Psychology Department Web site. Most were undergraduates, with a mean age of 23.8 years (range, 18–43 years). Six male participants and 1 female participant failed to take part in the required number of practice sessions, so the results are from 31 participants (17 female, 14 male). Analyses revealed no performance criteria linked to a dropout bias. Participants were tested according to all applicable regulations.

Tasks and Apparatus

The study consisted of three tasks. Stimuli consisted of a black background with a reference image on the left-hand side of a central fixation point and a comparison image on the right and were pre-

sented on a Macintosh computer (with a 16-in. screen). For each task, we presented comparable numbers of easy, medium, and hard trials.

MRT. A computerized version of the classic Shepard and Metzler (1971) task measured the ability to compare a pair of pictures of 3-D objects at different orientations and to decide whether they were identical or mirror images (see Figure 1). We created 48 basic block configurations of 3-D figures, resulting in 288 unique items (see also Peters & Battista, 2008, reporting 16 new base figures). The three levels of difficulty corresponded to whether the angular disparity between the objects in a pair was 50º (easy), 100º (medium), or 150º (hard).

MPFT. We prepared a computerized adaptation of the visuospatial task designed by Shepard and Feng (1972). In addition to the original 2-D unfolded cube templates, we included a reference 3-D cube image to create a comparison task analogous to the other tasks reported here (see Figure 2). Shepard and Feng prepared 165 unique stimuli (based on 11 structurally different configurations); we created 255 unique stimuli across difficulty levels. The difficulty levels corresponded to whether a single square (easy), two or three squares (medium), or between four and seven squares (hard) had to be carried.

VAT. This task was based on one devised by Morrison et al. (2004; see Figure 3). The participants were asked to compare relationships between two words in each of two simultaneously presented word pairs, to decide whether the relationship between the words in the left-hand pair was the same as the relationship between those on the right. We categorized the trials into three difficulty levels according to mean RTs for correct solutions, gathered from previous testing.

Each task included 228 trials in total. Pilot testing allowed us to equate the three tasks (as much as possible) for difficulty, and thus only a subset of the total number of stimuli created for each task was
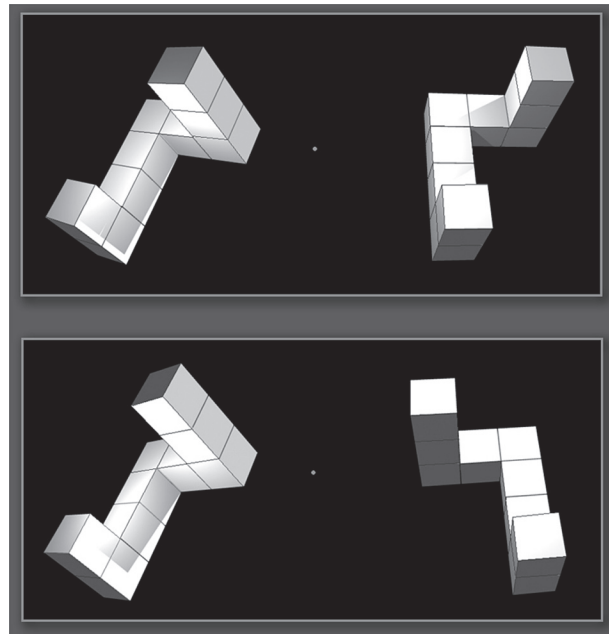
Figure 1. Example of mental rotation task (MRT) stimulus. Participants are asked to compare the images of these objects (which have a 3-D appearance) and to decide, using mental rotation, whether they represent the same object (when properly aligned) or whether they are different (i.e., mirror reversals of each other). The three levels of difficulty corresponded to whether the angular disparity between the objects in a pair was 50º (easy), 100º (medium), or 150º (hard). Nine female and 8 male participants were trained on the MRT. In the example on the top, the objects are the same. In the example on the bottom, the objects are different and cannot be rotated into congruence.
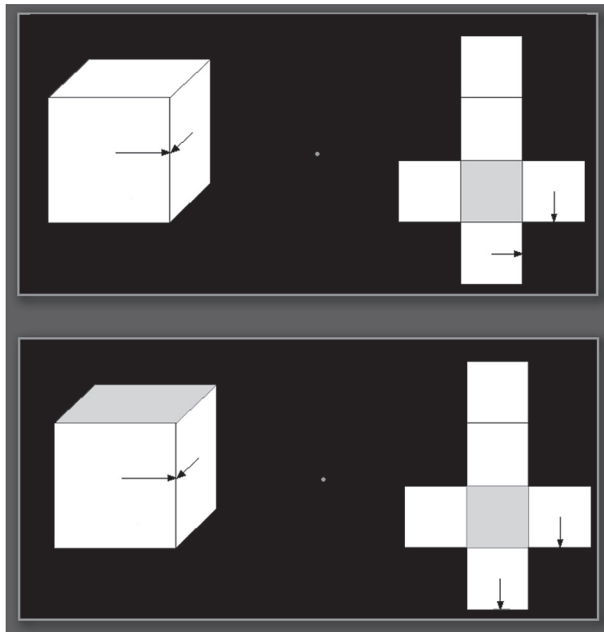
Figure 2. Example of mental paper-folding task (MPFT) stimulus. Participants are asked to mentally fold up the unfolded cube template and to decide, once they have done so, whether the arrows on the unfolded cube template would match (come together in the same way as) the arrows on the folded cube. Average latencies are linearly related to the number of squares carried along for each fold (Shepard & Feng, 1972). The difficulty levels corresponded to whether a single square (easy), two or three squares (medium), or four to seven squares (hard) had to be carried. Eight female and 6 male participants were trained on the MPFT. In the case represented on the top panel, the gray square represents the bottom of the cube (not visible on the folded cube because of occlusion, in the object on the left), and an upward mental folding of the two squares with arrows on the unfolded template is required to determine that the arrows on both models would come together in the same way. In the case depicted on the bottom panel, a similar mental folding of the two squares with arrows on the unfolded template is required in order to determine that the arrows on both models would not meet in the same way.

used. Trials were subdivided for presentation throughout different phases of the study (see Figure 4). Trials within each phase were distributed according to correct response (approximately half *same* and half *different*) and difficulty level. At most, three trials of any type (e.g., *same* or *different*; easy, medium, or hard) were presented in sequence.

Procedure

Initial laboratory session. The participants completed all three tasks described above. Tasks were presented in a counterbalanced order, with each of the three tasks occupying each ordinal position across participants. The participants sat 50 cm from the monitor and first completed 12 familiarization trials, with stimuli different from those in the experimental trials. Responses were made by pressing keys labeled "Same" and "Different." The participants were asked to perform the tasks as quickly and accurately as possible.

Practice phase. After completing the initial laboratory session, participants were assigned to one of two groups: those practicing only the MRT and those practicing only the MPFT. This phase of the study was conducted over the Internet. The participants completed daily practice sessions over 21 consecutive days. The practice phase consisted of 114 trials (as did the initial and final laboratory phases) and required the participants to spend about 15–20 min per day on the task. Trials were presented in a random sequence on each

session. The participants could miss up to three sessions without forfeiting their place in the study.

Final laboratory session. The day after completing the practice phase, the participants returned to the laboratory to be retested on all three tasks. The procedure was identical to that of the initial laboratory session (the three tasks were presented in the same order for a given participant), except that about half of the stimuli in each task were completely novel, and about half had been encountered just once, in the initial laboratory session (see Figure 4).

Results

Results are collapsed over *same* and *different* trials. Only trials with correct responses were included in the analyses. Participants had a 6-sec limit in which to respond. Outlier trials (fewer than 5%) contained RTs that were greater or less than 2.5 standard deviations (*SD*s) from the corresponding mean cell value for that participant. We calculated slopes and intercepts for each participant for each task. Slopes and intercepts were obtained from the regression line through the means of each of the three levels of difficulty. We performed 2 (practice group) $\times$ 2 (sex) $\times$ 2 (session) $\times$ 3 (difficulty level) ANOVAs; if the data are broken down into slope or intercept, we used a 2 (practice group) $\times$ 2 (sex) $\times$ 2 (session) design.

Effects of Practice

The effects of practice first needed to be documented; otherwise, we would not be justified in examining possible transfer of practice.

Data from the 3-week practice phase failed to meet Mauchly's test of sphericity, so the degrees of freedom were adjusted according to the Greenhouse–Geisser epsilon. Overall RTs decreased across successive sessions
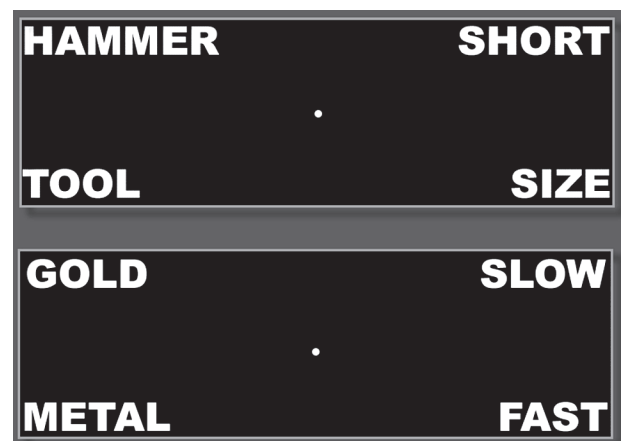
Figure 3. Example of verbal analogies task stimulus. Participants were asked to compare the word pairs on the left and right sides of the screen and to determine whether the relationship between the words on the right side of the screen was the same as the relationship between the words on the left side of the screen. We categorized the trials into three difficulty levels according to mean response times (RTs) for correct solutions, gathered from previous testing. The pilot work had ensured that RTs and errors were comparable to those from the two spatial tests, to enhance comparability of effects. In the case represented in the top panel, the relationship is the same (one word is an item within the wider category represented by the other word). In the case at the bottom, the relationship between the words is different.
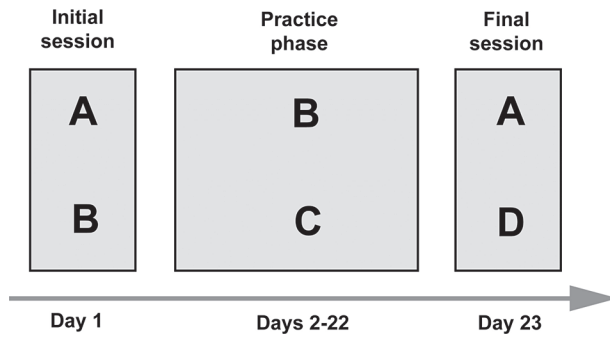
Figure 4. Counterbalancing of subsets of trials across the initial session, practice phase, and final session for the mental paper-folding task and mental rotation task training conditions (the verbal analogies task was not used in the practice phase; hence only the initial and final arrangements are applicable to it). The letters A, B, C, and D refer to groups of stimuli that were presented at each phase. In the initial session, for both tasks, participants first view stimuli (A and B). During the 3-week practice phase, about half of the stimuli (B) have already been seen in the initial session and about half (C) are completely novel. In the final lab session, there was a mix of stimuli previously seen in the initial session (A) and novel stimuli (D).

$[F(6,149) = 86.09, p < .001; \eta_p^2 = .76]$. This decrease—documenting an effect of practice—was due to changes in both slopes and intercepts $[F(5,127) = 10.25, p < .001; \eta_p^2 = .28,$ and $F(5,127) = 21.13, p < .001; \eta_p^2 = .44,$ respectively].

However, the MRT had shallower RT slopes than the MPFT did $[F(1,27) = 15.89, p < .001; \eta_p^2 = .37]$ and a more modest decline over practice $[F(5,127) = 4.37, p = .001; \eta_p^2 = .14]$. The MRT initially had larger intercepts than did the MPFT $[F(1,27) = 81.88, p < .001; \eta_p^2 = .75]$; this difference remained stable over practice. In addition, participants who practiced the MRT generally took longer than those who practiced the MPFT $[F(1,27) = 4.51, p = .04; \eta_p^2 = .14]$. Overall, errors did not decrease across the practice phase $[F(1,32) = 1.53, p = .23; \eta_p^2 = .05]$. Error slopes decreased over sessions $[F(6,154) = 5.21, p < .001; \eta_p^2 = .16]$, although there was no change in error intercepts $[F(7,183) = 0.79, p = .59; \eta_p^2 = .03]$.

Moreover, over both practice groups, participants were faster $[F(1,27) = 347.1, p < .0001; \eta_p^2 = .93]$ and made fewer errors $[F(1,27) = 81.12, p < .0001; \eta_p^2 = .75]$ during the final laboratory session than during the initial session. (This result also held within each of the tasks individually.) This was true in spite of the fact that about half of the stimuli in each task were completely novel during the final session, and the remainder had been seen only once during the initial laboratory session. Thus practice effects could not depend on memory for particular items (see Tables 1A and 1B for condition means). Slope and intercept analyses revealed that participants were faster on both measures for both practice groups. Errors declined for the MPFT group in terms of slopes, and for the MRT group in terms of intercepts (see Table 2 for statistics).

Transfer to Nonpracticed Spatial Task

Having found robust effects of practice, we assessed transfer across tasks. We began by comparing results from the practiced and nonpracticed spatial tasks in initial and

Table 1A
Mean Task Response Times (RTs, in Milliseconds), Slopes, and Intercepts, With Standard Errors, for the
Mental Paper-Folding Task (MPFT), the Mental Rotation Task (MRT), and the Verbal Analogies Task (VAT)

	MPFT						MRT						VAT					
	RT		Slope		Intercept		RT		Slope		Intercept		RT		Slope		Intercept	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Initial Session																		
MPFT group	2,965	184	758	99	1,448	170	3,166	132	240	37	2,686	152	2,933	155	237	30	2,459	173
MRT group	2,676	211	675	87	1,327	175	3,132	182	257	46	2,618	195	2,970	159	325	31	2,319	152
Final Session																		
MPFT group	1,079	126	398	106	282	106	2,338	170	261	55	1,817	160	2,721	134	210	39	2,301	121
MRT group	1,883	191	551	85	781	125	1,480	177	140	38	1,200	150	2,671	148	229	28	2,212	132

Note—Slope values represent mean increase per level of difficulty. Results are subdivided according to practice group (MPFT, $n = 14$; MRT, $n = 17$).

Table 1B
Mean Percentages of Error (%E, With Standard Errors) for the Mental Paper-Folding Task (MPFT),
the Mental Rotation Task (MRT), and the Verbal Analogies Task (VAT)

	MPFT						MRT						VAT					
	RT		Slope		Intercept		RT		Slope		Intercept		RT		Slope		Intercept	
	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE
Initial Session																		
MPFT group	18.7	2.0	16.3	2.7	−13.8	4.2	16.4	2.4	4.0	1.4	8.3	2.4	19.4	1.8	11.8	1.3	−4.2	2.7
MRT group	15.7	2.2	12.5	2.2	−9.3	3.1	16.5	2.3	4.3	1.3	7.9	2.6	16.4	1.4	12.8	1.1	−9.2	2.1
Final Session																		
MPFT group	5.7	1.0	5.0	1.4	−4.2	2.2	12.0	1.4	5.6	1.1	0.9	1.6	19.2	2.2	10.0	0.8	−0.7	2.0
MRT group	11.3	1.6	10.3	1.6	−9.3	2.0	5.6	1.5	3.4	0.8	−1.3	1.1	15.9	1.7	9.4	1.0	−2.9	1.8

Note—Slope values represent mean increase per level of difficulty. Results are subdivided according to practice group (MPFT, $n = 14$; MRT, $n = 17$).

Table 2
F Statistics, p Values, and Effect Sizes (η_p^2) Associated With Changes in Response Time (RT) and Error Performance Between the Initial and Final Sessions for the Practiced Task

	df	F	p	η_p^2
		RTs		
MPFT group				
Slope	1,12	8.13	.01	.40
Intercept	1,12	21.13	.0004	.66
MRT group				
Slope	1,15	5.53	.03	.27
Intercept	1,15	81.09	<.0001	.84
		Errors		
MPFT group				
Slope	1,12	14.00	.003	.54
Intercept	1,12	3.76	.08	.79
MRT group				
Slope	1,15	.24	.63	.02
Intercept	1,15	29.30	<.0001	.66

Note—MPFT, mental paper-folding task; MRT, mental rotation task.

final sessions. For overall RT (see Figure 5), we found an interaction between task (practiced vs. nonpracticed) and session (initial vs. final) [$F(1,27) = 102.86$, $p < .0001$; $\eta_p^2 = .79$]. Although the interaction revealed greater gains for the practiced task, planned contrasts revealed gains for both tasks ($p < .0001$; η_p^2 was .60 for the unpracticed task and .87 for the practiced task). The pattern held for both the MPFT and MRT groups. Analysis of slopes revealed transfer only for the nonpracticed MPFT in the MRT-practiced group. However, transfer was also found in the intercepts for the nonpracticed tasks in both groups (see Table 3). For errors (see Figure 6 and Table 3), an interaction between task and session was found [$F(1,27) = 29.88$, $p < .0001$; $\eta_p^2 = .53$], indicating greater gains for the practiced task. Nevertheless, transfer was evident in

the nonpracticed task for both MRT- and MPFT-practiced groups. When analyzed separately, there were no error slope or intercept gains for the nonpracticed task (see Table 3; also see Tables 1A and 1B for RT and error means in initial and final sessions for each task).

Transfer to the VAT?

Practicing a computerized spatial task might improve performance on another spatial task because of general factors, such as improved ease with the testing format, rather than because of shared skills between tasks. However, such interpretations are inconsistent with the greater improvement between initial and final sessions for the unpracticed spatial task, as compared with the VAT, for both RT [$F(1,27) = 34.04$, $p < .0001$; $\eta_p^2 = .56$] and error [$F(1,27) = 9.65$, $p = .004$; $\eta_p^2 = .26$]. See Table 4 for group transfer data for the unpracticed spatial task compared with that for the VAT, broken down by slope and intercept.

In order to ensure that this effect was not due to scale differences between the conditions, we normalized the data by calculating RT means and SDs for each participant in each condition (we focused on RTs because we could calculate SDs for each participant) and by extracting z scores of improvement from the initial to final session, with the values now on the same scale for each condition. Each practice group was examined separately, and we confirmed greater improvement on the unpracticed spatial task than on the VAT ($p < .0001$ in both cases; MRT group, $\eta_p^2 = .42$; MPFT group, $\eta_p^2 = .51$).

Comparison of Previously Seen With Novel Stimuli Within Task

Within each of the tasks, we also compared novel stimuli presented in the final session with stimuli that were also presented in the final session, but that had been seen
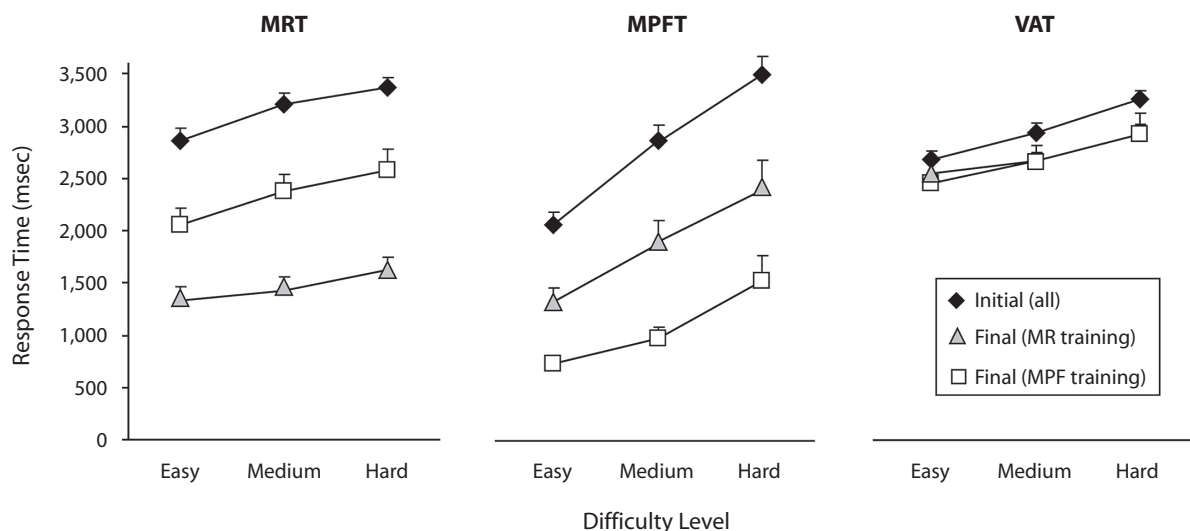
Figure 5. Response times in the three tasks in the initial versus final laboratory sessions. Note the reversal in the improvement pattern between the two tasks, depending on practice group. The amount of transfer from practicing the alternate task was approximately halfway between the initial values (equivalent to no practice) and the values associated with extensive practice on the target task. Error bars represent standard errors of the mean. MRT, mental rotation task; MPFT, mental paper-folding task; VAT, verbal analogies task.

Table 3
***F* Statistics, *p* Values, and Effect Sizes (η_p^2) Associated With Changes in Response Time (RT) and Error Performance Between the Initial and Final Sessions for the Unpracticed Spatial Task**

	df	F	p	η_p^2
		RTs		
MPFT group				
Slope	1,12	.08	.78	.0007
Intercept	1,12	22.60	.0005	.65
MRT group				
Slope	1,15	7.03	.02	.32
Intercept	1,15	14.08	.002	.49
		Errors		
MPFT group				
Slope	1,12	.40	.54	.03
Intercept	1,12	2.81	.12	.19
MRT group				
Slope	1,15	3.67	.07	.20
Intercept	1,15	.01	.92	.0007

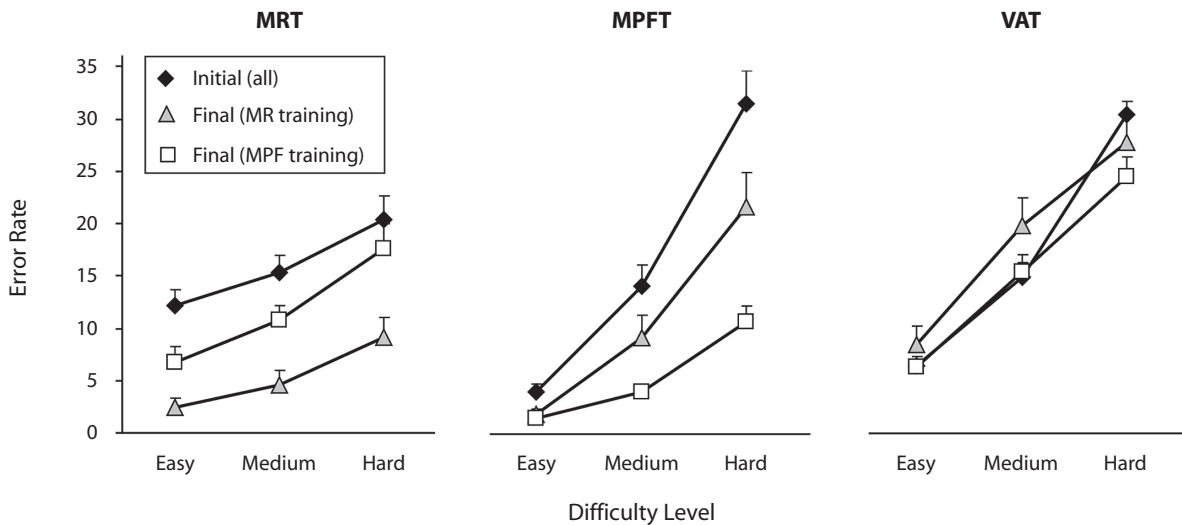Note—MPFT, mental paper-folding task; MRT, mental rotation task.

previously in the initial session (i.e., final-session A and D stimuli from Figure 4). Virtually complete transfer to novel stimuli was found for the MPFT, with no differences in overall RT or error between the novel and previously encountered trials in the final session. This held for both slopes and intercepts and for both practice groups ($p > .1$ in all cases). Thus, gains due to practice, whether directly obtained or due to transfer, did not occur because participants learned particular stimuli. For the MRT, completely novel stimuli required more time than did those encountered during the initial session [$F(1,27) = 9.16$, $p = .005$; $\eta_p^2 = .25$]. There was no difference for slopes, but intercepts in the MPFT-practiced group were higher for novel stimuli [$F(1,27) = 8.22$, $p = .007$; $\eta_p^2 = .23$]. The MRT-practiced group had an intervening period of 3 weeks of practice with a large number of stimuli, so dif-

ferences between those seen in the first session and those that were novel may have been less salient. In contrast, because these participants had less experience, stimuli encountered once before may have been more salient for the MPFT-practiced group. Although there was a trend for more errors to be made on novel stimuli [$F(1,27) = 3.9$, $p = .06$; $\eta_p^2 = .13$], there were no effects on slope or intercept when examined separately.

Transfer to Novel Stimuli?

Given the evidence above for instance-based learning in the MRT, we compared performance in the initial lab session (i.e., for all trials, A and B stimuli in the initial session; see Figure 4) with performance in the final lab session for only the novel trials (i.e., D stimuli in the final session; see Figure 4), in order to investigate whether process-based transfer would be evident for the novel stimuli taken alone. Participants showed performance transfer to novel trials [$F(1,27) = 194.35, p < .0001; \eta_p^2 = .88$]. There was greater transfer for those who practiced the MRT [$F(1,27) = 22.38, p < .0001; \eta_p^2 = .45$], although transfer was seen in both groups separately [MRT group, $F(1,15) = 168.07$, $p < .0001; \eta_p^2 = .92$; MPFT group, $F(1,12) = 47.85, p < .0001; \eta_p^2 = .80$]. Separate analyses showed that RT slopes flattened for the practiced MRT group [$F(1,15) = 6.66$, $p = .03; \eta_p^2 = .31$] but not for the MPFT group. RT intercepts decreased for both practice groups [MRT group, $F(1,15) = 79.49, p < .0001; \eta_p^2 = .84$; MPFT group, $F(1,12) = 29.39, p < .001; \eta_p^2 = .71$].

In addition, participants made fewer errors on the novel final-session trials than they did on the initial trials [$F(1,27) = 19.53, p < .0001; \eta_p^2 = .42$]. However, the advantage was greater for the MRT-practiced group [$F(1,27) = 5.65, p = .03; \eta_p^2 = .17$]. The MRT—but not the MPFT—group had relatively fewer errors in novel final trials [$F(1,15) = 23.58, p < .001; \eta_p^2 = .61$]. There was no effect on slope for either group; however, there was

Figure 6. Errors in the three tasks in the initial versus final laboratory sessions. A pattern similar to that seen for response times is visible. Error bars represent standard errors of the mean. MRT, mental rotation task; MPFT, mental paper-folding task; VAT, verbal analogies task.

Table 4
F Statistics, p Values, and Effect Sizes (η_p^2) Associated With Changes in Response Time (RT) and Error Performance Between the Initial and Final Sessions for the Unpracticed Spatial Task Versus the Verbal Analogies Task

	df	F	p	η_p^2
RTs				
MPFT group				
Slope	1,12	.31	.59	.03
Intercept	1,12	11.47	.005	.49
MRT group				
Slope	1,15	.08	.78	.005
Intercept	1,15	10.73	.005	.42
Errors				
MPFT group				
Slope	1,12	3.09	.10	.20
Intercept	1,12	6.55	.03	.35
MRT group				
Slope	1,15	.46	.51	.03
Intercept	1,15	4.57	.05	.23

Note—MPFT, mental paper-folding task; MRT, mental rotation task.

evidence of transfer to the error intercepts for both groups [MPFT group, $F(1,12) = 7.63$, $p = .02$; $\eta_p^2 = .39$; MRT group, $F(1,15) = 7.47$, $p = .02$; $\eta_p^2 = .33$]. These results demonstrate that process-based transfer occurred above and beyond any instance-based learning, for both tasks.

Sex Differences

Although sex was examined as a variable in all analyses, sex differences were not pervasive. Of most interest, the results suggest that such differences can be attenu-

ated or even eliminated with practice. In analyses of the 3-week practice sessions, RT slopes for women practicing the MRT were initially steeper than those for men [$t(15) = 2.28$, $p = .04$; $\eta_p^2 = .14$], but this sex difference disappeared with practice. However, in analyses of errors that focused on initial, rather than final, laboratory sessions, a task × session × gender interaction was found [$F(1,27) = 5.98$, $p = .03$; $\eta_p^2 = .18$; see Tables 5A and 5B and Figure 7 for details]. In addition, males had higher RTs than did females for the VAT [$F(1,29) = 12.52$, $p = .002$; $\eta_p^2 = .30$]. It is worth noting that although sex was included as a variable in all transfer analyses, no other main effects or interactions occurred with this variable.

DISCUSSION

Although numerous researchers have documented that practice improves performance on particular spatial tasks, there has been disagreement as to whether practice transfers to other spatial tasks. Our findings showed symmetric transfer of practice between spatial tasks. Crucially, we found greater improvement in the nonpracticed spatial task than in the VAT, which implicates spatial processes rather than general factors.

Why has prior work sometimes failed to find evidence of the sorts of transfer we document here? Evaluating the mixed findings on transfer of practice is hampered by lack of homogeneity in methods. For example, the choice of transfer task varies: Some studies employ novel stimuli taken from the practice task, some employ new tasks designed to assess the same cognitive skills as those implicated in

Table 5A
Mean Task Response Times (RTs, in Milliseconds), Slopes, and Intercepts, With Standard Errors, for the Mental Paper-Folding Task (MPFT), the Mental Rotation Task (MRT), and the Verbal Analogies Task (VAT)

	MPFT						MRT						VAT					
	RT		Slope		Intercept		RT		Slope		Intercept		RT		Slope		Intercept	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Initial Session																		
Male	2,948	213	706	98	1,536	212	3,105	127	254	48	2,597	192	3,164	113	278	39	2,609	149
Female	2,690	179	718	86	1,255	154	3,182	151	245	51	2,692	170	2,780	98	292	26	2,196	95
Final Session																		
Male	1,686	227	496	94	694	158	1,832	151	207	53	1,419	132	2,983	97	215	34	2,552	100
Female	1,384	168	471	90	442	108	1,896	189	185	30	1,527	162	2,455	79	225	27	2,005	58

Note—Slope values represent mean increase per level of difficulty. Results are subdivided by sex (male, $n = 14$; female, $n = 17$).

Table 5B
Mean Percentages of Error (%E, With Standard Errors) for the Mental Paper-Folding Task (MPFT), the Mental Rotation Task (MRT), and the Verbal Analogies Task (VAT)

	MPFT						MRT						VAT					
	RT		Slope		Intercept		RT		Slope		Intercept		RT		Slope		Intercept	
	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE	%E	SE
Initial Session																		
Male	19.3	2.9	14.1	2.6	−8.9	3.7	14.2	2.6	3.0	1.6	8.3	3.2	18.9	1.7	12.1	1.2	−5.4	2.6
Female	15.2	1.5	14.3	2.1	−13.4	3.0	18.3	2.1	5.2	1.3	8.0	2.5	16.8	1.1	12.5	1.0	−8.3	2.0
Final Session																		
Male	9.2	2.2	8.1	2.0	−7.0	2.2	7.8	1.6	3.8	0.9	0.2	1.1	18.3	2.1	9.3	0.9	−0.2	2.0
Female	8.4	1.1	7.7	1.5	−7.0	2.1	9.1	1.6	4.9	0.8	−0.7	1.2	16.6	1.4	9.9	1.0	−3.2	1.9

Note—Slope values represent mean increase per level of difficulty. Results are subdivided by sex (male, $n = 14$; female, $n = 17$).
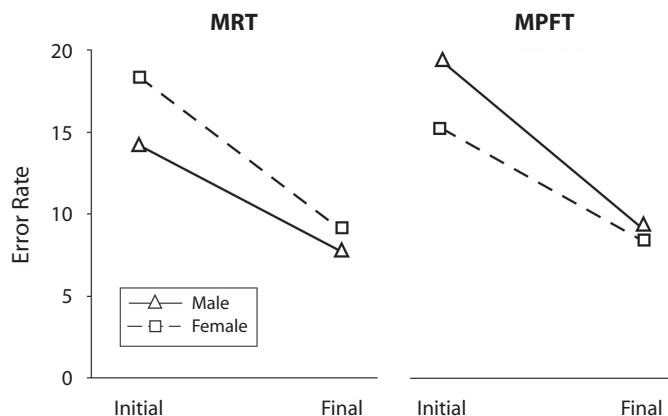
Figure 7. Sex differences in the initial versus final sessions for the mental paper-folding task (MPFT) and the mental rotation task (MRT). Error bars are not shown to minimize clutter. Initially, male participants outperformed females on the MRT and female participants outperformed males on the MPFT. By the final session, however, these differences had disappeared. This symmetric result suggests that proper training can reduce or eliminate such differences, at least with these spatial tasks. The interaction was not found when slopes and intercepts were examined on their own.

the practice tasks, and others use different but cognitively related tasks. Identifying patterns in the rate and degree of transfer within such studies requires careful consideration of the level of cognitive overlap between tasks, along with the relative amounts of practice, as well as other dimensions (Barnett & Ceci, 2002). Differences in the frequency and spacing of practice sessions may also affect the speed of skill acquisition and transfer within cognitive tasks.

In the present study, transfer effects occurred after approximately 7 h of practice across a 3-week period. Similarly, other studies have produced transfer effects following substantial practice. For example, Terlecki, Newcombe, and Little (in press) found transfer to a nonpracticed task following 12–14 h of practice across 14 weeks. Failures of transfer with substantial practice could result from other aspects of the methodology. For example, Heil, Rösler, Link, and Bajric (1998) found no significant transfer to novel objects or novel perspectives of the same object after 4 days of practice, but this appears to reflect the fact that retesting allowed the control group to improve on the transfer items.

An intriguing question with respect to the transfer effects is why they were not greater in the slopes, traditionally considered to specifically index spatial processes. Indeed, we found that effects occurred reliably only on the intercepts. One possibility is that much of the improvement is in the encoding of the stimulus and in the initiation of the mental rotation or folding process. To perform a mental transformation, participants must first encode the internal spatial relations among the parts of the stimulus (Hyun & Luck, 2007) and then correctly initiate transformation (e.g., they must begin rotating the stimulus in the correct direction). These processes would be reflected in the intercept, rather than in the slope. The transformation process itself (rotation or folding per se), once initiated, may not be subject to levels of improvement as dramatic as would encoding and transformation initiation. Here, we evaluated practice

and training effects by decomposing slope and intercept on both RT and error measures. The finding that spatially specific transfer effects appear reliably on intercepts has important theoretical implications because it suggests a reconsideration of the traditional slope–intercept decomposition: More specific spatial skills may be reflected in the intercept than is traditionally believed. This result also suggests that interventions could target the encoding of spatial stimuli, as well as actual spatial transformation. For example, Amorim, Isableu, and Jarraya (2006) found that changes in stimuli that improve object shape matching led to facilitation of mental rotation.

Using methodology guided by the five criteria summarized at the outset, we have shown robust, specific spatial transfer to new stimuli and tasks, in contradiction to a prevalent recent view (CSTS, 2006). Future work might examine the precise circumstances in which such transfer is obtained and the factors that govern its magnitude.

REFERENCES

AMORIM, M.-A., ISABLEU, B., & JARRAYA, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, **135**, 327-347.
BAENNINGER, M., & NEWCOMBE, N. [S.] (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, **20**, 327-344.

BARNETT, S. M., & CECI, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, **128**, 612-637.

BETHELL-FOX, C. E., & SHEPARD, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 12-23.

COMMITTEE ON SUPPORT FOR THINKING SPATIALLY (2006). *Learning to think spatially: GIS as a support system in the K–12 curriculum*. Washington, DC: National Academies Press.

COOPER, L. A., & SHEPARD, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75-176). New York: Academic Press.

DE LISI, R., & CAMMARANO, D. M. (1996). Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behavior*, **12**, 351-361.

HALPERN, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Erlbaum.

HEIL, M., RÖSLER, F., LINK, M., & BAJRIC, J. (1998). What is improved if a mental rotation task is repeated—the efficiency of memory access, or the speed of a transformation routine? *Psychological Research*, **61**, 99-106.

HYUN, J.-S., & LUCK, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin & Review*, **14**, 154-158.

KAIL, R. (1986). The impact of extended practice on rate of mental rotation. *Journal of Experimental Child Psychology*, **42**, 378-391.

LEONE, G., TAINE, M. C., & DROULEZ, J. (1993). The influence of long-term practice on mental rotation of 3-D objects. *Cognitive Brain Research*, **1**, 241-255.

MORRISON, R. G., KRAWCZYK, D. C., HOLYOAK, K. J., HUMMEL, J. E., CHOW, T. W., MILLER, B. L., & KNOWLTON, B. J. (2004). A neuro-computational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, **16**, 260-271.

PETERS, M., & BATTISTA, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain & Cognition*, **66**, 260-264.

SHEPARD, R. N., & FENG, C. (1972). A chronometric study of mental paper folding. *Cognitive Psychology*, **3**, 228-243.

SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

STERICKER, A., & LEVESCONTE, S. (1982). Effect of brief training on sex-related differences in visual–spatial skill. *Journal of Personality & Social Psychology*, **43**, 1018-1029.

TERLECKI, M. S., NEWCOMBE, N. S., & LITTLE, M. (in press). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*.

WALLACE, B., & HOFELICH, B. G. (1992). Process generalization and the prediction of performance on mental imagery tasks. *Memory & Cognition*, **20**, 695-704.