Dimensional transformation in tests of spatial and environmental cognition

Kashka Iwanowska · Daniel Voyer

Abstract In previous research, it has been argued that spatial performance on psychometric tests might be accounted for, in part, by the need for test-takers to transform mentally two-dimensional (2-D) test items into a three-dimensional (3-D) representation. With this in mind, the Landscape Perception Test (LPT) was designed to isolate the 2-D to 3-D (and vice versa) transformational aspect of spatial cognition. Gender differences were used as an indirect means to examine the contribution of the 2-D to 3-D transformation to spatial performance. Since the LPT was designed by means of an image relevant to geospatial cognition, measures of environmental cognition, as well as experience in various activities, were also examined. One hundred thirty undergraduate students (66 females and 64 males) completed the Childhood Activities Questionnaire, the Santa Barbara Sense of Direction Test, the Visualization of Views (VV) Test, the Perspective Taking/Spatial Orientation Test, and the LPT. Results showed higher scores for men than for women on LPT items requiring 2-D to 3-D conversion (LPT23), but not on those requiring the reverse operation. In addition, only the LPT23 produced significant indirect effects of gender on VV Test performance. Performance on the tests was also correlated with environmental cognition and previous experience with spatial activities. The results suggest that dimensional transformation might be a factor significantly contributing to spatial performance. Psychometric aspects of the LPT are also discussed.

Keywords Three-dimensional representation · Environmental cognition · Spatial abilities · Gender differences

K. Iwanowska · D. Voyer (✉)
Department of Psychology, University of New Brunswick,
PO Box 4400, Fredericton, NB, Canada E3B 5A3
e-mail: voyer@unb.ca

Spatial thinking is used implicitly and explicitly in most of our daily activities, such as tying our shoes, as well as when we play music, pack a suitcase, or navigate around our environment. Traditionally, spatial abilities have been investigated using a geometric abstraction of our physical environment, rather than a reasonable approximation of the actual three-dimensional (3-D) environment we live in (Lobben, 2007). In this context, spatial abilities are referred to as "the ability to generate, retain, retrieve and transform well-structured visual images" (Lohman, 1996, p. 98). This kind of research is often carried out using psychometric approaches that emphasize the ability to mentally rotate two-dimensional (2-D) and 3-D shapes and objects (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002). There is an abundance of psychological literature utilizing these mental rotation tests as a measure of spatial ability (Voyer, Voyer, & Bryden, 1995). In fact, mental rotation is considered by some researchers to be the "standard" measure of spatial ability (Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005).

Despite this perception that mental rotation assesses crucial aspects of spatial ability, these tasks are also believed to involve multiple processes beyond the need to mentally rotate (Hegarty & Waller, 2004), and it is possible that some, but not all, of these components relate to spatial cognition in our everyday environment. In fact, evidence from research conducted in virtual environments might help elucidate the processes required on such tests, above and beyond mental rotation. For instance, Parsons et al. (2004) suggested that the increase in the complexity of the stimuli from 2-D to 3-D might activate the use of different strategies in adults when they approach the task. This may partially explain why performance is facilitated in a 3-D virtual environment. Parsons et al. also pointed out that a virtual environment precludes the need to mentally perform 2-D to 3-D transformations—a purported source of difficulties for the pen-and-paper versions of mental rotation tasks (McWilliams, Hamilton, & Muncer, 1997; Voyer & Hou, 2006). Moreover, virtual environments offer an immersive sensory and

perceptual experience, which perhaps reflects human spatial thinking in the real-world environment more accurately.

Considering spatial cognition from a real-world perspective has inspired several researchers to investigate the concept of environmental cognition (Hegarty et al., 2002; Hegarty & Waller, 2004; Kozhevnikov& Hegarty, 2001). Simply speaking, environmental cognition refers to the process of spatial thinking in the geographic context and to the ability to make spatial relations and maintain relationships between objects in geographical space (Golledge, Dougherty, & Bell, 1995; Self & Golledge, 1994). As such, environmental cognition is involved in processes such as orientation of the self in the environment, wayfinding, route tracing, landmark recognition, and self-orientation (Hegarty et al., 2002). The common element among these processes is that they tend to involve mental operations in large- rather than small-scale spaces and are self-referential and egocentric, in that they rely on updating one's location in space on the basis of dynamic positioning of self (Hegarty et al., 2002). Consequently, performance on tasks tapping these environmental, as opposed to abstract, cognitions seems to be better predicted by self-report measures (Kozlowski & Bryant, 1977), rather than by psychometric tests of spatial ability (Hegarty et al., 2002). One possible reason for this finding is that both self-report measures and environmental cognition tasks might involve familiar situations that have a clear meaning for test takers.

From this perspective, it is interesting that several studies point to the idea that meaning and familiarity may facilitate better performance on tests of spatial rotation. Smith and Dror (2001) examined the relative influences of stimulus meaning and familiarity on performance in mental rotation. They concluded that the presence of semantic meaning in a stimulus seems to facilitate effective mental rotation, in that it allows for greater flexibility in choice of strategies. According to Smith and Dror, meaningful stimuli tend to be processed holistically when simple, but in a piecemeal fashion when more complex. Meaningless stimuli, however, tend to be rotated through a holistic process. However, it should be noted that Smith and Dror's interpretation that piecemeal rotation allows for a greater flexibility and accuracy in performing mental transformations has been contradicted by more recent research (Alexander & Son, 2007). Nevertheless, the fact remains that meaningful stimuli are typically processed more accurately and efficiently than meaningless stimuli (see also Amorim, Isableu, & Jarraya, 2006).

Considering the research presented so far, it is possible to propose a framework for investigating the potential influence of factors not related directly to the mental rotation process without the need to resort to a virtual environment. Specifically, we would argue that mental rotation tests typically involve meaningless stimuli that require the mental transformation of 2-D stimuli into 3-D representations. The present study focuses on this mental transformation aspect and proposes an indirect way to examine its influence. Specifically, the finding of substantial gender differences in favor of men in psychometric tests of spatial ability is seen by many researchers as one of the defining characteristics of these measures (Amponsah & Krekling, 1997; Moffat, Hampsom, & Hatzipantelis, 1998; Peters et al., 1995; Voyer et al., 1995). It is interesting to note, however, that these gender differences, while unequivocal in pen-and-paper tests, have been shown to disappear in equivalent tests conducted in virtual reality environments. For example, Parsons et al. (2004) conducted a study with 44 participants utilizing both pen-and-paper and virtual environment versions of mental rotation tasks. The geometric shapes used in both test versions were designed to approximate each other as closely as possible. The virtual environment test required the participants to interact with the working stimulus by rotating it, as quickly and efficiently as possible, to match the target stimulus. Significant gender differences were reported on the pen-and-paper version, but not on the virtual environment test. Similarly, Piburn et al. (2002) administered pen-and-paper and computer-generated tests of mental rotation and surface development (mental paper folding). The tests were equivalent in content. However, the paper-and-pen task had a time limit, whereas the computer task had no explicit time limit. The researchers obtained significant gender differences on the pen-and-paper, but not the computerized, test. In this case, the time pressure difference was confounded with the format manipulation and could explain the lack of gender differences in the computer tests (see Voyer, 2011). In any case, these findings suggest the possibility that much of the gender differences in mental rotation might be due to the need to mentally transform 2-D stimuli into 3-D representations (see also Neubauer, Bergner, & Schatz, 2010, on this point). Therefore, demonstrating that gender differences are significantly reduced in magnitude when this mental transformation is accounted for would support the more general notion that mental rotation test performance partly reflects this process. Following the logic presented by Preacher and Hayes (2004), this is similar to stating that the 2-D to 3-D transformation partly mediates gender differences in psychometric tests of mental rotation. Essentially, support for such mediation would demonstrate the importance of this transformation in general.

In addition to the 2-D to 3-D transformation component, the abstract nature of most standard tests is illustrated by the fact that tests of mental rotation typically use stylized 2-D representations of 3-D objects, such as block configurations with minimal depth cues. In fact, occlusion is often the only cue available, and it actually has been shown to hinder performance on the task (Voyer & Hou, 2006). Accordingly, such images have little bearing on the type of mental rotation

tasks one might need to perform in the real world. In most cases, this means that completing a mental rotation test with 2-D stimuli requires the construction of a 3-D representation on the part of the test-taker on the basis of inappropriate depth cues. In addition, this aspect (2-D to 3-D conversion) was presented by Voyer and Hou as an important component of the observed better performance for males than for females in mental rotation. It is therefore possible that spatial thinking investigated by psychometric tests and real-world spatial thinking might be partly governed by different cognitive processes. This difference could then account for the better performance of men than of women on paper tests, but not in environments more closely resembling the real-world view, purportedly as a function of the dimensional abstraction required in task completion.

The research presented here investigated the relation between abstract spatial ability, as measured with standard spatial tests, and geospatial cognition, as measured with a new test using images of landscapes relying on shading as a depth cue. In the present experiment, we aimed to examine more closely the potential role of the 2-D to 3-D conversion on gender differences in mental rotation as a handle for understanding the role of this conversion in mental rotation performance. A test of landscape perception that offers a closer approximation of an intuitive view of a landscape was designed specifically for this research. A computer-generated rendition of a natural landscape was synthesized in an attempt to depict that landscape in a way that is generally familiar to the viewer (requiring no or minimal abstract image construction). The aim was to present the participant with a view of a natural terrain corresponding to a landscape rather than a line representation of an abstract geometric object. That terrain (a mountainous landscape) was visualized using standard 3-D graphic sun-illumination techniques, with shading proportional to local surface slope and directional light source. Thus, the only clues to relative changes in elevation came from intuitive hill shading. The task required mapping the 3-D view into a 2-D view and vice-versa. In this way, we hoped to have an indirect measure of the 2-D to 3-D conversion process with a minimal mental rotation component. Construction of such a test was necessary since, to our knowledge, past research of seeming relevance has focused on map-reading skills (e. g., Goldberg & Kirman, 1990), thereby involving only 2-D processing. An example of one item from this Landscape Perception Test (LPT) is presented in Fig. 1.

An informal analysis of the items on the LPT allows inferences concerning the processes that might be involved in this test. In a general sense, the test requires skills that cartographers might apply in their everyday work. Specifically, the task involves reading the topography of a terrain by translation of a 2-D target into a 3-D response alternative (and vice versa). This is the base process that the task purports to measure, and it would allow a cartographer to have a sense of the elevation and shape of the terrain. In fact, this would also allow a hiker to obtain a "read" of the terrain if the image referred to a trail map, for example. This emphasizes the real-world relevance of the LPT. However, it is possible that some participants might use mental rotation to some extent in completing the task—for example, to match line position on the landscape with the terrain profile. In addition, some participants might choose to alter their perspective on the alternatives in the course of task completion. In fact, participants could potentially rotate themselves around the alternatives in imagination (perspective change) or rotate the alternatives (mental rotation), as is always a possibility in this type of task (Hegarty & Waller, 2004). Therefore, the LPT should show a relatively large correlation with any mental rotation measure, due mostly to the common 2-D to 3-D transformation component and, to a much smaller extent, to a minimal mental rotation component. In addition, it should show a small significant correlation with a measure of perspective taking.

Another important aspect of the LPT is that half the items were laid out as the sample in Fig. 1 (2-D target and 3-D alternatives), whereas the remaining items required translating a 3-D target into 2-D alternatives. Only the former should reflect the critical process that we intend to measure, but this characteristic also allows a dissociation of the effects of interest. Specifically, if the 2-D to 3-D transformation is actually involved in mental rotation, only 2-D to 3-D items on the LPT should produce results supportive of the role of this factor, and items reflecting the need to perform a 3-D to 2-D transformation should not.

Therefore, our aim in constructing this new test was to answer three specific questions. Specifically, the first question was whether there would be gender differences in performance on the LPT and, if so, what would be their magnitude, as compared with those found on standard tests of spatial ability. Considering that 2-D to 3-D conversion should be the main process involved in the LPT, it was expected that men would perform better than women, but only on items requiring 2-D to 3-D transformation, not on those where the reverse translation (3-D to 2-D) is required (hypothesis 1). However, since the LPT should have a minimal mental rotation component, the magnitude of this gender difference should be smaller than that on a pen-and-paper test of mental rotation, as measured with the Visualization of Views (VV) Test (Hegarty, Keehner, Khooshabeh, & Montello, 2009; hypothesis 2).

The second question focused on whether a correlation would be found between performances on a mental rotation measure and the LPT. We hypothesized that there would be a positive correlation between performances on a mental rotation test and the LPT, presumably due mostly to the common 2-D to 3-D conversion component. Accordingly,
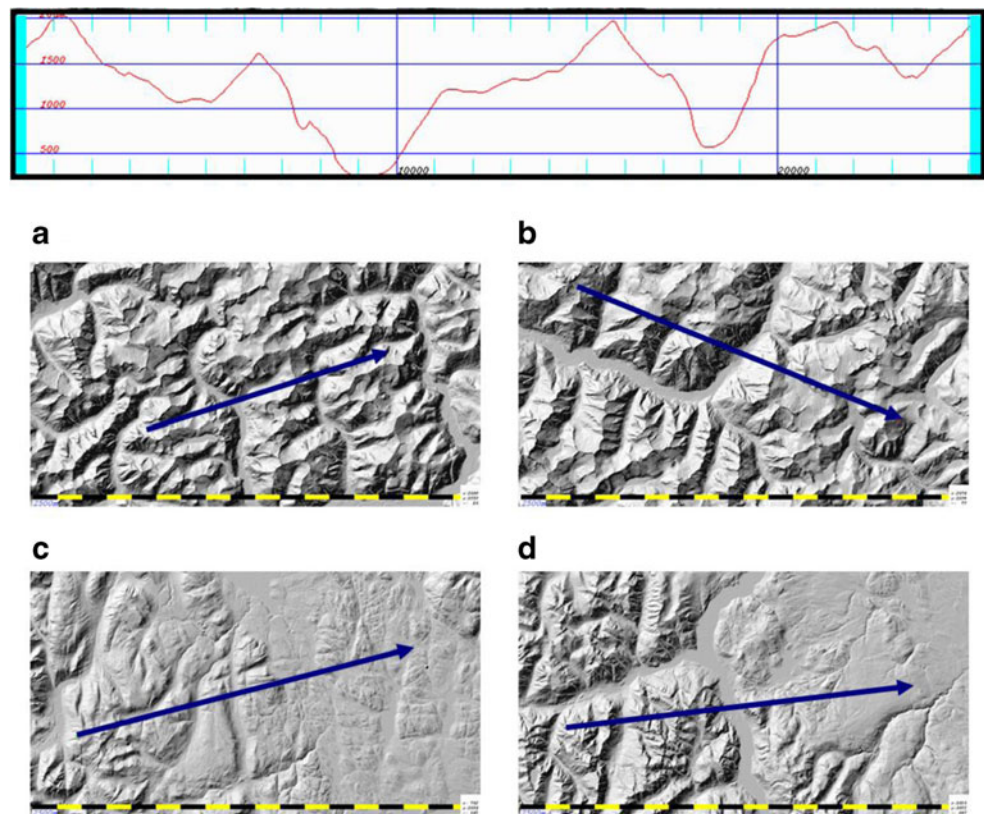
this correlation should be larger for items requiring translation from 2-D to 3-D than the reverse (hypothesis 3).

The third question concerned the possibility that processes underlying performance on the landscape test mediate performance on a test of mental rotation, but only on 2-D to 3-D items, and this should be reflected in the pattern of gender differences. Accordingly, it was hypothesized that performance on 2-D to 3-D items, but not 3-D to 2-D items, should produce significant indirect effects in the relation between gender and performance in mental rotation (hypothesis 4).

In addition to the tests central to the research questions, several other measures were administered to allow a more refined characterization of the processes involved in the LPT. Specifically, a measure of environmental cognition, the Santa Barbara Sense of Direction (SBSOD) Scale (Hegarty et al., 2002), was administered to verify that the landscape test was indeed relevant to environmental cognition. Accordingly, there should be a significant correlation between the SBOD Scale and both components (2-D to 3-D, 3-D to 2-D) of the LPT. In addition, a measure of perspective taking, the Perspective Taking/Spatial Orientation Scale (Hegarty & Waller, 2004), was administered, since it has been shown to assess abilities that are separate from mental rotation, although it does correlate with mental rotation ability. This should verify the expectation that, although the 2-D to 3-D component of the LPT should show a

relatively large correlation with a mental rotation measure, it should also show a smaller correlation with a measure of perspective taking (hypothesis 5).

From the perspective of the positive manifold of intelligence, many of the correlations predicted so far would be expected if only because the various cognitive measures used reflect a general intelligence factor (e.g., van der Maas et al., 2006). However, differences in the magnitude of these correlations are predicted on the basis of the putative components involved in the LPT. In addition, a noncognitive measure assessing childhood experience with various activities, the Childhood Activities Questionnaire (CAQ; Cherney & Voyer, 2010; Doyle, Voyer, & Cherney, 2012), was included in an attempt to replicate the findings of a correlation between this measure and spatial performance. This would confirm the role of experiential factors in the tests included here.

Method

Participants

Sixty-four males and 66 females were recruited from separate sections of introductory psychology classes. The mean age in the sample was 19.28 years ($SD = 2.94$, range = 17–35). Participants received bonus credit in their introductory

psychology course for their voluntary participation. Participants were tested according to proper ethical guidelines, and approval was obtained from the institutional Research Ethics Board.

Materials

All participants were asked to complete a demographic questionnaire that included questions about gender and whether or not they had taken any prior geography courses at high school, college, or university levels.

The SBSOD Scale (Hegarty et al., 2002) is a self-report scale pertaining to environmental spatial cognition and consisting of 15 statements about spatial and navigational abilities, preferences, and experiences. The items are scored on a Likert-type scale from 1 (*strongly agree*) to 7 (*strongly disagree*). Seven of the items are phrased positively (e.g., "I am very good at . . ."), whereas the other eight are phrased negatively (e.g.,, "I am not good at . . ."). The positive items are reverse scored, so that higher total scores on this scale correspond to a better self-reported sense of direction. This measure demonstrated good internal consistency in the present sample ($\alpha = .85$).

The VV Test (Hegarty et al., 2009) was selected as measure of mental rotation.[1] This test consists of 24 questions that ask a participant to identify a viewing position from which a picture of a 3-D object was taken. The object is drawn hovering in the middle of a "glass cube" whose edges are delineated by a dashed line. The target object, from a different viewing perspective, is drawn underneath the cube. The participants are to "move around the cube" to determine from which corner the picture of the object was taken and to circle that corner. The participants are given 8 min to complete the 24 questions. The score is obtained by calculating the number of correct responses, subtracting the number of incorrect responses, which has been divided by six (to correct for guessing). This measure also demonstrated good internal consistency in the present sample ($\alpha = .90$). A sample item from this test is presented in Fig. 2.

---

[1] Some readers might have expected selection of the Mental Rotations Test (Vandenberg & Kuse, 1978) as a measure of mental rotation, since it is likely the most widely used test in the literature (e.g., Peters, 2005; Voyer et al., 1995). However, issues with impossible rotations for response alternatives that cannot be rotated into congruence with the target figure have brought into question the actual contribution of the mental rotation process on this test (Bors & Vigneau, 2010; Kerkman, Wise, & Harwood 2000). In fact, Bors and Vigneau did not find the expected relation between angular disparity and accuracy when analyzing Mental Rotations Test performance. In contrast, in the VV Test (Hegarty et al., 2009), it is always possible to rotate the two block configurations into congruence. Therefore, the VV Test was deemed likely to provide a good measure of mental rotation abilities for the present study, since it also involves a 2-D to 3-D conversion, as is hypothesized in our main research question.
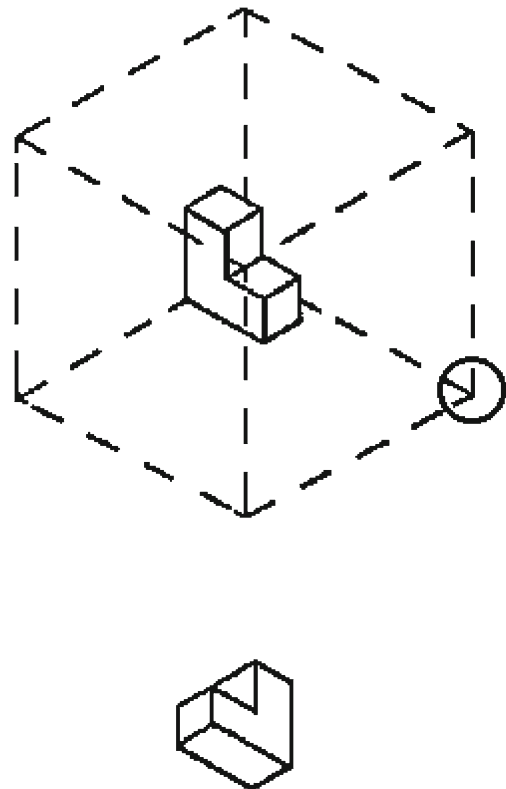
Fig. 2 Sample item from the Visualization of Views Test. The correct response for this item is circled. (Reproduced with permission from Hegarty et al., (2009))

The CAQ (Cherney & Voyer, 2010; Doyle et al., 2012) contains 27 items reflecting the participants' engagement during childhood (3–12 years old) in masculine–spatial, masculine–nonspatial, feminine–spatial, and feminine–nonspatial activities. The items are scored on a visual analog scale represented by a 100-mm-long line. The participants are asked to place an X along the line at the location corresponding to their degree of involvement in that activity during childhood. The starting and end points of the line are labeled *Never* and *Always*, respectively. The location of the X is measured from the start (score of 0) and end (score of 100) in millimeters. The scores on the items reflecting each of the four activity types (masculine–spatial, masculine–nonspatial, feminine–spatial, and feminine–nonspatial) are averaged to obtain the total score corresponding to each of the factors. Doyle et al. used masculine and spatial preference scores, where Masculine $= ln(M_{\text{masculine activities}}/M_{\text{feminine activities}})$ and Spatial $= ln(M_{\text{spatial activities}}/M_{\text{nonspatial activities}})$, to validate this measure. However, Doyle et al. reported that these measures are highly correlated. Therefore, following their approach, only the spatial score was used in data analysis, since it still reflects all the items in the questionnaire. Internal consistency for this measure was very good in the present sample ($\alpha = .83$).

The Perspective Taking/Spatial Orientation Scale (Hegarty & Waller, 2004) consists of 12 questions that test one's ability to imagine relative azimuthal orientation in space. Each question consists of a random collection of objects drawn at the top of the page. At the bottom half of the page, there is a reference circle with an arrow pointing at 0° (12 o'clock). The participants are asked to identify the direction between some of the objects. For example, they are asked to imagine standing at one object (drawn in the middle of the reference circle below), while facing another (drawn on the circle at the top of the 0° line), and then to determine the direction to a third object from that viewing perspective. That direction is to be marked by an arrow from the middle of the circle to a location on the circumference of the circle, corresponding to the azimuth of the direction between the two objects. The participants are allocated 5 min for this test. The score represents the total number of correct responses. No correction for guessing was applied to the score on this test. This measure produced an acceptable level of internal consistency in the present sample (α = .73).

As was previously mentioned, the LPT was developed for the present study. Its aim is to present the participant with a view of a natural terrain corresponding to a landscape. The landscape is visualized using standard 3-D graphic sun-illumination techniques (315° azimuth, 45° elevation angle). The only cues to relative changes in elevation come from shading, which is an important pictorial cue to depth (Foley & Matlin, 2010).

Digital terrain models with resolutions ranging from 25 to 100 m were created using publicly available topographic data from U.S. (http://eros.usgs.gov/#/Home) and Canadian (http://geogratis.cgdi.gc.ca/geogratis/en/index.html) landscapes. A fixed aspect area (910 × 500 pixels) was derived, and a sun-illuminated image was created from this area using the addSUN function from the UNB Ocean Mapping Group swathed software suite (in-house, unpublished). The surface gray-level intensity was derived from the cosine of the angle between the sun vector and the facet surface normal. This replicates a lambertian surface typical of nonspecular illumination models. Cast shadows were not included in the algorithm. Two-dimensional topographic profiles were extracted along sections from 20 to 70 km long of arbitrary azimuth. The sections were plotted with an automatically adjusted vertical scale to maximize the vertical exaggeration.

The test involves a 3-D to 2-D spatial transformation (and vice versa) and includes 20 multiple-choice questions. Participants are told that the 2-D profile represents a cut through the landscape, as if one could take a vertical slice along the indicated line (see top of Fig. 1). Half the questions (even-numbered items) involve matching one of four cross-section profiles to a reference line through a sun-illuminated digital terrain model (3-D to 2-D transformation). The other

half of the questions (odd-numbered items) apply the task in reverse: Given a 2-D cross-section profile, the participants are asked to identify which one of four indicated reference lines across the digital terrain model constitutes the source topography (2-D to 3-D transformation; see Fig. 1). Participants are allowed 10 min to complete this task. The score is computed as the number of correct responses minus the number of incorrect responses, which has been divided by four to correct for guessing. This novel measure demonstrated good internal consistency in the present sample (α = .84).

Procedure

Participants were tested in small groups ranging in size from 1 to 10 persons. Upon arriving at the lab, participants were seated in small cubicles separated by screens, and they completed a consent form. Participants were then given test booklets containing a demographics questionnaire, the SBSOD Test, the CAQ, the VV Test, the Perspectives Taking/Spatial Orientation Test (PTSOT), and the LPT.

The experiment was divided into two parts. The first part was self-paced and involved completing the demographic questionnaire, the SBSOD Test, and the CAQ. The second part commenced upon the signal from the experimenter after all the participants had completed part I. It contained the three timed spatial ability tasks: the VV Test, the PTSOT, and the LPT. The order of the tests in part II was fully counterbalanced.

Results

Preliminary analyses showed that whether participants had taken geography courses in the past had no effect on performance for any of the measures considered here. In addition, a similar proportion of men and women had taken geography courses, $\chi^2(1) = 0.89$, $p > .34$. Accordingly, this variable was excluded from further analyses.

Hypotheses 1 and 2 predicted that men would perform better than women on the LPT, but only on items requiring 2-D to 3-D transformation (hereafter, LPT23), not on those requiring the reverse conversion (3-D to 2-D; hereafter, LPT32) (hypothesis 1), and that this gender difference would be smaller than that on a pen-and-paper test of mental rotation, as measured with the VV Test (hypothesis 2). These hypotheses were examined by means of a multivariate ANOVA. Gender was the independent variable, and the scores on the SBSOD Test, the VV Test, the LPT, and the PTSOT and the spatial score on the CAQ were the dependent variables.

The multivariate test of significance revealed a significant main effect of gender, $F(6, 123) = 19.70$, $p < .01$, based on Pillai's trace. Follow-up univariate F tests revealed a

significant main effect of gender for the SBSOD, $F(1, 128) = 24.70$, $p < .01$, the VV, $F(1, 128) = 36.12$, $p < .01$, the LPT23, $F(1, 128) = 11.84$, $p < .01$, and the CAQ spatial score, $F(1, 128) = 82.25$, $p < .01$, but not for the LPT32, $F(1, 128) = 2.82$, $p > .09$, or the PTSOT, $F(1, 128) = 1.70$, $p > .19$. As can be seen in Table 1, relevant means showed higher scores for men in all cases where gender differences were significant (with higher scores reflecting better performance on the VV and the LPT23 and preference for spatial rather than nonspatial activities on the CAQ). In particular, these findings confirmed hypothesis 1, since significant differences were obtained for LPT23, but not LPT32. Examination of the second hypothesis required calculation of Cohen's d (Cohen, 1977), reflecting the difference between the scores of men and women on the VV and the LPT. This value is also shown for the other measures as a point of information. Cohen's d showed a large effect size for gender differences in VV scores ($d = 0.94$) but a medium effect size for gender differences in LPT23 scores ($d = 0.58$) and a small effect size on LPT32 scores ($d = 0.30$). The second hypothesis is therefore confirmed. It should be noted that the SBSOD and CAQ also produced large effect sizes of 0.80 and 1.24, respectively, whereas for the PTSOT, the effect size ($d = 0.23$) was, not surprisingly, small.

Hypothesis 3, predicting that the correlation between the LPT23 and the VV should be larger than that between LPT32 and VV, can be examined by a direct comparison of the correlations (see Table 2). However, the differences between these two correlations were significant only with a one-tailed test, $z = 1.78$, $p < .038$, one-tailed. This finding supports hypothesis 3. As further points of information, it should be noted that all but three correlations presented in Table 2 (for the SBSOD with LPT32 and PTSOT and for CAQ with PTSOT) were found to be significant.

Hypothesis 4 required a demonstration that LPT23, but not LPT32, mediates the relation between gender and VV

Table 1 Mean scores on dependent variables as a function of gender (standard deviations in parentheses)

Dependent variable	Men	Women	d
VV	14.77 (7.21)	7.70 (6.17)	0.94
SBSOD	56.08 (10.41)	46.11 (12.36)	0.80
PTSOT	7.47 (2.57)	6.71 (3.90)	0.23
LPT23	4.75 (2.15)	3.41 (2.26)	0.58
LPT32	5.12 (2.41)	4.39 (2.50)	0.30
CAQ	0.19 (0.53)	−0.53 (0.38)	1.24

Note. VV = Visualization of Views Test; SBSOD = Santa Barbara Sense of Direction Scale; PTSOT = Perspective Taking/Spatial Orientation Scale; LPT23 = Landscape Perception Test, 2-D to 3-D; LPT32 = Landscape Perception Test, 3-D to 2-D; CAQ = Childhood Activities Questionnaire spatial score. d represents men's scores minus women's scores, divided by the pooled standard deviation (see J. Cohen, 1977).

Table 2 Correlations among the various measures ($N = 130$)

	LPT32	VV	PTSOT	SBSOD	CAQ
LPT23	.510*	.463*	.266*	.272*	.330*
LPT32		.323*	.248*	.103	.246*
VV			.370*	.289*	.332*
PTSOT				.082	.139
SBSOD					.410*

Note. LPT23 = Landscape Perception Test, 2-D to 3-D; LPT32 = Landscape Perception Test, 3-D to 2-D; VV = Visualization of Views Test; PTSOT = Perspective Taking/Spatial Orientation Scale; SBSOD = Santa Barbara Sense of Direction Scale; CAQ = Childhood Activities Questionnaire spatial score.

* $p < .01$

performance. Essentially, this hypothesis would be confirmed by demonstrating significant indirect effects with LPT23, but not LPT32, as mediator. The analysis required for a test of hypothesis 4 was computed as recommend by Preacher and Hayes (2004) and relied on the SPSS macros provided by these authors. Results of this analysis showed significant indirect effects when LPT23 was used as a mediator of the relation between gender and VV performance (estimate of indirect effect = 1.5652; 95 % confidence intervals based on 1,000 bootstrap resamples = 0.6532–2.6062; $p < .01$). In contrast, indirect effects were not significant when LPT32 was used as mediator (estimate of indirect effect = 0.5737; 95 % confidence intervals based on 1,000 bootstrap resamples = −0.1978–1.4030; $p > .14$). This confirms hypothesis 4. However, it should be noted that gender differences on the VV remained significant even after accounting for LPT23 performance, $b = 5.50$, $t(126) = 4.83$, $p < .01$. The present results would therefore be described as reflecting partial mediation (Preacher & Hayes, 2004).

Finally, hypothesis 5, predicting that the 2-D to 3-D component of the LPT23 would show a relatively large correlation with a mental rotation measure (VV) but a smaller correlation with a measure of perspective taking (PTSOT), was confirmed by showing that the differences between these two correlations were significant, $z = 2.19$, $p < .03$, two-tailed. The remaining correlations between the LPT23 and all the other measures support the notion that the cognitive tests involve similar processes (i.e., positive manifold), whereas its correlation with the CAQ suggests an experiential basis to performance.

Discussion

The purpose of the present experiment was to investigate the relationship between abstract spatial ability, as represented in standard tests of these abilities, and environmental cognition, as represented by a self-report measure and a novel

test designed to tap the ability to convert 2-D stimuli into 3-D representations (and vice versa). More specifically, the aim was to isolate the 2-D to 3-D component in a measure of abstract spatial ability by means of a novel measure that integrated this component in a geospatial task. The well-known finding of gender differences in spatial performance was exploited to go beyond simple correlations and provide an indirect way to measure the role of the 2-D to 3-D conversion required in performing an abstract spatial task. Building on research by Parsons et al. (2004) and Piburn et al. (2002), who reported gender differences on pen-and-paper tests of spatial ability, but not on equivalent tests conducted in virtual reality environments, we posited that these findings were due to the differential ability regarding the 2-D to 3-D mental transformation. The LPT, designed specifically for the present study, focused on this abstraction and tested performance on 2-D to 3-D and 3-D to 2-D transformation tasks.

As was expected, men performed better than women on items requiring 2-D to 3-D transformation, not on those requiring the reverse conversion on the LPT, and this gender difference was smaller than that on the VV Test. These findings support the notion that men are more efficient than women on the processes underlying the conversion from a 2-D image to a 3-D representation, suggesting that gender differences in abstract tasks of spatial abilities might be partly due to the required 2-D to 3-D conversion in many of them. At a more general level, the present results fit well with findings that have emphasized the 3-D nature of the stimuli in explaining gender differences in mental rotation (Voyer & Hou, 2006).

The observed correlations suggest that participants also had to understand the perspective that the line in the landscape represented as reflected in the significant correlation between performance on both components of the LPT and performance on the PTSOT. Finally, the presence of a minimal mental rotation component on the LPT is supported by the finding that the VV Test produced larger gender differences than did our novel test. Therefore, our results support the expectations that the LPT includes mostly the transformation component, with minimal mental rotation and perspective-taking components (based on the observed pattern of correlations).

As was predicted, the 2-D to 3-D component of the LPT produced a larger correlation with the VV Test than did the 3-D to 2-D component. The fact that this was only significant on the basis of a one-tailed test of significance could be considered a partial confirmation, although, since this was the predicted direction of the effect, a one-tailed test was also sufficient (Furlong, Lovelace, & Lovelace, 2000). This suggests that the 2-D to 3-D component accounts for variance in performance over and above the mental rotation component that would be common to both types of items.

From this perspective, it would be interesting to have participants complete both tests (VV and LPT) in a virtual reality setting, which would preclude the need to mentally perform the 2-D to 3-D conversion. Presumably, these two tests should still be correlated significantly on both components (2-D to 3-D and 3-D to 2-D). However, the correlation for 2-D to 3-D items would likely drop to the same level as that observed for 3-D to 2-D items. In addition, if gender differences in 2-D to 3-D transformation are critical, the gender difference in favor of men should disappear on the 2-D to 3-D items, and it should be substantially reduced on the VV Test. Future work should investigate this question.

Likely the most critical finding observed here was that performance on 2-D to 3-D items, but not 3-D to 2-D items, produced significant indirect effects in the relation between gender and performance on the VV Test. This further supports the notion that component processes involved in performance of the LPT are relevant to performance on the VV Test. This finding is easily explained with the converging evidence presented so far. Specifically, it is plausible to believe that the need to convert a 2-D image to a 3-D representation is common to both tasks. Therefore, statistically controlling for this aspect significantly decreases the magnitude of gender differences on the VV Test. However, the fact that gender differences did remain significant on the VV Test suggests that components over and above the 2-D to 3-D conversion account for gender differences on this test. Of course, the mental rotation component itself is one of the most obvious candidates on this point. In fact, although the VV Test has been presented by some as a measure of perspective taking (e.g., Hegarty et al., 2009), only a mental rotation component could account for the large magnitude of gender differences observed here on this test, especially considering the absence of gender differences on the PTSOT in the present study (on this point, see also Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Zacks, Mires, Tversky, & Hazeltine, 2000). This suggests the possibility that participants favor a mental rotation strategy rather than a perspective change strategy on the VV Test.

The 2-D to 3-D component of the LPT showed a relatively large correlation with the VV Test but a significantly smaller correlation with the PTSOT. This suggests that, as was expected, the LPT likely has a perspective-taking component in common with the PTSOT. However, as also was expected, the combination of mental rotation and 2-D to 3-D conversion components in common with the VV test produced a larger correlation than that observed when only the perspective component was in common.

Confirmation of all five hypotheses therefore supports the results of our informal analysis of the processes involved in the LPT (major 2-D to 3-D or 3-D to 2-D transformation component, minimal mental rotation, and perspective-taking

components). In addition, the results support the notion that 2-D to 3-D conversion is an integral part of performance (and gender differences) in a psychometric measure of mental rotation. This was demonstrated through a task that, we felt, had real-life relevance, since it involves skills that might be used by cartographers or hikers in their respective activities. Accordingly, the present results can be added to the body of evidence demonstrating that tasks with real-life applicability relate to psychometric test performance (Hambrick et al., 2012; Liu, Oman, Natapoff, & Coleman, 2008; Menchaca-Brandan, Liu, Oman, & Natapoff, 2007).

Aside from those of relevance to the hypotheses, other correlations require some brief discussion. Specifically, the correlation between the SBSOD Test and the PTSOT failed to achieve significance. This likely reflects the fact that the Santa Barbara test has a strong real-life navigation component (Hegarty et al., 2002), whereas the PTSOT measure is more relevant to an abstract manipulation of egocentric orientation (Hegarty & Waller, 2004). In contrast, the correlations presented in Table 2 suggest that the PTSOT might share a common component with the VV Test and the 2-D to 3-D aspect of the LPT. It is also interesting to note that both the VV Test and the 2-D to 3-D LPT seem to include a navigational component, as is shown in their relatively strong correlation with the SBSOD Test. Inasmuch as the Santa Barbara test reflects real-life spatial abilities in general (Hegarty et al., 2002), the pattern of findings would support the conclusion that the VV Test and the 2-D to 3-D component of the landscape test are relevant to real-life spatial skills. However, the lack of correlation of the 3-D to 2-D component with the SBSOD Test might simply reflect the fact that it is quite rare to have to perform a 3-D to 2-D transformation from a picture. Therefore, this transformation would have little in common with the real-life skills that its authors claimed are assessed by the SBSOD Test.

It is worth noting that all the measures administered here were significantly correlated with the score on the CAQ, except for the PTSOT. This provides further support for the notion that spatial performance in adulthood is affected by activities preference and practice in childhood (Baenninger & Newcombe, 1989; Doyle et al., 2012).

Another finding requiring some discussion is that the SBSOD Scale produced higher scores in men than in women. This suggests that environmental spatial abilities also favor males. However, it is important to remember that, since this is a self-report measure, it would also be affected by confidence and the influence of social desirability associated with stereotype. Specifically, men are typically more confident than women concerning their spatial abilities (Cooke-Simpson & Voyer, 2007), and this is likely to be reflected in their self-report concerning related performance. Similarly, women might stereotypically view themselves as less skilled than men on the particular items they have to

complete on the Santa Barbara tests (Wraga, Duncan, Jacobs, Helt, & Church, 2006), and it is possible that it affected responses on this test for both men and women. Experimental manipulation of this aspect with the SBSOD Scale, similar to the approach used by Wraga et al. with mental rotation, would elucidate this point.

In conclusion, the present study relied on a newly developed test presumed to tap more directly into processes required in converting a 2-D image into a 3-D representation. Altogether, the results support our contention that the 2-D to 3-D component influences performance on a mental rotation test and, potentially, on other psychometric measures of spatial ability relying on 3-D stimuli, although this remains an empirical question. In addition, the findings suggest better performance for men than for women on the 2-D to 3-D component, and they support the notion that this component might also partly underlie gender differences in a task with a strong mental rotation component. Future work with the novel task should attempt to ascertain the role of the alleged transformation component (from 2-D to 3-D) by implementing its administration in a virtual reality setting, for example. In addition, a full construct validity study of the LPT would provide more in-depth information concerning the processes it involves. However, in the meantime, the study presented here establishes the usefulness of this test as a measure of geospatial cognition and as a means for isolating the 2-D to 3-D component in psychometric measures of cognitive abilities.

References

Alexander, G. M., & Son, T. (2007). Androgens and eye movements in women and men during a test of mental rotation ability. *Hormones and Behavior, 52,* 197–204. doi:10.1016/j.yhbeh.2007.01.011

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. doi:10.1037/0096-3445.135.3.327

Amponsah, B., & Krekling, S. (1997). Sex differences in visual-spatial performance among Ghanaian and Norwegian adults. *Journal of Cross-Cultural Psychology, 28,* 81–92. doi:10.1177/0022022197281005

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20,* 327–344.

Bors, D. A., & Vigneau, F. (2010). Sex differences on the mental rotations test: An analysis of item types. *Learning and Individual Differences, 21,* 129–132. doi:10.1016/j.lindif.2010.09.014

Cherney, I. D., & Voyer, D. (2010). Development of a spatial activity questionnaire I: Items identification. *Sex Roles, 62,* 89–99. doi:10.1007/s11199-009-9710-9

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Vol. 2nd). New York: Academic Press.

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17,* 181–186. doi:10.1016/j.lindif.2007.03.009

Doyle, R. A., Voyer, D., & Cherney, I. (2012). The relation between childhood spatial activities and spatial abilities in adulthood. *Journal of Applied Developmental Psychology, 33,* 112–120. doi:10.1016/j.appdev.2012.01.002

Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior, 47*(3), 326–335. doi:10.1016/j.yhbeh.2004.11.013

Foley, H. J., & Matlin, M. W. (2010). *Sensation and perception* (5th ed.). New York: Pearson Education.

Furlong, N., Lovelace, E., & Lovelace, K. (2000). *Research methods and statistics: An integrated approach.* New York: Wadsworth.

Goldberg, J., & Kirman, J. M. (1990). Sex-related differences in learning to interpret Landsat images and in road map reading in young adolescents. *Journal of Geography, 89,* 15–25.

Golledge, R. G., Dougherty, V., & Bell, S. (1995). Acquiring spatial knowledge: Survey versus route-based knowledge in unfamiliar environments. *Annals of the Association of American Geographers, 85,* 134–158.

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., et al. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology. General, 141,* 397–403. doi:10.1037/a00025927

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial ability enhances, and is enhanced by, dental education. *Learning and Individual Differences, 19,* 61–70. doi:10.1016/j.lindif.2008.04.006

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34,* 151–176. doi:10.1016/j.intell.2005.09.005

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30,* 425–447. doi:10.1016/S0160-2896(02)00116-2

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32,* 175–191. doi:10.1016/j.intell.2003.12.001

Kerkman, D. D., Wise, J. C., & Harwood, E. A. (2000). Impossible "mental rotation" problems: A mismeasure of women's spatial abilities? *Learning and Individual Differences, 12,* 253–269. doi:10.1016/S1041-6080(01)00039-5

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object-manipulation spatial ability and spatial orientation ability. *Memory and Cognition, 29,* 745–756. doi:10.3758/BF03200477

Kozlowski, L. T., & Bryant, K. J. (1977). Sense of direction, spatial orientation, and cognitive maps. *Journal of Experimental Psychology. Human Perception and Performance, 3,* 590–598. doi:10.1037/0096-1523.3.4.590

Liu, A. M., Oman, C. M., Natapoff, A., & Coleman, C. (2008). Spatial ability as a predictor of space robotics training performance. *Aviation, Space, and Environmental Medicine, 79,* 288–289.

Lobben, A. (2007). Navigational map reading: Predicting performance and identifying relative influence of map-related abilities. *Annals of the Association of American Geographers, 97,* 64–85. doi:10.1111/j.1467-8306.2007.00524.x

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Mahwah, NJ: Erlbaum.

McWilliams, W., Hamilton, C. J., & Muncer, S. J. (1997). On mental rotation in three dimensions. *Perceptual and Motor Skills, 85,* 297–298.

Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., & Natapoff, A. (2007). Influence of Perspective-Taking and Mental Rotation Abilities in Space

Teleoperation. *Proceedings. of the 2007 ACM Conference on Human-Robot Interaction, Washington, DC,* pp. 271–278.

Moffat, S. D., Hampsom, E., & Hatzipantelis, M. (1998). Navigation in a virtual maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior, 19,* 73–87.

Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence, 38,* 529–539. doi:10.1016/j.intell.2010.06.001

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J. G., et al. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42*(4), 555–562. doi:10.1016/j.neuropsychologia.2003.08.014

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57,* 176–184. doi:10.1016/j.bandc.2004.08.052

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28,* 39–58. doi:10.1006/brcg.1995.1032

Piburn, M, D., Reynolds, S. J., Leeds, D. E., McAuliffe, C. M., Birk, J. P., & Johnson, J. K. (2002). The hidden Earth: Visualization of geologic features and their subsurface geometry. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching. New Orleans, 7–10 April. Retrieved from http://reynolds.asu.edu/pubs/NARST_final.pdf

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717–731. doi:10.3758/BF03206553

Self, C., & Golledge, R. (1994). Sex-related differences in spatial ability: What every geography educator should know. *Journal of Geography, 93,* 234–243.

Smith, W., & Dror, I. E. (2001). The role of meaning and familiarity in mental transformation. *Psychonomic Bulletin & Review, 8*(4), 732. doi:10.3758/BF03196211

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113,* 842–861. doi:10.1037/0033-295X.113.4.842

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599–604.

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin and Review, 18,* 267–277. doi:10.3758/s13423-010-0042-0

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. *Canadian Journal of Experimental Psychology, 60,* 91–100. doi:10.1080/13803390591004310

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250–270. doi:10.1037/0033-2909.117.2.250

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13,* 813–819. doi:10.3758/BF03194002

Zacks, J. M., Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformation of objects and perspective. *Spatial Cognition and Computation, 2,* 315–332. doi:10.1023/A:1015584100204